Artificial Intelligence and Consciousness

Antonio Chella[1], Riccardo Manzotti[2]

[1]Department of Computer Engineering, University of Palermo
Viale delle Scienze, 90128, Palermo, Italy
chella@unipa.it
[2]Institute of Communication, Enterprise Consumption and Behaviour,
IULM University, via Carlo Bo, 20143, Milan, Italy
riccardo.manzotti@iulm.it

Morpheus : We have only bits and pieces of information, but what we know for certain is that some point in the early twenty-first century all of mankind was united in celebration. We marvelled at our own magnificence as we gave birth to AI
Neo : AI - you mean Artificial Intelligence?
Morpheus : A singular consciousness that spawned an entire race of machines.

from The Matrix script, 1999

Abstract

Consciousness is no longer a threatening notion in the community of AI. In human beings, consciousness corresponds to a collection of different features of human cognition. AI researchers are interested in understanding and, whether achievable, to replicate them in agents. It is fair to claim that there is a broad consensus about the distinction between the phenomenal and the cognitive aspects of consciousness (sometimes referred to as P- and A-Consciousness). The former problem is somehow related with the so-called hard problem and requires deep theoretical commitments. The latter problems are more frequently if not successfully addressed by the AI community. By and large, several models are competing to suggest the better way to deal with cognitive aspects of consciousness allegedly missing from traditional AI implementations. The AAAI Symposium on AI and Consciousness presents an overview on the current state of the art of consciousness inspired AI research.

Introduction

In the last ten years there has been a growing interest towards the field of consciousness. Several researchers, also from traditional Artificial Intelligence, addressed the hypothesis of designing and implementing models for artificial consciousness (sometimes referred to as machine consciousness or synthetic consciousness) – on one hand there is hope of being able to design a model for consciousness, on the other hand the actual implementations of such models could be helpful for understanding consciousness (Minsky, 1985; Baars, 1988; Franklin, 1995; McCarthy, 1995; Aleksander, 2000; Edelman and Tononi, 2000; Jennings, 2000; Aleksander,

2001; Baars, 2002; Franklin, 2003; Kuipers, 2005; Adami, 2006; Minsky, 2006; Chella and Manzotti, 2007).

At the same time, trying to implement a conscious machine is a feasible approach to the scientific understanding of consciousness itself. Edelman and Tononi wrote that: "to understand the mental we may have to invent further ways of looking at brains. We may even have to synthesize artifacts resembling brains connected to bodily functions in order fully to understand those processes. Although the day when we shall be able to create such conscious artifacts is far off we may have to make them before we deeply understand the processes of thought itself." (Edelman and Tononi, 2000).

The field of artificial consciousness naturally emerges from AI and consciousness debate.

According to Holland (Holland, 2003), it is possible to distinguish between *Weak Artificial Consciousness* and *Strong Artificial Consciousness*. Holland defines them as follow:

1. *Weak Artificial Consciousness*: design and construction of machine that simulates consciousness or cognitive processes usually correlated with consciousness.

2. *Strong Artificial Consciousness*: design and construction of conscious machines.

Most of the people currently working in AI could embrace the former definition. Anyhow, the boundaries between the twos are not easy to define. For instance, if a machine could exhibit all behaviors normally associated with a conscious being, would it be a conscious machine? Are there behaviors which are uniquely correlated with consciousness?

Other authors claim that the understanding of consciousness could provide a better foundation for complex control whenever autonomy has to be achieved. In this respect, consciousness-inspired architectures could be, at least in principle, applied to all kind of complex

systems ranging from a petrochemical plant to a complex network of computers. The complexity of current artificial systems is such that outperforms traditional control techniques.

According to Sanz, there are three motivations to pursue artificial consciousness (Sanz, 2005):

- implementing and designing machines resembling human beings (cognitive robotics);

- understanding the nature of consciousness (cognitive science);

- implementing and designing more efficient control systems.

The current generation of systems for man-machine interaction shows impressive performances with respect to the mechanics and the control of movements; see for example the anthropomorphic robots produced by the Japanese companies and universities. However, these robots, currently at the state of the art, present only limited capabilities of perception, reasoning and action in novel and unstructured environments. Moreover, the capabilities of user-robot interaction are standardized and well defined. A new generation of robots and softbots aimed at interacting with humans in an unconstrained environment shall need a better awareness of their surroundings and of the relevant events, objects, and agents. In short, the new generation of robots and softbots shall need some form of artificial consciousness.

AI has long avoided facing consciousness in agents. Yet consciousness corresponds to many aspects of human cognition which are essential to our highest cognitive capabilities: autonomy, resilience, phenomenal experience, learning, attention. In the past, it was customary to approach the human mind as if consciousness was an unnecessary gadget that could be added at the end. The prevailing attitude regarded consciousness as a confused name for a set of ill-posed problems. Yet, classic AI faced many problems and, under many respects, similar problems are now facing neural networks, robotics, autonomous agents and control systems. The fact is that AI systems, although a lot more advanced than in the past, still fall short of human agents. Consciousness could be the missing step in the ladder from current artificial agents to human like agents. A few years ago, John Haugeland asked whether (Haugeland, 1985/1997, p. 247): "Could be consciousness a theoretical time bomb ticking away in the belly of AI?"

Consciousness inspired AI research

Consciousness inspired AI research appeared in many different contexts ranging from robotics to control systems, from simulation to autonomous systems.

Epigenetic robotics and synthetic approaches to robotics based on psychological and biological models have elicited many of the differences between the artificial and mental studies of consciousness, while the importance of the interaction between the brain, the body and the surrounding environment has been pointed out (Chrisley, 2003; Rockwell, 2005; Chella and Manzotti, 2007; Manzotti, 2007).

In the field of AI there has been a considerable interest towards consciousness. Marvin Minsky was one of the firsts to point out that "some machines are already potentially more conscious than are people, and that further enhancements would be relatively easy to make. However, this does not imply that those machines would thereby, automatically, become much more intelligent. This is because it is one thing to have access to data, but another thing to know how to make good use of it." (Minsky, 1991)

The target of researchers involved in recent work on artificial consciousness is twofold: the nature of phenomenal consciousness (the so-called *hard* problem) and the active role of consciousness in controlling and planning the behaviour of an agent. We do not know yet if it is possible to solve the two aspects separately.

Most mammals seem to show some kind of consciousness – in particular, human beings. Therefore, it is highly probable that the kind of cognitive architecture responsible for consciousness has some evolutionary advantage. Although it is still difficult to single out a precise functional role for consciousness, many believe that consciousness endorses more robust autonomy, higher resilience, more general capability for problem-solving, reflexivity, and self-awareness (Atkinson, Thomas et al., 2000; McDermott, 2001; Franklin, 2003; Bongard, Zykov et al., 2006).

There are a few research areas that cooperate and compete in order to outline the framework of this new field: embodiment; simulation and depiction; environmentalism or externalism, extended control theory. None of them is completely independent of the others. They strive to reach a higher level of integration.

Embodiment tries to address the issues of symbol grounding, anchoring, and intentionality. Recent work emphasizing the role of embodiment in grounding conscious experience goes beyond the insights of Brooksian embodied AI and discussions of symbol grounding (Harnad, 1990; Harnad, 1995; Ziemke, 2001; Holland, 2003; Bongard, Zykov et al., 2006). On this view, a crucial role for the body in an artificial consciousness is to provide the unified, meaning-giving locus required to support and justify attributions of coherent experience in the first place.

Simulation and depiction deal with synthetic phenomenology developing models of mental imagery, attention, working memory. Progress has been made in understanding how imagination- and simulation-guided action (Hesslow, 2003), along with the *virtual reality metaphor* (Revonsuo, 1995), are crucial components of being a system that is usefully characterized as conscious. Correspondingly, a significant part of the recent resurgence of interest in machine consciousness has focused on giving such capacities to robotic systems (e.g., Cotterill, 1995;

Stein and Meredith, 1999; Chella, Gaglio et al., 2001; Ziemke, 2001; Hesslow, 2002; Taylor, 2002; Haikonen, 2003; Holland, 2003; Aleksander and Morton, 2005; Shanahan, 2005).

Environmentalism focuses on the integration between the agent and its environment. The problem of situatedness can be addressed adopting the externalism view where the vehicles enabling consciousness extend themselves to part of the environment (Drestke, 2000; O' Regan and Noe, 2001; Nöe, 2004; Manzotti, 2006).

Finally, there is a strong overlapping between current control theory of very complex system and the role that is played by a conscious mind. A fruitful approach could be the study of artificial consciousness as a kind of extended control loop (Chella, Gaglio et al., 2001; Sanz, 2005; Bongard, Zykov et al., 2006)

There have also been proposals that AI systems may be well-suited or even necessary for the specification of the contents of consciousness (synthetic phenomenology), which is notoriously difficult to do with natural language (Chrisley, 1995).

One line of thought (Dennett, 1991; McDermott, 2001; Sloman, 2003) sees the primary task in explaining consciousness to be the explanation of consciousness talk, or representations of oneself and others as conscious. On such a view, the key to developing artificial consciousness is to develop an agent that, perhaps due to its own complexity combined with a need to self-monitor, finds a use for thinking of itself (or others) as having experiential states.

AAAI Symposium on AI and consciousness

The AAAI Symposium on AI and Consciousness aims at providing an overview on the current state of the art of consciousness inspired AI research. In the course of the symposium, contributors present both the experimental result, the theoretical foundations of this emerging field, and their relationship with traditional Artificial Intelligence. As it is to be expected in a field as new and controversial like consciousness, many of the authors defend views which are conflicting. Yet this is positive and desirable since it is the hallmark of a new theoretical horizon. At the present stage, it is paramount that all possibilities are explored.

A first group of contributors focus on the theoretical issues highlighting the relation between consciousness and other approaches like artificial intelligence, cognitive science, cognitive science, neuroscience, and philosophy of mind. These authors offer a survey of recent research in the philosophy, psychology and neuroscience of consciousness and how it can inform AI, potential for AI to inform consciousness research, how the machine consciousness approach is more than just a re-packaging of work already being done in AI (Aleksander, 2000; Chrisley, 2003; Manzotti, 2006; Chella and Manzotti, 2007).

Ron Chrisley and **Joel Parthermore** address the issue of mental content. Linguistic means fails in capturing nonconceptul aspects of experience. They suggest using depictions in a strongly embedded context: the generation of depictions through the use of an embodied, perceiving and acting agent, either virtual or real. They rest on similar work of Igor Aleksander and aim at providing a working model of a robot exploiting synthetic phenomenology. And yet, what is an image or a representation?

Antonio Chella and **Salvatore Gaglio** focus on the role of self consciousness in AI. How to give a robot the capabilities of self-consciousness – i.e. to reflect about itself, its own perceptions and actions during its operating life. They claim that robot self-consciousness is based on higher order perception of the robot. In their model there is the outer world and the inner world inside a robot. Self consciousness is the perception of such inner world. They describe in details an implementation of such a model which rests on the robotic platform developed by the Robotics Lab of the University of Palermo. It is an architecture divided in three modules: a subconceptual, a conceptual, and a linguistic one. The robot is capable of perceiving its status and thus of creating a higher order representation of itself.

In their paper, **Riccardo Manzotti** and **Vincenzo Tagliasco** maintain that insofar consciousness has been scientifically intractable because an internalist stance had been uncritically adopted. Instead an externalist standpoint could be used. They trace an overview of different forms of externalism from Gibsons' ecological approach to David Chalmers and Andy Clark's extended mind. They suggest adopting a process oriented approach that does not require assuming a separation between the subject and the object. According to them the subject is a bundle of processes taking place between the environment and the internal structure of the agent's brain. After considering a series of classic cases (phenomenal experience, after images, dreams), they present a model of conscious agent. In order to be entangled in the right way with the environment, an agent needs to be teleologically open – namely capable of developing new goals and using them for its development. If they would prove right, their model would not require any kind of biological neural activity like those implicitly assumed by most of the neural correlates of consciousness biased literature. Therefore they claim that it is a viable approach for designing robots with phenomenal consciousness.

A strong critique of the link between AI and consciousness studies is advanced by **Stevan Harnad** and **Peter Scherzer**. They deal with the phenomenal nature of consciousness – to be conscious is to feel something. What is the nature of feeling and how can feeling have any causal efficacy? They challenge the AI attempt at providing models of consciousness, "there are no implementation issues inspired by consciousness. There are just internal structures and processes that we over interpret mentalistically." Luckily, they do not rule out the

possibility of building a device that will pass the human Turing test, at least in the far future.

On a more positive note, **Benjamin Kuipers** advances a strategy for sneaking up the hard problem. He believes that a computational approach could be the key for modelling a conscious agent. The trick could be the capability of using the huge amount of information of the sensory stream in order to develop a coherent sequential narrative describing the agent's sensorimotor interaction with the world. The agent will choose intentionally what events are more relevant for its future behaviour. According to him, a conscious agent with a subjective first-person view of the world is an embodied agent, with several processes, a coherent sequential narrative and the capability of continuously evaluating new hypothesis about the world. Kuipers discusses whether his proposal may solve the hard problem; he claims that it could help in understanding its computational foundations.

Sidney K. D'Mello and **Stan Franklin** present a dense analysis of the foundation of consciousness studies inside the AI framework. Are there theoretical foundations for the study of consciousness in AI systems? Can cognitive architectures that include consciousness be of use to AI? Can such AI cognitive architectures add to our knowledge of consciousness in humans and animals? Are phenomenally conscious AI systems even possible? The paper embodies the authors' suggestive, hypothetical and sometimes speculative attempts to answer these questions. Interestingly, although the authors rest most of their work on cognitive approaches to consciousness (IDA, CLARION, models based on Baars' Global Workspace), they believe that such implementations give some hope that phenomenally conscious cognitive robots might be designed and built

Two major obstacles stand in front of AI and consciousness: intentionality and phenomenal experience. **Igor Alexander** and **Helen Morton** defend an articulated theoretical standpoint that aims at solving both. Their view rest on the view the phenomenology and intentionality are closely related. Their paper concerns formalizing the 'gap' between first and third person visual experience in the context of AI. Their framework is called Axiomatic Consciousness Theory and is based on five introspectively-found interlocking components of experienced consciousness: presence, imagination, attention, volition, and emotion. The authors are among the few that explicitly address the issue of phenomenal consciousness by suggesting a hybrid approach merging introspective analysis of phenomenal experience and computational models. They also present a simulation of their model that could provide a useful starting point for research aimed at synthetic phenomenology, although serves mainly as an example. Since AI is often criticized for not addressing the issue of phenomenology, they hope that the Axiomatic Consciousness Theory concepts will take AI forward towards phenomenology.

Is there a link between AI and the hard problem? **Piotr Boltuc** and **Nicholas Boltuc** tries to answer to this question analyzing various foundational aspects of consciousness. What is the relation between cognition and consciousness and are there aspects of consciousness that cannot be reduced to cognitive aspects.

Fiora Pirri discusses the *neural correlates of consciousness* from a more classic AI point of view. She analyzes the fact that the *minimum* conscious event must be a *representational* event. Besides she refers to the debate between global and local content. Can we have simple explicit phenomenal representation of single events or do we need a global workspace of the state of the agents in order to have local representations of particular state of things?

Susan Stuart discusses the question "will conscious machines perform better than unconscious machines?" In this respect her main discussion topic is the connection between machine phenomenology and the dynamic coupling between the body and environment. She provides a through criticism of the conscious innessentialism, which claims that consciousness will not yield to better performance. Quoting extensively Damasio's theory of the somatic marker, Stuart argues that a true autonomous agent requires a self-directed active, dynamically-coupled agent with a capacity for affective bodily activity that makes possible the development of appropriate responses and adaptive behaviours, namely consciousness.

In the same line of thinking, **Domenico Parisi** and **Marco Mirolli** propose to analyze consciousness-related phenomena in artificial systems in order to decide if the system may be conscious or not. One phenomenon may be the knowledge of system's own body as something which is different from other physical objects. Their goal is to operationalize consciousness providing behavioural and computational definitions. Building artifacts which, according to these definitions, can be reasonably said to be conscious, the authors claim that the debate on consciousness can be framed on less metaphysical and more scientific grounds than it is typically the case in traditional philosophical debates. The question they are going to ask is "What are the conditions under which a machine can be reasonably claimed to be 'conscious'?"

A second group of contributors focus on specific implemented systems modelling consciousness.

Stan Franklin and coworkers present LIDA, a working model of machine consciousness. The architecture of LIDA is an implementation of the Global Workspace Theory of consciousness. The LIDA architecture include perceptual associative memory, episodic memory, functional consciousness, procedural memory and action-selection. The LIDA architecture is also compared with other models of consciousness. Franklin and coauthors discuss the architecture as a model that implements access consciousness but they leave out the problem of phenomenal consciousness.

Another implementation of the Global Workspace theory is proposed by **Dustin Connor** and **Murray Shanahan**. They support Global Workspace Theory since is of especial interest to AI researchers because it posits an essential link between consciousness and cognition. While LIDA is essentially an agent system, the implementation proposed by Connor and Shanahan is biologically oriented and it is based on a network of spiking neurons. In particular, they analyze the capabilities of a reverberating population of spiking neurons capable of competing together and of broadcasting the results in a way compatible with Baars' Global Workspace. With respects to previous work, the authors increased the similarity with biological neural networks. Although the present work is mostly a software model, the authors plan to use it in a robotic embedded system in the close future.

Lee McCauley goes hands down and describes a hybrid connectionist-symbolic system that implements a mechanism for consciousness inasmuch consciousness does not refer to any phenomenal aspect. He claims that a global broadcasting system will boost the learning rate of an autonomous system. A description is given relating both how this effect occurs and why the effect is produced. Yet he admits that the term "conscious" is used only to denote that the mechanism implements a portion of a psychological theory of consciousness.

Irene Macaluso and **Antonio Chella** discuss a model of robot perceptual awareness based on a comparison process between the effective and the expected robot input sensory data generated by a 3D robot/environment simulator. The paper contributes to the machine consciousness research field by testing the added value of robot perceptual awareness on an effective robot architecture implemented on an operating autonomous robot offering guided tours in a real archaeological museum. The robot perceptual awareness is based on a stage in which two flows of information, the internal and the external, compete for a consistent match. There is a strong analogy with the phenomenology in human perception. When a human subject perceives the objects of a scene he actually experiences only the surfaces that are in front of him, but at the same time he builds an interpretation of the objects in their whole shape. The authors maintain that the proposed system is a good starting point to investigate robot phenomenology. As described in the paper it should be remarked that a robot equipped with perceptual awareness performs complex tasks as museum tours, because of its inner stable perception of itself and of its environment.

Ricardo Sanz, Ignacio Lopez and **Carlos Hernandez** describes the approach taken in the Autonomous Systems Laboratory of the Universidad Politecnica de Madrid for the development of technology of full bounded autonomy. Their approach is based on the analysis of phenomenal consciousness in relation with the construction of mechanisms for system self-awareness. In particular, they describe the SOUL project for a generic architecture for self-aware autonomous systems.

The study of consciousness could be helpful to improve man-machine interactions. This is the approach addressed by **Daniel Dubois** and **Pierre Poirier**. Stressing the fact that current artificial agents cannot cope with the complexity of real world they aim at designing conscious artificial tutors (CTS) for training of humans. In order to design a psychological implementation of a functional model of Ned Block's access-consciousness, they developed a complete model of most consciousness related cognitive functions.

Catherine Macarelli and **Jeffrey L. McKinstry** exploit the promising relationship between simulation and consciousness. They point out that consciousness can be detected on the basis of three approaches: verbal report, neurological measurements and behaviour. The first twos are unsuited to artificial agents. How is it possible to detect consciousness by observing behaviour? The authors' answer is that insight learning is the key. According to them, conscious thought is related with the ability to mentally simulate the world in order to optimize behaviour. In order to test their hypothesis, the authors developed a computer simulation of an autonomous agent in which the agent had to learn to explore its world and learn to achieve a certain goal. Afterward, the agent should address a new situation by means of a conscious simulation of it (insight learning). The authors believe that this approach could be fruitfully applied for testing consciousness in machines and animals.

Alexei V. Samsonovich stresses the link between learning and consciousness. He suggests to model a universal learner as a cognitive agent that can learn arbitrary schemas as well as associated values, experiences and linguistic primitives, with the help of a human instructor. He maintain that this phenomenon can be regarded as an emergent computational consciousness.

Self-recognition is the topic addressed by **Pentti Haikonen**. Is the mirror test useful to check self consciousness in an agent? In the past, this ability has been taken as a demonstration of self-consciousness. Yet, the author claims that very simple machinery is able to pass the mirror test and, consequently, he argues that the mirror test by itself cannot show the presence of self-consciousness. It could be easy to construct robots that have a somatosensorily grounded non-conscious self-reference with mirror self-recognition. On the other hand, it is possible that raising the criteria bar for self recognition will lead to an improved version of the mirror test.

What is the inner world of an agent? **Germund Hesslow** and **Dan-Anders Jirenhed** wonder whether internal simulation could be used to create agents with an inner world. Resting on their previous work, the authors suggest a strong dependence between the capability of simulating the outer world and consciousness. But they also make a bolder claim – namely that a robot in which perception can be simulated has an inner world and subjective experience in the same sense as humans. Two still open issues are: i) do such simulation mechanisms exist in human beings? ii)

are the simulation patterns similar or different from those of normal perception? A further and even more worrisome issue is whether simulation is responsible *per se* for conscious perception or whether it only triggers consciousness. The authors admit that most readers could be rather uncomfortable with the idea that a simulating agent has an inner world. Yet, they boldly conclude that they "have not yet encountered any convincing reason to deny that [a robot with a inner simulation] has the basic mechanism of an inner world and that in this crucial respect, it is closer to a human".

In a position paper, **Christophe Menant** proposes to shift the attention in the study of self consciousness in relation to the study of the evolution of representations and to take into account the results for designing artificial conscious artefacts.

In another position paper, **Rafal Rzepka and Kenji Araki** briefly present the GENTA project aiming at adopting Internet as a medium to extract and imitate the "typical" interacting behaviours of conscious persons.

Finally, the symposium is enriched by two keynotes speakers (Giulio Tononi and Aaron Sloman) that comment on the general perspectives of consciousness and AI.

Giulio Tononi introduces an interesting line of research still in infancy concerning the study of consciousness from a theoretical point of view. The main item of this research is: what are the characteristics for a generic system, both natural and artificial, so that the system itself may be considered conscious? Tononi (Tononi 2006; Tononi and Edelman 1998) proposes a theoretical measure of consciousness based on the capability of a system to integrate information. He observes that two key features of a conscious agent are differentiation and integration. According to him, the quality of phenomenal experience is determined by the informational relationships among the elements of a complex. Hence he presents several neurobiological observations concerning consciousness including the association of consciousness with certain neural systems; the fact that neural processes underlying consciousness can influence or be influenced by neural processes that remain unconscious. He believes that consciousness is a fundamental quantity and that it should be possible to build conscious artifacts.

Aaron Sloman discusses about the needs of a general theory for designing and implementing conscious machines that should incorporate some form of phenomenal consciousness. In this way, it should be possible to analyze the trade off for a machine of having or not having "qualia". Sloman suggests that phenomenal experience could depend on the existence in the machine of a 'meta-management' subsystem, which has self-monitoring and 'meta-semantic' capabilities and is capable of developing an ontology to describe some of its own internal states and processes. Sloman hopes that the design and implementation of such machines, and analyses of their tradeoffs, could help to unify philosophy, psychology, psychiatry, neuroscience, studies of animal cognition, and of course AI and robotics.

Conclusions

Only a few years ago, Jerry Fodor could wrote that (Fodor, 1992, p. 5) "Nobody has the slightest idea how anything material could be conscious. Nobody even knows what it would be like to have the slightest idea about how anything material could be conscious." To be honest, the theoretical landscape has not changed much (Koch, 2004). Although there surely survive many ingenuities and many untested assumption, the alliance between AI and consciousness could prove to be a necessary and fruitful step to the understanding of the nature of consciousness, subjectivity and autonomy in agents. Up to now, there is now widespread consensus of what consciousness is and how it can be applied to artificial agents. Yet, it is paramount to check all possibilities since it is difficult to underestimate the advantage of an artificial agent equipped with a human-like consciousness whatever this entails.

Acknowledgements

References

Adami, C. (2006). "What Do Robots Dreams Of?" *Science* 314 (5802): 1093-1094.

Aleksander, I. (2000). *How to Build a Mind*. London, Weidenfeld & Nicolson.

Aleksander, I. (2001). "The Self 'out there'." *Nature* 413: 23.

Aleksander, I. and H. Morton (2005). "Enacted Theories of Visual Awareness, A Neuromodelling Analysis". in BVAI 2005, LNCS 3704.

Aleksander, I. and H. Morton (2007). "Axiomatic consciousness theory for visual phenomenology in artificial intelligence." This volume.

Atkinson, A. P., M. S. C. Thomas, et al. (2000). "Consciousness: mapping the theoretical landscape." *Trends in Cognitive Sciences* 4 (10): 372-382.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge, Cambridge University Press.

Baars, B. J. (2002). "The Conscious Access Hypothesis: origins and recent evidence." *Trends in Cognitive Sciences* 6 (1): 47-52.

Boltuc, N. and P. Boltuc (2007). "Replication of the hard problem of consciousness in AI and Bio-AI: an early conceptual framework." This volume.

Bongard, J., v. Zykov, et al. (2006). "Resilient Machines Through Continuous Self-Modeling." *Science* 314 (5802): 1118-1121.

Chella, A., S. Gaglio, et al. (2001). "Conceptual representations of actions for autonomous robots." *Robotics and Autonomous Systems* 34 (4): 251-264.

Chella, A. and S. Gaglio (2007). "A cognitive approach to robot self-consciousness." This volume.

Chella, A. and R. Manzotti (2007). *Artificial Consciousness*. Exeter (UK), Imprint Academic.

Chrisley, R. (1995). "Non-conceptual Content and Robotics: Taking Embodiment Seriously". In: *Android Epistemology*. F. K, G. C and H. P., Cambridge, AAAI/MIT Press**:** 141-166.

Chrisley, R. (2003). "Embodied artificial intelligence." *Artificial Intelligence* 149: 131-150.

Chrisley, R. and J. Parthemore (2007). "Robotic specification of the non-conceptual content of visual experience." This volume.

Connor, D. and M. Shanahan (2007). "A simulated global neuronal workspace with stochastic wiring." This volume.

Cotterill, R. M. J. (1995). "On the unity of conscious experience." *Journal of Consciousness Studies* 2: 290-311.

Dennett, D. C. (1991). *Consciousness explained*. Boston, Little Brown and Co.

D'Mello, S. and S. Franklin (2007). "Exploring the complex interplay between AI and consciousness." This volume.

Drestke, F. (2000). *Perception, Knowledge and Belief*. Cambridge, Cambridge University Press.

Dubois, D. and P. Poirier (2007). "What does consciousness bring to CTS?" This volume.

Edelman, G. M. and G. Tononi (2000). *A Universe of Consciousness. How Matter Becomes Imagination*. London, Allen Lane.

Franklin, S., U. Ramamurthy, S.K. D'Mello, L. McCauley, A. Negatu, R. Silva L., V. Datla (2007). "LIDA: a computational model of global workspace theory and developmental learning." This volume.

Fodor, J. A. (1992). "The big idea: Can there be a science of mind?" *Times Literary Supplement*: 5-7.

Franklin, S. (1995). *Artificial Minds*. Cambridge (Mass), MIT Press.

Franklin, S. (2003). "IDA: A Conscious Artefact?" in *Machine Consciousness*. O. Holland. Exeter (UK), Imprint Academic.

Haikonen, P. O. (2003). *The Cognitive Approach to Conscious Machine*. London, Imprint Academic.

Haikonen, P. (2007). "Reflections of consciousness; the mirror test." This volume.

Harnad, S. (1990). "The Symbol Grounding Problem." *Physica* D (42): 335-346.

Harnad, S. (1995). "Grounding symbolic capacity in robotic capacity". in *"Artificial Route" to "Artificial Intelligence": Building Situated Embodied Agents*. L. Steels and R. A. Brooks. New York, Erlbaum.

Harnad, S. and S. Scherzer (2007). " First, scale up to the robotic Turing test, then worry about feeling." This volume.

Haugeland, J. (1985/1997). "Artificial Intelligence: The very Idea". in *Mind Design II*. Cambridge (Mass), MIT Press.

Hesslow, G. (2002). "Conscious thought as simulation of behaviour and perception." *Trends in Cognitive Sciences* 6 (6): 242-247.

Hesslow, G. (2003). Can the simulation theory explain the inner world? Lund (Sweden), Department of Physiological Sciences.

Hesslow, G. and D.-A. Jirenhed (2007). " Must machines be zombies? Internal simulation as a mechanism for machine consciousness." This volume.

Holland, O., (2003). *Machine consciousness*. New York, Imprint Academic.

Jennings, C. (2000). "In Search of Consciousness." *Nature Neuroscience* 3 (8): 1.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood (Col), Roberts & Company Publishers.

Kuipers, B. (2005). "Consciousness: drinking from the firehose of experience". in National Conference on Artificial Intelligence (AAAI-05).

Kuipers, B. (2007). "Sneaking up on the hard problem of consciousness." This volume.

Macaluso, I. and A. Chella (2007). "Machine Consciousness in CiceRobot, a Museum Guide Robot." This volume.

Manzotti, R. (2006). "An alternative process view of conscious perception." *Journal of Consciousness Studies* 13 (6): 45-79.

Manzotti, R. (2007). "From Artificial Intelligence to Artificial Consciousness". in *Artificial Consciousness*. A. Chella and R. Manzotti. London, Imprint Academic.

Manzotti, R. and Tagliasco, V. (2007). " Towards artificial consciousness." This volume.

Marcarelli, C. and J.L. McKinstry (2007). "Testing for machine consciousness using insight learning." This volume.

McCarthy, J. (1995). "Making Robot Conscious of their Mental States". in *Machine Intelligence*. S. Muggleton. Oxford, Oxford University Press.

McCauley, L. (2007). "Exploring the complex interplay between AI and consciousness." This volume.

McDermott, D. (2001). *Mind and Mechanism*. Cambridge (Mass), MIT Press.

Menant, C. (2007). "Proposal for an approach to artificial consciousness based on self-consciousness." This volume.

Minsky, M. (1985). *The Society of Mind*. New York, Simon & Schuster.

Minsky, M. (1991). "Conscious Machines". in Machinery of Consciousness, National Research Council of Canada.

Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, Simon & Schuster.

Nöe, A. (2004). *Action in Perception*. Cambridge (Mass), MIT Press.

O' Regan, K. and A. Noe (2001). "A sensorimotor account of visual perception and consciousness." Behavioral *and Brain Sciences* 24 (5).

Parisi, D. and M. Mirolli (2007). "Steps towards artificial consciousness: a robot's knowledge of its own body." This volume.

Pirri, F. (2007). " The Prince of Holmberg I." This volume.

Revonsuo, A. (1995). "Consciousness, dreams, and virtual realities." *Philosophical Psychology* 8: 35-58.

Rockwell, T. (2005). *Neither ghost nor brain*. Cambridge (Mass), MIT Press.

Rzepka, R. and K. Araki (2007). "Consciousness of crowds – The Internet as a knowledge source of human's conscious behavior and machine self-understanding." This volume.

Samsonovich, A.V. (2007). "Universal Learner As an Embryo of Computational Consciousness." This volume.

Sanz, R. (2005). "Design and Implementation of an Artificial Conscious Machine". in IWAC2005, Agrigento.

Sanz, R., I. Lopez and C. Hernandez (2007). "Self-awareness in real-time cognitive control architectures." This volume.

Shanahan, M. P. (2005). "Global Access, Embodiment, and the Conscious Subject." *Journal of Consciousness Studies* 12 (12): 46-66.

Sloman, A. (2007). " Why some machines may need qualia and how they can have them: including a demanding new Turing test for robot philosophers." This volume.

Sloman, A. and R. Chrisley (2003). "Virtual Machines and Consciousness." *Journal of Consciousness Studies* 10 (4-5).

Stein, B. E. and M. A. Meredith (1999). *The merging of the senses*. Cambridge (Mass), MIT Press.

Stuart, S. (2007). "Conscious machines: memory, melody and imagination." This volume.

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

Tononi, G. (2004). "An information integration theory of consciousness." *BMC Neuroscience* 5:42.

Tononi, G. (2007). "Consciousness as integrated information." This volume.

Tononi, G. and G.M. Edelman (1998). "Consciousness and Complexity." Science 282, 1846-1851.

Ziemke, T. (2001). "The Construction of 'Reality' in the Robot: Constructivist Perpectives on Situated Artificial Intelligence and Adaptive Robotics." *Foundations of Science* 6 (1-3): 163-233.