ConsScale: A Plausible Test for Machine Consciousness?

Raúl Arrabales Moreno, Agapito Ledezma Espino and Araceli Sanchis de Miguel

Universidad Carlos III de Madrid. Departamento de Informática.
{rarrabal,ledezma,masm}@inf.uc3m.es

Abstract

Is consciousness a binary on/off property? Or is it on the contrary a complex phenomenon that can be present in different states, qualities, and degrees? We support the latter and propose a linear incremental scale for consciousness applicable to artificial agents. *ConsScale* is a novel agent taxonomy intended to classify agents according to their level of consciousness. Even though testing for consciousness remains an open question in the domain of biological organisms, a review of current biological approaches is discussed as well as their possible adapted application into the realm of artificial agents.

Regarding to the always controversial problem of phenomenology, in this work we have adopted a purely functional approach, in which we have defined a set of architectural and behavioral criteria for each level of consciousness. Thanks to this functional definition of the levels, we aim to specify a set of tests that can be used to unambiguously determine the higher level of consciousness present in the artificial agent under study. Additionally, since a number of objections can be presumably posed against our proposal, we have considered the most obvious critiques and tried to offer reasonable rebuttals to them. Having neglected the phenomenological dimension of consciousness, our proposal might be considered reductionist and incomplete. However, we believe our account provides a valuable tool for assessing the level of consciousness of an agent at least from a cognitive point of view.

1. Introduction

Testing for consciousness remains an open problem even when it is aimed at humans or other mammals. The scientific study of consciousness differs from any other scientific scrutiny due to the fact that consciousness is private and subjective while traditional scientific methods are based on public and objective observations. However, most contemporary neuroscientists agree that consciousness can be tackled scientifically using alternative strategies. Additionally, we believe that establishing a framework for classifying, measuring, and testing the level of consciousness of a subject is of central importance in the scientific quest for consciousness. Provided that first-person analysis alone cannot offer a convincing scientific value, alternative third-person standpoints can be adopted in terms of behavior and neurobiological structures analysis. Although these alternative third-person approaches cannot offer a direct inspection of consciousness, we believe they are of useful scientific value. In fact, analogous indirect observations have been extensively applied in other scientific domains like atomic physics. Basically, the hypothesis that we support in this work is that the level of consciousness of a biological organism could be accurately assessed by correctly analyzing two aspects: the presence of Neural Correlates of Consciousness (NCC) and the presence of characteristic behaviors. When it comes to artificial agents, we adopt the same methods but they cannot be directly applied due to the following reasons: (1) Artificial agents have different underlying machinery, instead of NCC we should refer to functionally analogous Computational Correlates of Consciousness (CCC) (Atkinson, Thomas & Cleeremans 2000). (2) Artificial agents produce different behavioral patterns. As embodiment in artificial agents differs significantly from humans or any other biological organism, the generated behavior, which is deeply influenced by the body physical characterization, also differs.

While some authors advocate for the necessity of a specific physical substrate for consciousness, see for instance Hameroff and Penrose (1996), others support the idea that consciousness is produced by specific processes that could be reproduced in different substrates, e.g. Edelman and Tononi (2000), Crick and Koch (1990), and Grossberg (2003) amongst others. The latter approaches allow the possibility of Machine Consciousness and this is indeed the hypothesis that we support. Although the precise way in which the brain produces consciousness is not clear yet, we argue that specific aspects of consciousness can be analyzed and

sorted adopting an evolutionary perspective, from phylogenetically older functions to the most modern features observed in human adults.

Taking into account the former considerations, a scale designed to test the level of consciousness of artificial agents called *ConsScale* has been recently proposed (Arrabales, Ledezma & Sanchis 2008). *ConsScale* is a taxonomy, actually defined as an ordered list, intended to classify artificial agents according to their level of consciousness. *ConsScale* is also intended to serve as a reference framework for analyzing the existing correlation between consciousness and cognitive skills; furthermore, it is expected to be a realistic reference to determine the current state of the art in the field of Machine Consciousness.

The following sections introduce the context and objectives of *ConsScale*. Firstly, current methods for assessing the level of consciousness are discussed in section 2. Our computational approach to consciousness in artificial agents is explained in section 3. *ConsScale* levels are briefly described in section 4. A short discussion on the theoretical foundations of the scale and possible objections is presented in section 5. Finally, section 6 includes a consclusion and future work.

2. Current methods for assessing the level of consciousness

Determining the level of consciousness of a living organism is a hard problem. The actual definition of levels is indeed an open question and clinical diagnosis methods for humans like the Glasgow Coma Scale (GCS) or Simplified Motor Score (SMS) do not cover the broad range of consciousness functional components (Gill et al. 2007; Van de Voorde et al. 2008). There also exist psychological scales which are focused in specific aspects of consciousness like the Private Self-Consciousness Scale (Fenigstein, Scheier & Buss 1975). However, neither neurological nor psychological tests of this sort can be directly applied or plausibly adapted to artificial agents.

From a cognitive point of view, one could think that some sort of Turing test might be a plausible solution (Turing 1950). However, such a test cannot provide a measurement of the level of consciousness for subjects not reaching the *human-like* level. Even for that level, mere observation of external behavior is not enough (Haikonen 2007b).

From the neuroscience perspective, Seth, Baars, and Edelman (2005) propose a set of criteria for consciousness in humans and other mammals. A number of these criteria are based on neurobiological aspects. If the NCC and the associated activity patterns that give place to consciousness are identified, then we can look for them in animals endowed with a central nervous system. Additionally, if some behavioral patterns are identified as uniquely produced by conscious subjects, we can design experiments where these behaviors are tested.

Other proposals exist to test the presence of consciousness in artificial agents. However, they are not designed as gradual scales but unitary tests. The most remarkable work in this area is the set of axioms proposed by Aleksander and Dunmall (2003) and Aleksander and Morton (2007). According to the authors, a minimal set of axioms (depiction, imagination, attention, planning, and emotion) are required in order to consider an agent conscious. Additionally, Haikonen has pointed out that the ability to report mental content (to itself and to others) is also a requirement for consciousness. Inner speech with *grounded meaning*s is one manifestation of mental content and the system's ability to report this would thus be an indicator of the presence of consciousness (Haikonen 2007b). As explained below, these criteria are also considered in *ConsScale*.

3. A computational approach to consciousness in artificial agents

In order to characterize consciousness as a property of agents we need to formally define the basic components of an artificial situated agent, i.e. the conceptual building blocks integrated in a possible artificial consciousness implementation. Then we could test the presence of these functional components, their interrelation, and the corresponding behavioral outcome in order to assess the level of machine consciousness.

An agent interacts with the environment by retrieving information both from its own body and from its surroundings, processing it, and acting accordingly. Following Wooldridge's definition of abstract architectures for intelligent agents (Wooldridge 1999), and taking into account the embodiment aspect of situated agents, we have identified a set of essential architectural components: sensors, sensorimotor coordination, internal state (including memory), and effectors. These subsystems implement the basis of the

following processes: perception, reason, and action. Cognition and learning can develop in an agent on top of the former processes during the interaction with the external world and their own inner state. Consequently, the following abstract architectural components can be identified:

- Body (***B***). Embodiment is a key feature of a situated agent. Agent's body can be physical or software simulated (as well as its environment). A boundary is established between agent's body and its environment (***E***). The rest of components are usually located within this boundary.
- Sensory Machinery (**S**). Agent's sensors are in charge of retrieving information from the environment (exteroceptive sensors) or from the agent's own body (propioceptive sensors).
- Action Machinery (***A***). In order to interact with the environment the agent uses its effectors. Agent's behavior is composed of the actions ultimately performed by this machinery.
- Sensorimotor Coordination Machinery (***R***). From purely reactive agents to deliberative ones, the sensorimotor coordination system is in charge of producing a concrete behavior as a function of both external stimuli and internal agent's state.
- Memory (***M***). Internal agent's state is represented both by its own structure and stored information. Memory is the mean to store both perceived information and new generated knowledge. We consider that even agents that do not maintain state can be said to have a minimal state represented by its own structure, i.e. preprogrammed sensorimotor coordination rules.

The former components refer to an abstract architecture; therefore, we are not considering here any particular agent implementation or concrete sensorimotor machinery. Using the presented abstract architecture allows us to define consciousness levels independently of particular implementations. As Wooldridge (1999) has pointed out, different classes of agents could be obtained depending on the concrete implementation of the abstract architecture. It is also important to note that no specific component of this architecture is responsible for the production of consciousness. Instead, we support that consciousness could emerge from the interaction of the specialized processes present in the agent.

In computational terms, consciousness can be regarded as a unique sequential thread that integrates concurrent multimodal sensory information and coordinates voluntary action. Hence, consciousness is closely related with sensorimotor coordination. Our aim is to establish a classification of agents according to the realization of the functions of consciousness in the framework of agent's sensorimotor coordination.

Out of the set of cognitive functions that an intelligent agent could potentially exhibit, the following group of functions specifically characterizes the behavior of a conscious agent: Theory of Mind (ToM), Executive Function (EF), and modulating function of emotions. ToM is the ability to attribute mental states to oneself and others (Vygotsky 1980). From a human developmental standpoint, Lewis (2003) suggests four stages in the acquisition of ToM: (1) "I know", (2) "I know I know", (3) "I know you know", and finally (4) "I know you know I know". Undoubtedly, ToM is required in order to implement any sort of social learning. The term EF includes all the processes responsible for higher level action control, in particular those that are necessary for maintaining a mentally specified goal and for implementing that goal in the face of distracting alternatives (Perner, Lang 1999). Attention is an essential feature of EF. It represents the ability of the agent to direct its perception and action, i.e. selecting the contents of the working memory out of the entire mind's accessible content. Planning, coordination, and set shifting (the ability to move back and forth between tasks) are also key processes included in EF. Emotions play a key role in the generation and modulation of behavior even in organisms that lack self-consciousness. In organisms with higher levels of consciousness, like humans, the feeling of emotions enables the interaction and competition between emotional and rational responses. Effective learning in complex and unstructured environments requires these cognitive skills. We argue that the effective integration of all of these cognitive functions could build an artificial conscious mind. However, each of the mentioned functions could also be implemented independently or partly integrated with other cognitive functions, thus giving place to different levels of implementation of artificial consciousness as discussed in the next section.

4. *ConsScale*

The fact that consciousness is an incremental property can be easily observed in both human ontogeny and other mammals' phylogeny. Most complex learning capabilities and higher cognitive skills are generally observed in those species with higher levels of consciousness. The level or degree of consciousness that a biological creature is endowed with seems to increment correlated with a concrete path of evolution, showing the highest levels in quadruped mammals and great apes, and reaching its maximum expressions in

the case of Homo sapiens sapiens. The definition of concrete and functionally discrete levels of consciousness permits us to have a pragmatic reference framework for classifying artificial agents. Table 1 describes *ConsScale*, which is a sorted list of potential levels of consciousness for artificial agents. This scale has been defined in terms of reference abstract architectures, characteristic behaviors, and cognitive skills. As illustrative analogy, machine consciousness levels are assigned a comparable level of consciousness in biological phylogeny and human ontogeny.

The first level in the scale, level -1 or *Disembodied*, refers to a 'proto-agent' and serves as an initial reference that remarks the importance of a body as a requirement for defining a situated agent. The rest of the scale comprises a set of twelve ranks, where lower levels are subsumed by higher ones. Therefore, each stage of the incremental development of an artificial agent could be identified by a concrete level.

Level 0, *Isolated*, is also a conceptual reference which helps characterizing situatedness in terms of the relation with the environment. It represents an inert body lacking any functionality or interaction with the medium except the inevitable derived from the physical properties of its proper inactive body.

Level 1, *Decontrolled*, represents an agent which is endowed with some sensorimotor machinery, which for some reason is not functional at all (or it lacks a functional relationship between $\boldsymbol{S}$ and $\boldsymbol{A}$). Therefore, this is also a conceptual reference level. Systems which belong to either level 0 or level 1 cannot be defined as situated agents.

Level 2, *Reactive*, defines a classical reactive agent which lacks any explicit memory or learning capabilities. From level 2 onwards the agents make use of the environment as the mean to close the feedback loop between action and perception. Hence, all agent types above level 1 can be regarded as situated agents. The characteristic behavior of this level is the reflex, hence an agent able to autonomously react to any given environment situation is said to comply with level 2.

Level 3, *Adaptive*, can be identified as the simplest form of an adaptive agent. At this level, the agent's internal state is maintained by a memory system and sensorimotor coordination ($\boldsymbol{R}$) is just a simple function of both perceived and remembered information. Propioceptive sensing can be present at this level; however, it is not producing any self-awareness. At this level, learning mechanisms are possible as new reflective behaviors can be acquired. When the response to a given environment state is not fixed, but it is a function of both the information acquired by $\boldsymbol{S}$ and agent's internal state ($\boldsymbol{M}$), then the agent is said to comply with level 3 (note that some propioceptive sensing mechanism is required to make agent's internal state available in $\boldsymbol{R}$, so it can be an input of the sensorimotor coordination function). Level 3 can also be seen as an evolution of level 2 in which a capability for learning new reflexes has been acquired.

Level 4, *Attentional*, is characterized by an attention mechanism, which allow the agent to select specific contents both from the sensed and stored state information. This means that explicit learning is directed toward selected objects or events. However, implicit learning mechanisms also exist, like the acquisition of reflective strategies which is also a characteristic of the former level. If the agent is able to direct attention to a selected subset of the environment state ($\boldsymbol{E}_i$) while other environmental variables are also sensed but ignored in the explicit processing of $\boldsymbol{R}$, and the selected perception is automatically evaluated in terms of agent's goals allowing subsequent responses to be adapted (emotions), then the agent is said to comply with level 4. Attentional agents are able to show specific attack or escape behaviors and trial and error learning. The ability to pay attention toward specific objects or events gives place to the formation of directed behavior, i.e. agent can develop behaviors clearly related to specific targets, like following or running away. Additionally, level 4 agents can have primitive emotion mechanisms in the sense that the objects to which attention is paid are elementally evaluated as positive or negative. A positive emotion triggers decrease of distance behavior or bonding to selected object, while negative emotion triggers increase of distance and reinforcement of boundaries toward selected object (Ciompi 2003). Moreover, a new relation between emotions and memory appears at this level: as demonstrated in biological organisms, emotions are deeply involved in the selection of what needs to be stored in memory (LaBar, Cabeza 2006). Basically, level 4 can be seen as an evolution of level 3 in which the attention capability has been acquired.

A level 5 agent, *Executive*, includes a more complex reasoning and internal state representation, which provides set shifting capabilities. The achievement of multiple goals is performed thanks to a higher coordination mechanism that shifts attention from one task to another. The agent is also endowed with a mechanism to evaluate the performance in achieving the pending goals. This mechanism is the self-status assessment provided by emotions (which was already present in the former level). The presence of emotions associated to objects and events jointly with the set shifting capability permits the development of reinforcement learning mechanisms. In addition to advanced planning, emotional learning is another characteristic that can be observed at this level, as the most emotionally rewarding tasks are assigned more

time and effort. In sum, level 5 can also be seen as an evolution of level 4 in which goal seeking and set shifting capabilities have been acquired.

Level 6, *Emotional*, is the first level in which an agent can be to certain extend regarded as conscious (but not self-conscious). The main characteristic of this level is the support for ToM stage 1, "I know". Feelings appear as representations of organism changes due to an emotion (Damasio 1999). As the effects of emotions in the organism are mapped, a sense of "I know" appears in the agent.

Level 7, *Self-Conscious*, corresponds to the emergence of self-consciousness. At this level the agent is able to develop higher order thoughts (Rosenthal 2000), specifically thoughts about itself. Consequently it presents support for ToM stage 2, "I know I know". This requires the presence of a model of self in the agent, which in turns permits advance planning as the proper agent is part of the plan. Therefore, learning mechanisms can operate now in the realm of anticipated future. The agent can plan about itself, and later learn if the plan was efficient or not. Recognizing the self as a character in the plan seems to be a key factor for learning to use tools (Sasaki et al. 2008). The reference behavior test for this level would be the mirror test, which although originally applied to primates (Gallup 1977), has also been adapted to other mammals and even artificial agents. Takeno et al. have proposed a specific experiment design to test whether a robot is able to recognize its own image reflected in a mirror (Takeno, Inaba & Suzuki 2005).

In level 8, *Empathic*, the internal representation of the agent is enriched by inter-subjectivity. In addition to the model of the self, others are also seen as selves; hence, they are consequently assigned a model of subjectivity. This is the seed for a complex social interaction. This capability in addition to the ability to hold a precise and updated map of body schema, i.e. body shape and posture (which is acquired in level 6), is necessary for the learning of tool usage and for the making of new tools (Maravita, Iriki 2004; Stout, Chaminade 2007). Being aware of others permits the conscious collaboration with other agents in the pursuit of common goals.

The next step is represented by level 9, *Social*, where ToM is fully supported. In this case, agents are strongly influenced by the social environment and a culture can be potentially developed. Characteristic behavior of this level is defined by sophisticated Machiavellian strategies (or social intelligence) involving social behaviors like lying, cunning, and leadership. In other words, an agent A could be aware that another agent B could be aware of A's beliefs, intentions, and desires. Advanced communication skills are the characterization of this level behavior, where, for the first time, an agent would be able to purposely tell lies. There exist mathematical models of the dynamics of Machiavellian intelligence that could be potentially used to test these sorts of behaviors with artificial agents (Gavrilets, Vose 2006).

Level 10, *Human-Like*, represents the sort of agents endowed with the same level of consciousness as a healthy adult human. Therefore, the formation of a complex culture is a feature of this level. This implies the usage of external complex tools for learning. The abstract architecture for both level 9 and 10 is the same. The real difference in level 10 comes from the fact that culture affects the mind, i.e. the way the brain is used. Accurate communications skills (language) and the creation of a culture would be a clear feature of this level. Other key characteristics are that level 10 agents are able to profoundly modify their environment and society. The fluidity between social and technical intelligence permits the extension of their own knowledge using external media (like written communication) and technological advances are also possible.

Finally, level 11 or *Super-Conscious*, refers to a kind of agent able to internally manage several streams of consciousness, while coordinating a single body and physical attention. A mechanism for coordination between the streams and synchronized access to physical resources would be required at this level. We cannot envisage any conclusive behavior test for level 11 due to the lack of known exemplifying references.

5. Theoretical foundations and application of *ConsScale*

The main point to take into account about *ConsScale* is that it is based on a functionalist approach and it is bio-inspired. The levels of artificial consciousness defined in *ConsScale* are characterized by abstract architectural components and agent's behavior. It must be noted that ***R*** and ***M*** do not represent centralized modules. They represent the substrate that supports functions that can be carried out by any distributed machinery. In other words, the proposed scale is based on specialized processes rather than specialized machinery. The architecture components represent functional subsystems whose integration makes possible the emergence of a characteristic behavior. Therefore, at least one behavior-based test might be associated to each level in order to assess if a particular agent fulfills the minimum required behavioral pattern for that level. In fact, an agent can only be assigned a concrete level if and only if it is able to show the behavioral

pattern of that level as well as the behavioral patterns of all lower levels. In other words, higher levels subsume all lower ones (except the three first levels). As discussed above, the three first reference levels (*Disembodied*, *Isolated*, and *Decontrolled*) are a special case as they do not actually describe situated agents. Therefore, it does not make sense to have behavioral tests associated to any of these first three levels. A given agent could be assigned either of these initial reference levels just by analyzing its architectural components. In contrast, from level 2 onwards a characteristic behavior pattern is defined per *ConsScale* level. This characteristic pattern should be taken as the base of any behavior test that can be assigned to a particular level.

An obvious critique to the proposed scale can be based on the fact that it merely follows a particular discretized linear path within the virtually infinite map of possible artificial agent implementations. Why have we chosen this particular arrangement of conscious levels? Why not some functionality could be in different levels? The short answer is that this is a bio-inspired scale and we have tried to specifically select and identify the most significant levels of phylogenetic and ontogenetic development that have led to human-like consciousness. The main reason is that evolutionary development of humans is the most advanced example we know of building a conscious agent. For instance, as observed in biological organisms, generally doing precedes understanding. As indicated by Haikonen (2007a), explorative actions seem to be a requirement to be able to learn to make sense of perceptions. Therefore, *ConsScale* considers capability for explorative actions in the lower levels of the scale. This ability is combined with cognitive skills like imagination at higher levels of the scale. Also, from the point of view of emotions, embodiment and action have to be present before any cognitive process takes place (James 1884), e.g. we learn to speak before learning syntax or grammar. Bodily representations of emotion follow perception, and as Damasio (1999) has pointed out, the feeling of the emotion is what finally causes the conscious state associated to the emotion.

The following features are added incrementally in *ConsScale*: perception and action capabilities, basic situatedness, basic emotions, basic adaptiveness, attention, explorative actions and set shifting, feelings, planning, imagination, Theory of Mind, language, and culture. These functions cannot be considered as unitary features, instead an integrated grand function of consciousness should be considered as being present to the degree specified at any given level. For instance, having the ability to develop imaginations but not being able to adapt to the environment and perform explorative actions is useless in terms of building a conscious agent. This is the reason why we have designed a linear scale, not taking into consideration any possible 'hybrid' levels in which the required sensorimotor machinery and behavior are not in the line that lead to human-like consciousness in biological evolution.

6. Conclusions and future work

We have proposed *ConsScale* as a machine consciousness taxonomy for artificial agents, which can be used as a conceptual framework for evaluating the potential level of consciousness of a given agent. To our best knowledge, most of current implementations of artificial agents fall between levels 2 and 4 inclusive. The classification of any current implementation as fully belonging to level 5 or higher could be thoughtfully discussed elsewhere; nonetheless, we think these kinds of agents are within current technology possibilities.

Identifying consciousness by means of interpreting behavior remains an open problem that is being currently addressed primarily in mammals, cephalopods, and birds (Seth, Baars & Edelman 2005). However, more effort should be put in the domain of artificial agents.

The controversial aspects of consciousness have been addressed in this work with a functional characterization of consciousness based on the notions of embodied intelligence and the integrated grand function of consciousness.

In order to effectively apply the proposed scale to a particular agent, concrete tests have to be defined. Learning capabilities associated to each level are the key to define such behavioral test cases. However, the cognitive foundations of advanced learning capabilities like tool usage and the making of tools remain controversial. The application of the proposed scale could help in the clarification of the role that higher cognitive functions play in advance learning. Therefore, future work regarding *ConsScale* includes designing experiments and reference behavior patterns for levels 2 to 11.

Acknowledgements

We wish to thank Pentti Haikonen for his kindly support at Nokia Research Center, his valuable comments and corrections, and granting us the opportunity to participate in the outstanding Nokia conference. This research has been also supported by the Spanish Ministry of Education and Science under CICYT grant TRA2007-67374-C02-02.

References

Aleksander, I. & Dunmall, B. (2003), Axioms and Tests for the Presence of Minimal Consciousness in Agents, *Journal of Consciousness Studies*, vol. 10, no. 4-5.

Aleksander, I. & Morton, H. (2007), Why Axiomatic Models of Being Conscious? In R. Chrisley, Clowes, R., and Torrance, S. (eds.) *JCS special issue on Machine Consciousness*. vol. 14. no. 7. pp. 15-27.

Arrabales, R., Ledezma, A. & Sanchis, A. (2008), Criteria for Consciousness in Artificial Intelligent Agents, *ALAMAS&ALAg Workshop at AAMAS 2008*, eds. F. Klügl, K. Tuyls & S. Sen, , pp. 57-64.

Arrabales, R., Ledezma, A. & Sanchis, A. (2007), Modeling Consciousness for Autonomous Robot Exploration, in *IWINAC 2007. Lecture Notes in Computer Science*, vol. 4527. pp. 51-60.

Atkinson, A.P., Thomas, M.S.C. & Cleeremans, A. (2000), Consciousness: Mapping the Theoretical Landscape, in *Trends in Cognitive Sciences*. vol. 4, no. 10, pp. 372-382.

Ciompi, L. (2003), Reflections on the role of emotions in consciousness and subjectivity, from the perspective of affect-logic, *Consciousness & Emotion*, vol. 4, no. 2, pp. 181-196.

Crick, F. & Koch, C. (1990), Towards a Neurobiological Theory of Consciousness, *Semin Neurosci*, no. 2, pp. 263–275.

Damasio, A.R. (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Heinemann, London.

Edelman, D.B., Baars, B.J. & Seth, A.K. (2005), Identifying hallmarks of consciousness in non-mammalian species, *Consciousness and Cognition*, vol. 14, no. 1, pp. 169-187.

Edelman, G.M. & Tononi, G. (2000), *A Universe Of Consciousness: How Matter Becomes Imagination*, Basic Books, NY.

Fenigstein, A., Scheier, M.F. & Buss, A.H. (1975), Private and public self-consciousness: Assessment and theory, *Journal of Consulting and Clinical Psychology*, no. 36, pp. 1241-1250.

Gallup, G.G. (1977), Self-recognition in primates: A comparative approach to the bidirectional properties of consciousness, *American Psychologist*, no. 32, pp. 329-337.

Gavrilets, S. & Vose, A. (2006), The dynamics of Machiavellian intelligence, *PNAS*, vol. 103, no. 45, pp. 16823-16828.

Gill, M., Martens, K., Lynch, E.L., Salih, A. & Green, S.M. (2007), Interrater Reliability of 3 Simplified Neurologic Scales Applied to Adults Presenting to the Emergency Department With Altered Levels of Consciousness, *Annals of Emergency Medicine*, vol. 49, no. 4, pp. 403-407.

Grossberg, S. (2003), The Brain's Cognitive Dynamics: The Link between Learning, Attention, Recognition, and Consciousness, *Knowledge-Based Intelligent Information and Engineering Systems; LNCS*, eds. V. Palade, R.J. Howlett & L.C. Jain, Springer, pp. 5.

Haikonen, P.O.A. (2007a), Essential Issues of Conscious Machines, *Journal of Consciousness Studies*, vol. 14, no. 7, pp. 72-84.

Haikonen, P.O.A. (2007b), *Robot Brains. Circuits and Systems for Conscious Machines*. John Wiley & Sons. UK.

Hameroff, S.R. & Penrose, R. (1996), Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness, *Toward a Science of Consciousness*, eds. S. Hameroff, A. Kaszniak & A. Scott , MIT Press.

James, W. (1884), What is an Emotion? *Mind*, no. 9, pp. 188-205.

LaBar, K.S. & Cabeza, R. (2006), Cognitive neuroscience of emotional memory, *Nature reviews. Neuroscience*, vol. 7, no. 1, pp. 54-64.

Lewis, M. (2003), The Emergence of Consciousness and Its Role in Human Development, *Annals of the New York Academy of Sciences*, vol. 1001, no. 1, pp. 104-133.

Maravita, A. & Iriki, A. (2004), Tools for the body (schema), *Trends in Cognitive Sciences*, vol. 8, no. 2, pp. 79-86.

Perner, J. & Lang, B. (1999), Development of theory of mind and executive control, *Trends in Cognitive Sciences*, vol. 3, no. 9, pp. 337-344.

Rosenthal, D.M. (2000), Metacognition and Higher-Order Thoughts, *Consciousness and Cognition*, vol. 9, no. 2, pp. 231-242.

Sasaki, S., Hongo, T., Naitoh, K. & Hirai, N. (2008), The process of learning a tool-use movement in monkeys, with special reference to vision, *Neuroscience Research*, vol. 60, no. 4, pp. 452-456.

Seth, A., Baars, B. & Edelman, D. (2005), Criteria for consciousness in humans and other mammals, *Consciousness and Cognition*, vol. 14, no. 1, pp. 119-139.

Stout, D. & Chaminade, T. (2007), The evolutionary neuroscience of tool making, *Neuropsychologia*, vol. 45, no. 5, pp. 1091-1100.

Takeno, J., Inaba, K. & Suzuki, T. (2005), Experiments and examination of mirror image cognition using a small robot, *Proceedings of the 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp. 493-498.

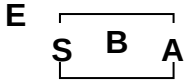
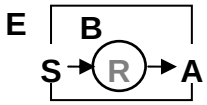
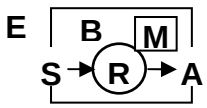
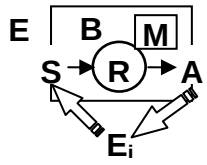
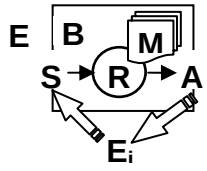
Turing, A. (1950), Computing Machinery and Intelligence, *Mind*.

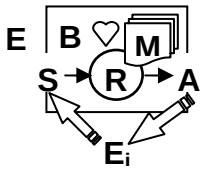
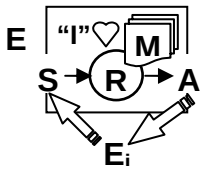
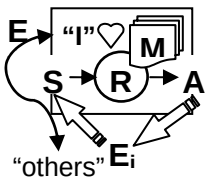
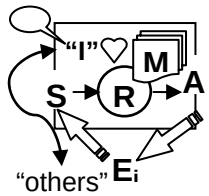
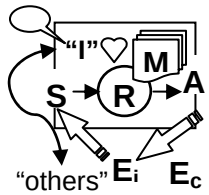
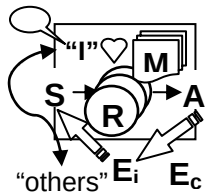
Van de Voorde, P., Sabbe, M., Rizopoulos, D., Tsonaka, R., De Jaeger, A., Lesaffre, E. & Peters, M. (2008), Assessing the level of consciousness in children: A plea for the Glasgow Coma Motor subscore, *Resuscitation*, vol. 76, no. 2, pp. 175-179.

Vygotsky, L.S. (1980), *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press.

Wooldridge, M. (1999), Intelligent Agents in *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, ed. G. Weiss, The MIT Press, pp. 27-78.

Table 1. ConsScale levels of consciousness.

Level of Machine Consciousness	Abstract Architecture	Short Description	Featured Behavior	Biological Phylogeny	Human Ontogeny
Level -1 *Disembodied*		Boundaries of the agent are not well defined. It can be confounded with the environment.	None. It is not a situated agent.	Amino acid as part of a protein.	n/a
Level 0 *Isolated*		There is an obvious distinction between body and environment, but no autonomous processing.	None. It is not a situated agent.	Isolated chromosome.	n/a
Level 1 *Decontrolled*		Presence of sensors and/or actuators, but no relation between them.	None. It is not a situated agent.	Dead bacteria	n/a
Level 2 *Reactive*		Fixed reactive responses. ***R*** establishes an output of ***A*** as a predetermined function of ***S***.	Primitive situatedness based on reflexes. Evolutionary learning.	Virus	n/a
Level 3 *Adaptive*		Actions are a dynamic function of both memory and current information acquired by ***S***.	Basic ability to learn and propioceptive sensing allow orientation and positioning behavior.	Earthworm	1 Month.
Level 4 *Attentional*		Attention mechanism selects E_i contents from ***S*** and ***M***. Emotions are present.	Ability to direct attention toward selected E_i allows attack and escape behaviors. Directed learning.	Fish	5 Months.
Level 5 *Executive*		Multiple goals can be interleaved as they are explicitly represented in memory.	Set shifting capability allows multiple goal achievement. Basic emotional learning.	Quadruped mammal	9 Months.

Level 6 *Emotional*		Feelings. Support for ToM stage 1: "I know".	Feelings provide a sense of self-status and influence behavior.	Monkey	1 Year.
Level 7 *Self-Conscious*		Support for ToM stage 2: "I know I know".	Self-reference makes possible advanced planning. Use of tools.	Monkey	1.5 Years.
Level 8 *Empathic*		Support for ToM stage 3: "I know you know".	Making of tools. Social behavior.	Chimpanzee	2 Years.
Level 9 *Social*		Support for ToM stage 4: "I know you know I know".	Linguistic capabilities. Ability to develop a culture.	Human	4 Years.
Level 10 *Human-Like*		Human like consciousness. Adapted Environment (E_c) and culture.	Accurate verbal report. Behavior modulated by culture (E_c).	Human	Adult
Level 11 *Super-Conscious*		Several streams of consciousness in one self.	Ability to synchronize and coordinate several streams of consciousness.	n/a	n/a