THE AUTONOMOUS SYSTEMS LABORATORY
UNIVERSIDAD POLITÉCNICA DE MADRID

# The Operative Mind: a functional, computational and modelling approach to machine consciousness

*Published in IJMC Vol. 1, Issue 1, Jun. 2009*

Carlos Hernández, Ignacio López, and Ricardo Sanz

ASLab A-2009-XX v 1.0 Final
September 2007

**Abstract**

The functional capabilities that consciousness seems to provide to biological systems can supply valuable principles in the design of more autonomous and robust technical systems. These functional concepts keep a notable similarity to those underlying the notion of operating system in software engineering, which allows us to specialize the computer metaphor of the mind into that of the operating system metaphor for consciousness. In this article, departing from these ideas and a model-based theoretical framework for cognition, we present an architectural proposal for machine consciousness, called the Operative Mind. According to it, machine consciousness would be implemented as a set of services, in an operative system fashion, based on modelling of the own control architecture, that supervise the adequacy of the system architectural structure to the current objectives, triggering and managing adaptativity mechanisms.

## 1 Introduction

In addition to the everlasting multidisciplinary quest for autonomy in technology, for which a universal design solution remains elusive despite the great efforts and

partial advancements in AI, robotics or control engineering, other aspects, such as the need for augmented dependabilitySanz et al. (2007a) or sustaining performance**?**, are crucial for building reliable artificial systems rendering higher levels of functionality and performance in hazardous and uncertain environments. We hypothesize that consciousness provides the most advanced biological systems—in terms of evolution of a central nervous system—with additional features that precisely help maintaining and even improving their operation in environments with a high degree of uncertainty.

In this article we do not mean to account for a theory of biological consciousness, but to provide a useful approach to the problem of machine consciousness. We are presenting architectural ideas for building machines possessing a certain functionality or properties scientists currently associate with conscious phenomena. The relevant point here, from our engineering perspective, is the advanced properties in terms of robust autonomy these machines will have as a result of implementing those *consciousness* characteristics.

We will not therefore try to address questions such as the ultimate nature of consciousness, or if is it a process or a state of the brain or the body, not only because they are somehow artificial questions given the complex and gray-scaled nature of the phenomenon termed as *consciousness*, but because as far as we are concerned in *machine consciousness*, it will be what we design or build. We leave to others—neuroscientists, biologists, psychologists, philosophers—to decide whether what we theorize is of relevance in the explanation of the biological phenomenon, and to anothers—engineers—if it is useful for building more robust and autonomous technical systems. Of course we hope our theory will be valuable regarding both.

The article is organized as follows: section 2 puts in a nutshell the concepts involved in conscious phenomena according to the literature, from a functional perspective. Section 3 is an analysis of the similarities between consciousness and operating systems that have motivated our approach to machine consciousness based in the *operating system metaphor*—akin the paradigmatic computer metaphor—, called the Operative Mind, which is presented in section 4. To conclude, in the last section we compare the presented approach with previous models of consciousness and other approaches to machine consciousness.

## 2 A Functional Approach to Consciousness

As many researchers claim, consciousness seems to have an evolutionary value Damasio (1998), i.e. it provides biological systems possessing it with some functions or capabilities that suppose an advantage over those lacking it. We will join their functional view of consciousness and leave at a side the problem of phenomenal consciousness and qualia in this reflection on machine consciousness. We hypothesize that those are not but the result of some of the functions involved in con-

sciousness operating within brains, so their explanation can be reduced[1] to them.

We can find a profusion of terms for referring to conscious phenomena in different disciplines: vegetative state, coma, sleep or wakefulness in medicine, reportability, attention or voluntary control in psychology or the mentioned phenomenology and qualia in philosophy are a few examples. They refer to precise clinical states—i.e. medical terms—, cognitive functions—in psychology— or simply buzzwords —i.e. self, awareness—. A good part of the confusion around consciousness is precisely due to the relaxed use of these concepts, which pertain to different levels of abstraction and domains, with specific context-dependent meanings and connotations. Terms and concepts from different realms usually refer to similar parts of the problem of consciousness, but their mapping is usually less than perfect so their loose use results misleading. As Sloman and Chrisley put it Sloman and Chrisley (2003):

> "it (*consciousness*) is a cluster concept, in that it refers to a collection of loosely-related and ill defined phenomena."

What follows is a reasonable list of the referred phenomena, synopsized from a cross-domain perspective:

**Awareness of the world:** It is usually explained as the access to some information that is used to control/generate behaviour Chalmers (1997). A theoretical approach already formulated by Craik Craik (1943) considers that this information is actually an internalised model the agent has of the surrounding world.

**Self(–awareness):** As commonly understood, the problem of the self has two strands: one involving the differentiation one's own from the rest of the world with the related sense of agency, and the other one comprising the identity of oneself as a result of development, like the record of autobiographical memories which render personality in humansDamasio (2000).

**Attention:** The term consciousness is often conflated with attention in the literature, thus promoting confusion **?**. However some authors neatly distinguish them while preserving their deep relationTaylor (2002): the psychological phenomenon of attention is generally regarded as the selective process responsible for deciding which contents in the mind become conscious. However, other authorsSommerhoff (2000) consider that attention is also a selective process not for deciding which enters consciousness, but for focusing on the more relevant contents within those already conscious, the rest forming a background.

[1] it is worth remarking that in our commitment to reductionism we are not making less of anything, but:
**re·duce – 5a:** to bring to a systematic form or character <*reduce natural events to laws*> *Merriam-Webster Ditionary*

**Voluntary control:** This aspect is closely coupled with the already referred sense of agency. There is a clear common sense distinction between involuntary actions, like kicking when hit in the knee or an spontaneous smile, and voluntary ones like rising an arm because of deciding so, the later "voluntary" control being equivalent to "conscious" control **?** in the line of James' *ideomotor theory* of controlling action as a result of bringing to consciousness the desired goal.

**Introspection:** Our ability to observe our own mental and emotional processes is one of the most puzzling aspects of consciousness, with many theories trying to account for mental states whose objects are other mental states while avoiding the *homunculus* trap. It is related to inner speech and imagery, and is also referred to as reflection, which can be considered a special kind of access to some intellectual resources **?**.

## 2.1 Functions of Consciousness

Notwithstanding the previous discussion, the relations and couplings between the above aspects of consciousness, and the interdisciplinar use of the terms used to talk about them suggest that there are some common core principles underlying these phenomena. Being engineers in the pursue of methods for building more autonomous technology by applying useful principles underlying natural systems, and not committed to the mimicking of the material realization in animals, we are interested in the functional concepts, rather than in any concrete physical substrate of consciousness[2].

Therefore we shall now make a summary of the more relevant functions, according to our perspective, identified so far by the theories developed in the search for explanation of the previous aspects of consciousness. We will try to present them as a sound set of functional concepts as general as possible, by abstracting from the domain specific details, and as far as possible clearly differentiated, separating intermingled concepts, while preserving the terminology used in the literature.

**Access:** Many theories on consciousness assign it the role a blackboard has in the so named architectures in artificial intelligence, which is that of allowing the different processes running in the system—i.e. the mind—to put their content at disposal of the rest by means of a broadcasting mechanism. This is, for example, the main hypothesis underlying Baar's Global Workspace Theory-Baars (2001).

**Sequentiality:** The serial and limited character of consciousness could be a mechanism for guaranteeing the *consistency* and unity in the mental contents**?**. The sequential character of conscious contents could also be involved in our sense

[2] referred to as the neural correlate in the specialized literature

of time, allowing for the temporal analysis of perceptions as AnceauAnceau (1999) proposes.

**Integration:** An important function attributed to consciousness is that of integrating multiple sensory input into a single unified experienceBaars (2002). SommerhoffSommerhoff (2000) states that consciousness is precisely an Integral Global Representation (IGR), a functional unit that integrates representations of the fact that first-order representations of sensory inputs and stimuli are part of the state of the organism.

**Meta-representation:** Other of the more common ideas about what consciousness is (or how does it work) is that of structures in the mind/brain representing other structures in the mind/brain. That is Damasio's idea of conscious creatures constructing images of a part of themselves forming images of something elseDamasio (1998), Singer's meta-representations of the brain's own computational operationsSinger (2000) or Sommerhoff's representations in the IGRSommerhoff (2000).

**Metareasoning:** This feature of consciousness is strongly related with the previous one, and refers to the capacity of conscious brains to operate upon their own operations, for example monitor and reason about them or evaluate their performance, as SingerSinger (2000) proposes. Other authors separate high-level (meta-) cognitive processes, such as reasoning or long term memory from consciousness, but keep them related because, due to these processes being very resource demanding, only conscious contents have access to themBaars (2001). François Anceau goes further by proposing that the role consciousness is providing the underlying mechanisms—i.e. the previously enumerated functions—that subserve the functioning of those high level processesAnceau (1999).

**Evaluation:** Some authors relate consciousness to value assigning systems in the brain, e.g. the mentioned view of Singer associating it to a certain monitoring at a metalevel which provides the brain with the capability of comparing the performance of its operationsSinger (2000), or the evaluation of plans by affective states, bringing up the close interconnection between emotion and consciousnessAleksander and Dunmall (2003).

**Learning:** while the learning process is itself unconscious, there is strong evidence for learning of conscious events and no robust one so far for long-term learning of unconscious inputBaars (2002). There seems to be a relation between events or entities entering consciousness and our capacity to learn themBaars (2001).

The previous list is not exhaustive, but it captures the functional hypothesis about consciousness put forward in the literature which we judge more valuable for our account of machine consciousness.

Some authorsChalmers (1997) claim that most of the presented concepts are functional notions whose phenomenological counterparts remain to be addressed. That point of view takes a dualist approach to consciousness, by splitting it into a functional and a phenomenological phenomenon. Since it is not a scientifically sustainable position—not actually much more than that of those rejecting a functional approach to natural phenomena in generalSearle (2000)—we will just expend the next paragraph to refute it.

Suppose we are natural philosophers back in the stone age and, instead of consciousness, we are currently much more interested in another natural phenomena: that of sunrise and sunset. Now consider we elaborate a ludicrous theory of it, involving living on a spheric gigantic object which is rotating, and the existence of a punctual source of light far away. It sounds really ridiculous, does not it? Any anti-functionalist would claim that sunrise is not the function of emitting light! And sure sunset is not that of rotating!

## 3 The Operating System Metaphor

Current paradigms in cognitive science, artificial intelligence, or cognitive robotics—i.e. embodiment, enactivism, etc.—tend to dismiss the computer metaphor of the mind by claiming that minds are not computers. It is obvious that there are no biological organisms controlled by a dual-core processor. However we do not find any solid argument preventing the use of computing technology to address the engineering problems of artificial intelligence or machine consciousness, and, however, we do see a solid argument for it: we have successfully mastered that technology, which appears the most suitable for these enterprises up to date.

Since the problem of consciousness has been identified and isolated as the core remaining problem of mind Sommerhoff (1996) still untamed **?**, some authors have specialized the metaphor to consciousness. These theories consider consciousness as the *operating system of the mind* because of its role of providing certain general services to the specific, unconscious processes running in the mindJohnson-Laird (1988) and to higher level cognitive functions, such as reasoning, language, etc.Anceau (1999)

From our engineering perspective, we join their position by noticing the similarities between the list of concepts, properties, mechanisms and functions related to consciousness presented in the previous section, and the functionality operating systems render in today's computer systems. In this section, after recapitulating the main concepts of operating systems, we will analyze in detail their resemblances to those of conscious phenomena.

## 3.1 Operating Systems

Firstly let us make a brief overview of what an operating system (OS) is in computer science, and what functionality it provides to computing systems, which, it is worth remembering, represent the result of the earlier attempts to create artificial intelligence—i.e. artificial minds**?**.

The definition of operating system is a quite fuzzy one—a first similarity with consciousness—, and there is not even full agreement on whether in a concrete computer system a certain piece of software is part of the OS or not. But the concept of OS is clearly characterized for what an OS *does*, rather by what it *is*Silberschatz and Galvin (1994)—second similarity, according to our functional approach to consciousness—. An operating system provides:

- efficient operation and
- convenience of use

of the computer system.

Regarding the first basic feature, an operating system can be considered as a *resource allocator*. Different processes running at a time in a computer may require concurrently the same resource: CPU time, memory space, access to input/output devices etc. The OS acts as a manager and allocates these resources to the programs requiring them as necessary, looking for the most efficient allocation so as to maximize computer's performance.

The convenience of use should not be regarded as a second order objective just concerning a friendly interface for human users. The user of the computing system can be any agent demanding a service from the computing system: it may be a human user, a process within the computing system or a process within another computer system connected through a network or even any other type of physical entity properly interfaced to it. The OS is responsible for furnishing them with services providing an appropriate execution and development environment. Two important parameters of an OS are the time between the service is requested and a response to it is initiated, and the time elapsed between the system starts processing a response to the demand and a result is delivered.

## 3.2 Operating Systems and Consciousness

We shall now examine the resemblances between the objectives and mechanisms of an operating system with those presented in section 2.1 associated to consciousness. As previously commented, in computing systems it is a core responsibility of the operating system to allocate computational resources—CPU cycles, memory, file storage, I/O access—to the multiple processes running at the same time,

in a way that optimizes their usage and the performance of the system. Likewise, as mentioned in section 2 *attention* is considered the selective processes that determines which contents—from perceptions, inner thoughts—become conscious, and hence which are allocated motor control, reasoning or any other high level resources. This selection process of attention is, notwithstanding, performed at an unconscious level, and it optimizes the use of the living system's resources—both physical and mentalHernández (2008)—in the purchase of its goals, ultimately survival, as an operating system does with computational resources.

The other main objective of an OS, convenience of use, involves providing an environment suited for the development and execution of programs, by issuing general services such as interface with hardware, synchronization, handling of errors, managing of the file system, etc. AnceauAnceau (1999) assigns a similar role to consciousness, specially from an evolutionary perspective:

> "Consciousness could be seen as an environment for high-level functions. This environment makes possible the very existence of high-level brain functions (intelligence, long-term memory, reasoning, etc.) by given synchronization mechanism to them."
>
> *F. Anceau*, Consciousness seen as a framework for high level cerebral functions, August 8, TSC 2001

Once presented the similitudes between the general purposes of consciousness and operating systems, let us now review in more detail the concrete functions an OS performs for realizing them, and at the same time analyze their correspondence to those, listed in 2.1, we presume underlay consciousness:

**Communication:** One of the responsibilities of an OS is providing communication mechanisms so that computing processes can interchange information and cooperate, such as message passing—allowing for implementation of unicast, multicast or broadcast modes—and synchronization services.

Consciousness, in its turn, possesses the characteristic feature of *access* as, by which it acts as a broadcasting channel so that unconscious processes can communicate with each others and cooperate: accessing the results of others, requesting help for an operation, etc.

**Timing:** An operating system provides timing services. It is not just that it supplies the current universal time, an OS also supplies coherent timestamping for serially ordering of processes or timers to avoid processes running indefinitely, besides *synchronization* methods for cooperation of processes, in relation to communication.

It could plausibly be the case that human consciousness subserved similarly timing consistency by reason of its serial character, as already mentioned.

**Coherence and consistency:** An OS provides mechanisms so that data accessed by different processes is modified consistently and data residing in different storage systems is properly updated in all of them.

Correspondingly, the limited capacity Baars (2001) and sequential Anceau (1999) characters of consciousness seem to guarantee the coherence and consistency of conscious contents in the human mind.

**Monitoring:** an operating system is responsible for monitoring the processes running so as to detect errors and handle them as soon as they occur. Errors can be due to many reasons: a process trying to access a prohibited memory location or issuing an illegal I/O instruction, a power failure, a damaged optical device. The OS must take the appropriate actions in each case to ensure correct and consistent computing.

Detection of errors in the mind seems to be performed by unconscious processes, however it causes the contents related to the error to become conscious, so as other processes can access that information in order to identify the error and correct it **?**. In this case the "consciousness" of the error provides more flexibility for the recovery processes that if it were performed unconsciously, because that way the error is broadcast to many cognitive processes that could elaborate a custom response, instead of being given a prefixed one. In addition, the conscious character of errors is related to our capacity of learning from them Baars (2002).

**Meta-management:** an operating system manages the execution of processes in the computer. In addition to monitoring or resource allocation, which can be considered as forms of meta-management, advanced OS such as those named real-time perform advanced meta-management such as priority handling, in which a processes is allocated resources by analysis of its relations to other processes.

The ideas of broad access, attention, meta-representation and metareasoning in the phenomenon of consciousness seem to play a similar role of deciding and handling the execution of the mental routines adequately for the circumstances.

Notwithstanding the previous similarities, it is worth pointing out that, differing from an operating system, consciousness does not seem to completely isolate body control from the cognitive processes in the mind by providing an intermediate and unavoidable layer, as most OS do separating programs from the computing hardware. In fact, most biological body regulation—breathing, heart pace, or internal homeostatic processes—is unconscious. However, voluntary (conscious) control does provide a high level control of motor actions whose execution rely upon unconscious movement patterns and reflexes, separating high level cognitive functions such as planning, simulation, reasoning, from the low-level control signals to muscles, which remain unconscious.

# 4 The Operative Mind

In this section we will present a theoretical approach to machine consciousness, which could be synthesized by the motto *consciousness as a model based operating system*. From a general perspective this can be labelled as a reductionist[3], functional, computational and model–based approach. We have already addressed and given arguments for all but the last adjective. Following we explain that one before concretely describing this proposal, the Operative Mind (OM), for building more robust autonomous machines by applying architectural ideas inspired on biological consciousness.

It is important to remark that the validity of our proposal for machine consciousness does not depend upon the extent of the resemblance the software concept of operating system may bear with the natural phenomena. It is only used as an explaining metaphor. As soon as it shall no longer be of clarification utility, we will freely abandon it. Our model should only be assessed i) in an engineering base, considering its usefulness for the design of artifacts with higher levels of robust autonomy, and ii) in an scientific basis, regarding the explanatory value of the underlying hypothesis about the natural phenomenon of consciousness, which are to be tested and validated upon experimental data. In this article only i) is addressed.

## 4.1 Model-based Cognition

As previously commented our Operative Mind architecture for machine consciousness is part of a modelling approach to cognition, in the line of RosenRosen (1993) or Conant&AshbyConant and Ashby (1970), which we have called the ASys Framework Sanz et al. (2007a). From a control engineering perspective, we consider cognition as the exploitation of knowledge—i.e. models—to realize control. We claim that we can equate knowledge and models and state that: Sanz et al. (2007b). Departing from the basic principle: *a system is said to be cognitive if it exploits models of other systems in their interaction with them*, we have started building up the ASys principled approach to cognition and consciousnessSanz et al. (2007b), which we shall summarize now:

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

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

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

[3] according to the entry by Merriam-Webster in page 3 footnote

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

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

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

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

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

From now on in this section we will use the terms awareness and consciousness strictly according to their definitions above. When it could lead to confusion because us referring to the common understanding of these terms, we will put them in *slanted type* to indicate they are used as normally literature.

## 4.2 The OM Architecture

The Operative Mind (OM) is an architectural framework—in the line of RCSAlbus (1991) or CogAffSloman and Chrisley (2003)—for engineering systems which implement, as we claim, analog functional capabilities to those listed in section 2.1 of biological consciousness, and, as a result, will have improved autonomy and robustness. *Consciousness* is implemented on it as a set of services, in an operative system fashion, based on modelling of the own control architecture, that supervise the adequacy of its structure to the current objectives in the given environment**?**, triggering and managing adaptativity mechanisms. This is the reason we condense the Operative Mind in the motto *consciousness as a model based operating system*.

To ground the presented ideas, we will give a sketchy view of how they will map in the case of developing the control architecture of a mobile robot, which is a very common testbed for research in machine consciousness and cognitive robotics in general; e.g. for the task of patrolling an area of military facilities.

The control architecture based on OM would consist of a network of "cognitive" *nodes*, similarly to Albus RCS architectureAlbus (1991), but in this case each one

realizing a control loop following the pattern of the epistemic control loop**?** (Fig. 1) and not necessarily organized into a hierarchy, but will be connected as required by the current task and global state—i.e. system state and environment state as perceived by the system—. These nodes will have different spatiotemporal resolutions and span several levels of abstraction. They could include for our example: a node controlling a pan-tilt camera, which would receive as input a stream of video from the camera and would output the location and shape of salient entities to other nodes, together with the commands to direct the camera towards the most relevant location, a node responsible for processing the sonar lectures to build a 2D map of the surrounding environment and generating reactive obstacle-avoidance motor responses, a higher abstraction node integrating information coming from the previous ones into a medium-term map and generating the path to follow, another node performing simultaneous localization and mapping, and so on.

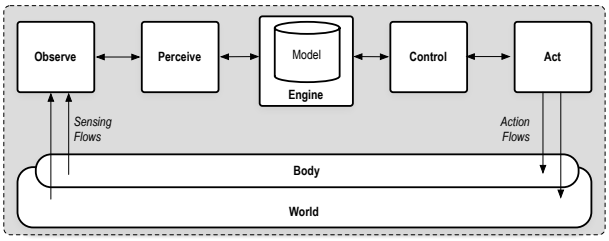


*Figure 1: The core epistemic control loop is the minimal structure providing general model-based cognition.*

**Modelling**

According to the ASys principled approach to cognition, each node in the architecture maintains and exploits models of the world and the robot physicality, to the extent relevant for its operation. For example, there could be a node maintaining a model of the environment consisting of 2D map of the whole area known by the robot, another maintaining a short term 3D map of the current surroundings, a different node could maintain a model of the state of the robot: battery charge, the pressure on the wheels, and so on. This models are *federated* through the network, so that any node can make use of them. This is the first reason why the models must be *explicit*, that is coded in a common modelling language as far as possible.

The OM system also maintains functional models of the nodes, containing information of what they do, what interfaces they have to other nodes, what computational resources they consume or their performance—execution time, algorithm optimality, etc.—. This models are used by *meta-nodes*, which this way can *monitor* and *control* the operation of nodes, and are also patterned after the epistemic control loop 2. More interestingly, meta-nodes are also explicitly modelled, so they can operate upon themselves, closing the controlling–the–controller regression loop.

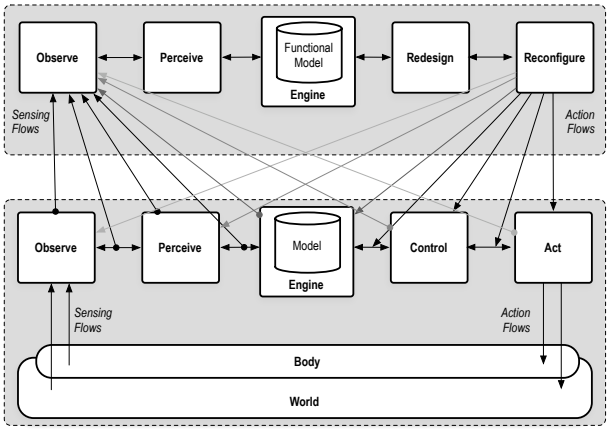


*Figure 2: A meta-node follows the same pattern of the epistemic control loop, but in this case its "plant", following control jargon, is not the physical system, but a controller node.*

The OM frameworks poses thus high level requirements to the modelling language to be used:

- It must be able to capture dynamics.
- It must be able to capture both physical and abstract or conceptual entities.
- It must be executable for simulation.
- It must be able to capture the self, so the same language used by engineers to design the system shall be used by the system to model itself.
- It must include quantifiable metrics for evaluation.

**Model evaluation: generating meaning**

Awareness involves evaluation models according to the current state of objectives. As that can spread to several nodes it is only possible as far as that models are explicit. Our so claimed meaning generation would not be produced by an static generation of value in virtue of the present state as captured by models within the system, but rather of dynamic evaluation of simulations of possible evolutions of the system, by forward execution of models of the robot as such in its environment and of the coupled controlling architecture within its computational environment, as "expected"[4] according to the system past. That evaluation realized in view of the current hierarchy of system's objectives, which is dynamic upon the task**?**. This is

[4] in the sense of Sommerhoff's *expectancies*.

in strong relation with the theory behind model predictive control systems. For example, suppose our patrolling robot pass by the open door of an equipment store while going to meet an scheduled patrolling check-point. Provided with an OM architecture the robot would evaluate the conflicting objectives of being on time or investigating the state of the store, by anticipating the delay it would incur if entering to inspect. In addition it should consider that the delay would be different if summarily checking the store by taking images at predefined points within the store, reached by simple path planning and reactive obstacle avoidance, or exploring it in detail performing SLAM. These alternatives would suppose different configurations of the controlling architecture—resource usage, interconnection of domain nodes, etc.— so that the simulation would include the execution of models of the architecture's own performance.

**_Consciousness_ as a matter of degree: attention to select modelling depth and breadth**

In our view of *cognition* and *consciousness* as modelling and self-modelling and evaluation, a key idea is that of consciousness as something which can vary within a degree; in an OM system it is a quantizable graduation in modelling. We shall better explain it with an example, e.g. the experience of perceiving a building by a *conscious* human and our OM-based robot. Of course we say we are *conscious* of it as a whole, walls, doors and windows included. Despite being part of the *conscious* experience, we do not experience one of the building's doors in the same way as we do when we focus our attention on it, our *expectancies* in Sommerhoff's senseSommerhoff (2000) are not the same. What we propose is that our robot would be conscious of the building by instantiating a model of it and the associated self-model and by performing evaluations on them, at the resolution level of the building concept. In this state the robot would exploit models about the building having doors and what are they for in relation to it, but it would not be till an attention shift towards a window, e.g. as a result of an inference process with the goal of searching for the presence of people in the building, that a more complex model of the door containing information of its functional state—closed, open, broken—, and the possibility to change it for example; and hence we would say the robot becomes aware of the door and, if including models and evaluations of its own processing in the perception of that door, eventually conscious. The selection of modelling depth and breadth directed by attentional mechanisms in OM are therefore our proposal to tackle the frame problem.

**Attention: managing limited resources**

Attentional mechanisms at the architecture level, implemented in meta-nodes, apply algorithms to the models of the nodes and, taking into account the current state of the system, obtaining by monitoring, and its current hierarchy of objectives, decide the configuration of the controlling architecture, in terms allocation of

resources—access to sensors and actuators, memory, CPU—, connections between nodes, including synchronization, communication mode—unicast, multicast, or even broadcast for critical error signals—. Since the number of possible configurations is $O(2^{num.nodes^2})$, the search of an optimal one by means of simulation—forward execution of models of nodes— is time consuming, and therefore it is a main responsibility for attentional mechanisms to reduce the search space and focus the simulation effort so as to reach acceptable solutions in time.

## 5 OM and previous models of consciousness

Our approach is inspired by different theories of consciousness, and shares some considerations with others. It is worthy characterizing our Operative Mind in relation to them, and also compare it with other frameworks for machine consciousness.

Due to the operating system metaphor, an immediate comparison of our approach is with that of Johnson-LairdJohnson-Laird (1988). However it is not difficult to tell them apart because he did not referred to the software concept of the term operating system when comparing consciousness with one, as we do. He used the term "operating system" for the highest process in a hierarchy of cognitive processes. It would receive messages that represent the world from the processors in lower levels and would send messages to them to communicate its plans. That conscious process is, according to him, ontologically different from the unconscious processes:

> The conscious process is the serial process of explicitly structured symbols, whereas the unconscious are parallel processing of distributed symbolic representations.
>
> *P. N. Johnson-Laird*Johnson-Laird (1988)

Despite sharing with Johnson-Laird the ontological difference between consciousness and the unconscious cognitive processes, we do not consider it the highest level in an hypothetical mental hierarchy but rather something transversal to it in the line of AnceauAnceau (1999): consciousness seems to provide the scaffolding for minds—running on brains—developing more complex and advanced cognitive capabilities and integrating them, as an OS is the infrastructure for more powerful computing applications running on computers. However Anceau refutes any direct relation between consciousness and meaning generation—further that those with any other high-level mental process—, while our modelling view builds up precisely on it. Besides, he considers the sequential character of consciousness as a key character of consciousness, whereas we consider that as a side-effect of limited

capacity in the human brain. The limited character that results in the need for attentional mechanisms is surely universal to any cognitive/computing system with finite resources, but the serial solution of the human consciousness for coping with it and guaranteeing coherence is not; software engineering provides other solutions, such as synchronization mechanisms, that preserves processing distribution and concurrence.

In relation to more biological accounts for consciousness, the Operative Mind framework for machine consciosness has been influenced by Sommerhoff's and Baars' theories. With Sommerhoff's theory of the Integrated Global RepresentationSommerhoff (2000) we agree in the need for integration of both models of the environment and agent's physicality and of the "mental" architecture of the agent. Baars' Global Workspace Theory and Franklin's implementation of it in the IDA architecture **??**, have been a major inspiration for our OM architecture, especially in relation with the view of consciousness as providing global services by broadcasting information.

Regarding other approaches to machine consciousness, our model-based framework for cognition shares some basic ideas with Holland's control engineering grounded proposal based on modellingHolland and Goodman (2003), but he advocates for a dissociation between the self-model of the agent and the model of the world, whereas we propose an integrated federation of explicit models. There is also a complementary resemblance with Sloman's view of mind as a *virtual machine*: the meta-management component of his CogAff schemaSloman and Chrisley (2003) performs much the same role as the services of our Operative Mind architecture.

## 6 Final comments

This article has analyzed to some detail the resemblance of biological consciousness to an hypothetical operative system for the mind, an idea already proposed by other authors. Departing from that similitudes and a modelling framework of cognition, we have developed an architectural framework for the design of conscious systems, called the Operative Mind. We shall now make a brief comment regarding some of the exposed ideas.

As a line of research in the pursue of building more autonomous and robust systems, machine consciousness must not be overrestricted by its biological counterpart, we do not want the conscious control system of a chemical plant not allowing for the production of more than a product at a time because of its limited attentional capacity, for example. We propose that, despite providing a plausible explanation of human consciousness in terms of mechanisms evolved for synchronization and adaptation to a causal world Anceau (1999), the sequential character of consciousness is not necessary for machine consciousness, maybe other techniques for assur-

ing consistency and cohesion can be used, together with alternative mechanisms for dealing with time, while maintaining distributed and parallel processing.

The presented operating system metaphor for consciousness should not limit the architectural guidelines for building artificial consciousness neither. The proposed Operative Mind framework could be considered an operative system for cognitive machines, but its pervasive modelling approach adds more powerful mechanisms, when compared to current "services" in software systems, that we claim will render true conscious capabilities.

## References


Albus, J. S. (1991). Outline for a theory of intelligence. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3):473–509.

Aleksander, I. and Dunmall, B. (2003). Axioms and tests for the presence of minimal consciousness in agents. *Journal of Consciousness Studies*, 10:7–18.

Anceau, F. (1999). *Vers une étude objective de la conscience*. Hermes.

Baars, B. J. (2001). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, 6(1):47–52. PMID: 11849615.

Chalmers, D. J. (1997). *The Conscious Mind*. Oxford University Press.

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

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

Damasio, A. R. (1998). Investigating the biology of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1377):1879–1882. PMC1692416.

Damasio, A. R. (2000). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Vintage.

Hernández, C. (2008). Adding consciousness to cognitive architectures. Master's thesis, Dpto. Automática, Ing. Electrónica e Informática Industrial, Universidad Politécnica de Madrid.

Holland, O. and Goodman, R. (2003). Robots with internal models. *Journal of Consciousness Studies*, 10(4-5):77–109.

Johnson-Laird, P. (1988). *The Computer and the Mind: An Introduction to Cognitive Science*. Harvard University Press.

Rosen, R. (1993). On models and modeling. *Applied Mathematics and Computation*, 2-3(56):359–372.

Sanz, R., López, I., and Hernández, C. (2007a). Self-awareness in real-time cognitive control architectures. In *Proc. AAAI Fall Symposium on Consciousness and Artificial Intelligence: Theoretical foundations and current approaches*, Washington DC. AAAI.

Sanz, R., López, I., Rodríguez, M., and Hernández, C. (2007b). Principles for consciousness in integrated cognitive control. *Neural Networks*, 20(9):938–946.

Searle (2000). Consciousness. *Annual Review of Neuroscience*, 23:557–578.

Silberschatz, A. and Galvin, P. (1994). *Operating System Concepts*. Addison Wesley, fourth edition.

Singer, W. (2000). *Phenomenal awareness and consciousness from a neurobiological perspective*, pages 121–137. The MIT Press.

Sloman, A. and Chrisley, R. (2003). Virtual machines and consciousness. *Journal of Consciousness Studies*, 10(4-5):133–172.

Sommerhoff, G. (1996). Consciousness explained as an internal integrating system. *Journal of Consciousness Studies*, 3(2):139–157.

Sommerhoff, G. (2000). *Understanding Consciousness. Its Function and Brain Processes*. SAGE Publications Ltd.

Taylor, J. G. (2002). Paying attention to consciousness. *Trends in Cognitive Sciences*, 6(5):206–210.