The Challenges for Implementable Theories of Mind

Pentti O. A. Haikonen

Department of Philosophy, University of Illinois at Springfield

Abstract
Implementable theories of mind would be of great value to the designers of artificial minds. Existing philosophical theories of mind tend to be loose and metaphorical and therefore do not provide very much guidance to a mind engineer. Unfortunately a complete implementable theory of mind does not yet exist even though there are several attempts toward that direction. The development of an implementable theory of mind faces several major challenges. Among these are the mind-body problem, the identification of the processes of mind, the problem of meaning and understanding, emotions, qualia and consciousness. These issues have been addressed via high-level algorithmic approach and low-level system approach and the combination of these, but each approach has proven to have its own challenges.

1 Introduction

Cognitive robots need brains and minds. Human brain has some 10^{14} synapses that are supposed to store memorized information. If one synapse were to store one bit then the brain's maximum memory capacity would be around 100 Terabits. On the other hand 32 Gigabyte (=2,56*10^{11} bits) miniature memory cards are now available and Terabyte memory cards are just around the corner. Biological synapses are not digital memory locations and their architectural organization is different from random access memories, but nevertheless the lesson is that semiconductor industry is now beginning to be able to produce devices with the circuit element density and complexity comparable to those of the brain. The brain is the site of the mind; does the aforesaid lead to the conclusion that artificial minds are just around the corner, too? The answer is a definite yes, provided that we are able to locate the correct corner. The correct corner is, of course, the implementable theory of mind. This, unfortunately, is not yet available in a concise, complete and tried engineering form even though several attempts towards this already exist. (e.g. Anderson et al 2004, Duch 2005, Haikonen 2003, 2007). Somehow it seems easier to explain the workings of the brain than to devise an engineering theory of mind that would allow the creation of a thinking machine. The brain operates independent of the correctness of the explanation, but a thinking machine will not work if the theory is not right.

What is a mind? What should a mind do? What kind of an information processing system can be called a mind? Should a mind be aware of itself, be self-conscious? What does it mean when something has a mind of its own? A theorist must look into these questions while looking for an implementable theory of mind. These are also issues that the philosophers of mind have treated over centuries. During these musings philosophers have stumbled on the mind-body problem; the apparent immateriality of mind and consequently, the apparent impossibility of interaction between the immaterial mind and the material body. It follows from the definition of material and immaterial substances that this problem is unsolvable, therefore implementable theories of mind cannot be dualistic ones in the sense of Descartes.

Whose mind is it? An artifact may behave as if it had a mind of its own, yet it may only be executing a collection of preprogrammed commands. In this case the artifact's operation reflects only the mind of the designer, not any of its own. Clearly no real mind has been designed or created. –A well is constructed for the water. However, a successful well digger does not supply the water, he only excavates a suitable hole for the water to seep in. In an analog way, a successful designer of mind should only design machinery that supports the mind and let the contents and caprices of the mind accumulate in the course of operation. In the following the constitutive aspects of an implementable theory of mind are examined.

2 What kind of a theory?

What kind of a theory would an implementable theory of mind be? Philosophical theories of mind tend to be abstract and metaphorical and consequently they are not very helpful for designers of artificial minds. Engineers are able to design systems as soon as the specifications for the system to be designed are given. A metaphor is not a proper specification, an algorithmic description of a desired function is. Thus, at first sight, it would seem that an implementable theory of mind should be algorithmic.

An algorithm is a sequence of instructions, which will lead to the desired outcome when executed properly. In a computer the instructions refer to the set of available operations such as the memory storage or recall, arithmetic operation, shifting a bit string, etc. A sequence of instructions that does not lead to a definite outcome should not be considered as an algorithm. Sometimes algorithms are seen as deterministic processes. However, this is not always the case as an algorithm may involve probabilistic and random operations, e.g. the utilization of randomly generated numbers.

The human mind appears to be non-deterministic; the mind is supposed to have "free will". Consequently the inaccurate idea that algorithms are necessarily deterministic may lead to the conclusion that the human mind must be non-algorithmic. For instance, Penrose has proposed that the mind would rely on non-algorithmic quantum mechanic processes (Penrose 1989). However, the operating temperature of the brain does not readily support quantum computing and the apparent freedom of will must have another explanation.

The operation of any system that obeys natural laws can, in principle, be simulated by algorithms; the accuracy of the simulation is another issue. The brain is such a system and the basic operation of individual neurons and synapses can be simulated with some accuracy. However, the real time computer simulation of a neural system with the complexity of the human brain and some 10^{14} synapses remains a really hard challenge.

High-level symbolic theories of mind are algorithmic and computational. These theories describe syntactic interactions between abstract entities, symbols, and in this way avoid the need to model and compute the operation of low-level units such as neurons and synapses. An early example of this approach is the computational theory of mind (CTM), proposed by Putnam (1961) and further developed by Fodor (1975). Newell and Simon (1975) had a similar idea. According to their Physical Symbol System Hypothesis a physical symbol system has the necessary and sufficient means for general human level intelligent action. Newell and Simon believed to have empirical evidence for this even though they admitted that the main evidence would be the absence of competing hypotheses, i.e. their proof was a proof by ignorance. Putnam and Fodor had a similar line of argument; they argued that mind is necessarily computational because symbolic computation is (as they claimed) the only known method to achieve results that otherwise can only be achieved via thinking; "it is the only game in town". However, so far nothing close to an artificial mind has materialized from these theories.

High-level symbolic theories provide algorithms that describe how further symbols are to be determined on the basis of given symbols. This computation is syntactic and as such does not require the grounding of meaning of these symbols, these do not have to refer to something. However, in practical applications, such as robots, the grounding of meaning is necessary. Robots are situated in and interact with the real word and consequently the mind of the robot must deal with real world entities. This leads to the practical problem: how the abstract symbols are to be derived from the information provided by the robot's sensors. This is a pattern recognition problem; the presence of an object is to be deduced from patterns of sensory signals. This is also a classification problem. Symbols stand for discrete well-classified entities that can be ordered into ontologies. This would work if it were possible to classify every entity in the world univocally. However, this is hardly the case, classes are artificial and arbitrary. Consequently, every object may be a member of not one but numerous classes (Clancey 1989).

The phenomenal aspects of mind such as the feel of pain, pleasure and perceptual qualia pose also a problem to symbolic theories, because these phenomena are supposed to take place at a sub-symbolic level.

High-level symbolic theories of mind can be formulated as computer programs and can be run on an ordinary computer.

Low-level sub-symbolic theories describe system reactions and interactions between low level signals in neural systems and architectures. The equivalents of higher level symbols may exist and may consist of a number of low level signals. Higher level symbols of this kind have fine structure and consequently modified symbols can depict modified entities. Absolute object recognition and classification is not necessary, an object may be seen in different roles depending on the context. Only the interactions between low-level signals are defined in algorithmic ways and are built in the neural architecture. Higher level cognitive functions arise from these via adaptation and learning. No implicit or preprogrammed algorithms for high-level operations are provided. The phenomenal aspects of the operation, if there will be any, are expected to be related to the dynamics of the system reactions that arise in the architecture. True realization of this approach calls for specific hardware that is able to support dynamic system reactions.

Low level sub-symbolic theories of mind can also be formulated as computer programs, which can be run on an ordinary computer. However, these executions should be seen only as simulations of the proposed neural hardware. The simulation of very large number of synapses usually calls for some simplifications and shortcuts in order to keep the processing time reasonable. Therefore these simulations do not necessarily produce all aspects of the theory and one should be critical and realistic when attributing phenomenal aspects to these simulations.

Which approach, the high level or low level, symbolic or sub-symbolic, would be the preferred one? Would a hybrid symbolic/sub-symbolic approach be able to combine the strengths of both approaches while avoiding their shortcomings? The brain is not a symbolic computer, but a biological neural network, which operates with sub-symbolic signals. Yet it manages to handle symbolic thought, too. Therefore, there must be a way in which a sub-symbolic system bridges naturally the gap between sub-symbolic and symbolic representations. For instance, Kelley has proposed that no gap actually exists, the sub-symbolic and symbolic representations are the ends of an intellectual continuum (Kelley 2003). In the same sense, Haikonen has proposed a way in which a neural system can utilize sub-symbolic representations as higher level symbols (Haikonen 2007).

Which aspects should an implementable theory of mind cover? Cognitive psychology has described many processes of mind and these can be used as a starting point. A successful theory should also explain meaning, qualia and consciousness in implementable terms.

An implementable theory of mind would be an engineering theory, which is described by commonly accepted engineering terms; mathematics, operational diagrams, circuit diagrams, system architectures and specifications. On the other hand, the aspects to be described belong to the realm of cognitive sciences. Here the interdisciplinary nature of this undertaking will be an interesting challenge and consequently engineers will have to study a bit of cognitive psychology and brain theories. Thereafter the engineering cycle of <identification of requirements – specification – design – test – revision> will hopefully meet this challenge.

3 The Processes of Mind

All animals that can execute motor responses have also more or less complicated nervous systems. One fundamental function of these nervous systems is the generation of motor response commands. In order to respond to something a nervous system must acquire information about that something.

Therefore some kinds of sensors that detect external and internal conditions are also necessary; the nervous system must be perceptive. In this kind of a system a motor response can be a reaction that is triggered by a sensory percept. Useful action may result, but sometimes blind reactive responses may be harmful or even fatal. A more complicated system may remedy this shortcoming by evaluating the fitness of the intended action with the help of experience; memorized instances of similar cases and the good/bad value of their outcomes. This calls for the ability to evoke memory-based imagery and to imagine itself executing the act. This, in turn, is related to the ability of "thinking about itself". If these capabilities were accepted as prerequisites for a mind then the minimum functions of a mind could be readily identified and they would be: Perception, reaction, deliberation and reflection. This conclusion has been reached and shared by Nilsson, Sloman and others including the author (Nilsson 1998, Sloman 2000a, 2000b). Thus the elementary functions of mind are seen as those of a controller and planner.

The above list of basic functions offers a good starting point, but a more detailed evaluation of the functions and processes is necessary for an implementable theory of mind. Cognitive psychology identifies the following processes of cognition: Perception, prediction, attention, learning, memory, understanding, reasoning, imagination, introspection, general intelligence, emotions, volition (See e.g. Aschraft 1998, Nairne 1997, Haikonen 2003). This list must be augmented with the additional functions and processes of pleasure, displeasure, pain, good/bad criteria and match/mismatch detection. Additionally, special and important hallmarks of human mind are the use of natural language and the flow of inner speech. However, these listings of cognitive functions should be mainly considered as kinds of check-lists; the listed functions and processes are not necessarily autonomous and independent of each other and some of them may be only loose descriptions of phenomena created by completely different processes. Nevertheless, the challenge for the potential developer of artificial mind theories becomes now visible; instead of actually clarifying the essential issues these lists highlight the wide spectrum of functions and processes to be quantified. Things are complicated further by the fact that these items relate to the functional layer of mind; the content layer is another story. Yet it is the content that determines what we are; our behavior, motives, values and culture. These are the subject of behavioral, social and cultural studies and go beyond the basic theory of mind.

4 Mind, Meaning and Understanding

Our thoughts are intentional; they are about something, they refer to something and have meaning that we understand. Likewise, a robot with an artificial mind should understand and have meaningful thoughts. Folk psychology has it that reasoning cannot take place without understanding and the utilization of meaning. However, it is known that mathematical and logical reasoning operates without meanings; no semantics, only syntactic rules. One plus one is two no matter what is being counted, be it apples or animals. It is exactly this abstraction property of mathematics and logic that make them so powerful. A computer works well without any grounding of meaning. Accordingly the computational theory of mind proposes that understanding can be effected via syntactic computation. This view has been criticized and opposed by e.g. Searle (1980, 1984, 1997). On the other hand the opponents of Searle have argued that syntax will somehow convey semantics if executed properly.

The question about semantics and syntax is a complicated one. In many cases it would seem that syntax would indeed suffice, but then there are cases that are not so clear. Consider the following examples:

- A candy bar has two sections. How many sections remain if one section is cut away? (two minus one is one).
- A candy bar has two ends. How many ends remain if one end is cut away? (two minus one is two).
- A triangle-shaped cookie has three corners. How many corners remain if one corner is cut away? (three minus one is four).

Mathematics may be context-free, but its application may not be. Simple arithmetic seems to fail in two examples here and correct answering seems to call for the visualization of the problems; the evocation of topological meaning. In general terms, in this example the "meaning" of an entity would seem to involve potential connections to a number of other mental concepts and physical world objects and "understanding" would seem to involve the proper activation of the relevant connections amongst all the possible connections and the consequent evocation of the relevant concepts.

An implementable theory of mind must address the problem of meaning and understanding properly. This requirement is especially apparent in the context of robotics. A cognitive robot must be able to understand what it is doing and why. Robots must also understand the commands given by their masters and they must be able to communicate their intentions to their masters. A robot cannot obey the command "go to the kitchen and bring me a soda can" if it does not understand the meanings of the words and the structure of the sentence, how these relate to the world and to the executable actions of the robot. But even this is not always sufficient. For instance, the master of the robot may give a verbal command: "Robot, please" or the master may simply snap his fingers. What is the robot supposed to do? This depends on the situation and context; perhaps the robot is expected to serve drinks to guests or escort somebody out. The robot must also understand the implicit conditions and limitations of each situation; while executing given commands the robot must not cause any collateral harm and damage.

5 Mind and Qualia

Human consciousness is characterized by qualia, the "phenomenal feel and quality" of every percept. Qualia are the way in which sensory information manifests itself in mind. Therefore, to be phenomenally conscious is to have qualia-based perception of the environment and self.

Qualia depict qualities of the sensed entities. The sensory faculties of vision and audition generate qualia that are related to the properties of the entities in the visual and auditory scene. It is known that visual and auditory stimuli are transformed by the eye and ear into neural signals that project into the depths of brain; yet the resulting qualia that depict visual objects and sounds seem to reside outside. In this way the individual comes to experience its existence as a center point in the world. This illusion does not readily take place in digital signal processing and therefore calls for a special explanation.

Qualia are subjective; there is no known way in which one's subjective experience, own feel of qualia can be transmitted to another person. However, the similarity of our biological built allows us to assume also similar feel of qualia. Thus we may assume with some confidence that a given real world quality such as that of a sound, taste or color will evoke same kind of qualia in different persons. But even here exceptions exist. For instance, a person with normal color vision has no way of knowing how a color blind person experiences the colors of red, green or brown. A color blind person may report no difference between these colors, but which would be the actual percept quale? Would it be the same as normal person's perception of red or perhaps green or brown? Or would it be something completely different?

Qualia are often associated with good/bad property; in fact they as themselves may feel pleasant or unpleasant. Music capitalizes on this property of qualia, the pleasantness of certain sounds, chords, rhythms and melodies. Without qualia music would be all but pointless. Thus, it seems that artificial minds that do not have qualia would not enjoy music in the same way as most humans do, if not at all, as the feel of enjoyment itself is based on qualia.

Computational theories of mind do not consider any feel of qualia as a necessary part of the cognitive process. In fact, it would be quite difficult to maintain that the execution of a computer program would involve any kind of subjective feel in a computer. Why would this feel be necessary anyway? Digital signal processing methods are quite able to handle qualities of the world. They can acquire and quantify information about physical qualities and represent these in numeric form. Powerful numeric algorithms for transformations, filtering, pattern recognition, motion detection and other signal processing tasks are available and can solve many of the related problems without any considerations of qualia. However, if necessary, computational qualia can be defined and represented as numeric values of variables: "if the variable p has the value ten, then the system is in great pain". But then, obviously this line of execution is an example of naïve anthropomorphism and should be recognized as such.

On the other hand, low-level sub-symbolic theories do not exclude the possibility of subjective qualia. For instance it has been proposed that the subjective feel of pain and pleasure would be related to system reactions in a system consisting of associative neuron networks (Haikonen 2003). Further research is called for also along this avenue.

At this moment the actual nature of qualia is still some kind of a mystery and a major challenge to any worthwhile theory of mind.

6 Mind and Emotions

Human mind is also characterized by the spectrum of emotions that can be triggered by various conditions. Emotions have been seen as non-rational states of mind that should not have any part in rational thinking. However, in recent years research has revealed that emotions do have an important role in cognition (LeDoux 1996, Damasio 2000, 2003). Percepts are seen to have emotional significance, which guides attention and modulates learning. Emotional significance is also seen to be an important factor in judgment and decision-making. Emotions seem to have motivational effects, too. Emotions have some connection to qualia; to be in an emotional state feels like something. In which way should emotions be incorporated in an implementable theory of mind? Could emotions be useful for a robot? Some attempts towards this direction already exist (e.g. Dodd and Gutierrez 2005, Haikonen 2003, 2007, Lee-Johnson and Carnegie 2006, Shirakura, Suzuki and Takeno 2006)

7 Mind, Consciousness and Self

Are mind theories also theories of consciousness? In nature minds and consciousness seem to go together, all beings that seem to have minds seem to be conscious, too. The content of consciousness is also mind's content at any moment even though mind is seen to involve also sub-conscious components. Usually mind and consciousness are attributed to an autonomous actor who is aware of itself, its mind and existence. A proper theory of mind should address also the problems of consciousness and self-consciousness.

The philosophy of consciousness divides the problem of consciousness into two parts, namely the so-called easy and hard problems (Chalmers 1995). The easy problem is related to the explanation of the cognitive functions that consciousness is supposed to execute or are otherwise associated with consciousness. The hard problem is related to the phenomenal aspect; consciousness as subjective qualia-based perception of the environment and self, the "feel". A developer of conscious machines may wish to define the focus of his pursuit along this demarcation. Machines that are supposed to execute the easy problem, but not the hard problem may be called "functionally conscious". Machines that execute also the hard problem may be called "phenomenally conscious".

The concept of "functional consciousness" is not without problems. This concept could be justified if consciousness actually executed a certain function. Consequently, a machine could be said to possess functional consciousness if it executed the same or a similar function. Baars (1997) proposes a number of functions for consciousness: Prioritization, access to unconscious resources (this is trivial tautology!), decision making and executive control, recruiting and controlling actions, error detection, understanding, and others. Given these functions there are two possibilities:

1. These functions are executed because the system is conscious, i.e. "consciousness executes these",

2. The system is conscious because these functions are executed by the system. In this case the style of execution may make the difference between a conscious and a non-conscious system.

Cognitive functions can be executed without consciousness and therefore are not a strong indication of any functionality of consciousness. On the other hand, decision making has been seen as a proof of the proposition that consciousness has a functional executive role. However, Libet's experiments and other studies (Libet 1993, Wegner 2003) seem to show that consciousness does not have decision-making power and decisions are made sub-consciously. Thus it may be possible that consciousness does not execute any function, instead it may be only an inner appearance in the system created by a special way of execution of the supposed functions of consciousness (Haikonen 2007). This leads to the following conclusion: If consciousness were only an inner qualia-based appearance with no function then no functional consciousness in the previous sense could exist. A system that would reproduce only the outer appearances of a naturally conscious system would not create the equivalent of the subjective qualia-based inner appearance of consciousness. Consequently no proper emulation or simulation of consciousness would take place. "Functionally conscious machines" would be functional but not conscious; the label would promise too much.

8 Mind and Inner Speech

Human mind is characterized by inner speech. In folk psychology inner speech is often seen as thinking and the main content of the human mind and is understood as a main difference between man, animals and machines.

The running of a computer program does not involve inner speech. Consequently inner speech has been largely ignored by AI researchers. However, cognitive psychology and neuroscience has seen inner speech as a key component of consciousness (e.g. Morin & Everett 1990, Morin 1993, 1999, 2003, 2005, Siegrist 1995, Schneider 2002). Recently also some machine cognition researchers have recognized inner speech as a relevant component of human-like cognition and consciousness (Clowes 2006, Haikonen 2003, 2006, 2007, Steels 2003a, 2003b). The relevance of inner speech to consciousness seems deceivingly obvious; how else could we know what we think if we did not hear our inner speech? This observation may easily lead to the conclusion that language and inner speech were necessary conditions for consciousness. However, this is not necessarily the case; there are also other forms of conscious thinking such as visual and kinesthetic imagination.

Humans explain their situation to themselves via the silent inner speech. Morin (2005) sees this self-talk as a device that can reproduce and extend social interactions leading to self-awareness. During social interactions people may receive comments about themselves, the way they are and behave. Inner speech may repeat these comments as such or as first-person transformations. This may lead to enhanced awareness of the commented features and to modified self-image.

Human mind is able introspect itself, it is self-aware. Duval and Wicklund (1972) define self-awareness as the state of being the object of one's own attention. This would include the paying of attention to one's own mental content such as percepts, thoughts, emotions, sensations, etc. Morin (1990, 2005) has seen inner speech as important means for introspection and processing of information about the self and the creation of self-awareness.

Inner speech utilizes a natural language. Natural language understanding and generation is a notoriously difficult discipline that, unfortunately, a mind theorist is not able to avoid. Existing machines do not "think" in a natural language and existing linguistic theories do not really help there. It may be possible that bold new approaches to linguistics will be called for.

9 Conclusions

An implementable theory of mind would be a theory that is expressed in engineering terms which allow the simulation of the processes of mind or the design of hardware systems that support the said processes. The contents and processes of mind should be meaningful, that is, mind objects should refer to real world entities. Thus perceptive processes are called for; these processes would execute symbol grounding. Cognitive psychology has identified several cognitive functions. It is obvious that an implementable theory of mind should be compatible with these.

Human mind operates with qualia. The act of perceiving the world through qualia seems to be the very essence of human consciousness. Artificial minds without qualia may be called functionally conscious, but not without problems; functional consciousness may be a valid concept only if consciousness actually executed some function and that function could be emulated.

Human mind utilizes inner speech. This inner speech is one hallmark of human consciousness that animals most probably do not share. Simple minds without inner speech can be envisioned, but a theory of mind may not be complete if it does not include the phenomenon of inner speech and allow communication via a natural language.

An implementable theory of mind should address the workings of the functional layer. Except for simple reactions and reflexes the behavior of the system with a mind would be determined by the mind's content; the accumulated experience, emotional states, motives and good/bad values. This process would be most interesting to observe in an artifact, yet it would belong to the realm of behavioral psychology and would be beyond the basic theory of mind.

The solving of the technicalities of mind would have important implications to information technology and also our own philosophical view about ourselves. The spectrum of unsolved issues provides great opportunities to creative researchers.

References

Anderson J. R., Bothell D., Byrne M., Douglass S., Lebiere Ch., Qin Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, 2004, Vol. 111, No 4, 1036 - 1060

Ashcraft M. H. (1998). *Fundamentals of Cognition*. New York: Addison Wesley Longman Inc.

Baars B. J. (1997). *In the Theater of Consciousness*. New York: Oxford University Press

Chalmers D. J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, Vol 2, No 3, 1995, pp. 200 – 219

Clancey W. J. (1989). The Frame of Reference Problem in Cognitive Modeling. *Proceedings of 11th Annual Conference of the Cognitive Science Society*, Ann Arbor. Lawrence Erlbaum Associates; 1989. pp. 107–114

Clowes R. (2006). The Problem of Inner Speech and its Relation to the Organization of Conscious Experience: a Self-Regulation Model. In T. Kovacs and J. Marshall (Eds.) *Proceedings of the AISB06 Symposium*. The Society for the study of Artificial Intelligence and the simulation of behaviour, UK. Vol. 1. pp. 117 – 126

Damasio A. (2000). *The Feeling of What Happens*. London: Vintage

Damasio A. (2003). *Looking for Spinoza*. USA: Harcourt Inc.

Dodd W., Gutierrez R. (2005). The Role of Episodic Memory and Emotion in a Cognitive Robot. *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication* (RO-MAN), Nashville, Tennessee 13 – 15 August 2005; 692–697

Duch W. (2005). Brain-Inspired Conscious Computing Architecture. In *The Journal of Mind and Behavior* Vol. 26 (1-2) 2005, pp. 1 - 22

Duval S., Wicklund, R. A. (1972). *A theory of objective self awareness*. New York: Academic Press

Fodor J. (1975). *The Language of Thought*. New York: Thomas Crowell

Haikonen P. O. (2003). *The Cognitive Approach to Conscious Machines*. UK: Imprint Academic

Haikonen P. O. (2006). Towards Streams of Consciousness; Implementing Inner Speech. In T. Kovacs and J. Marshall (Eds.) *Proceedings of the AISB06 Symposium*. The Society for the study of Artificial Intelligence and the simulation of behaviour, UK. Vol. 1. pp. 144 – 149.

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

Kelley T. D. (2003). Symbolic and Sub-symbolic Representations in Computational Models of Human Cognition. *Theory & Psychology* 2003;13(6):847–860

LeDoux J. (1996). *The Emotional Brain*. New York: Simon & Schuster

Lee-Johnson C. P., Carnegie D. A. (2006). Towards a Computational Model of Affect for the Modulation of Mobile Robot Control Parameters. *3rd International Conference on Autonomous Robots and Agents* (ICARA 2006) 12 – 14 December 2006, Palmerston North, New Zealand

Libet B. (1993). The neural time factor in conscious and unconscious events. *Experimental and theoretical studies of consciousness*, Wiley, Chichester (Ciba Foundation Symposium 174), 1993, pp. 123 – 146

Mckenna T. M. (1994). The Role of Interdisciplinary Research Involving neuroscience in the Development of Intelligent Systems. In: Honavar V, Uhr L, editors. *Artificial Intelligence and Neural Networks: Steps toward Principled Integration*. USA: Academic Press; pp. 75–92

Morin A. (1993). Self-talk and self-awareness: On the nature of the relation. *The Journal of Mind and Behavior*, 14: 223-234

Morin A. (1999). On a relation between self-awareness and inner speech: Additional evidence from brain studies. In *Dynamical Psychology: An Interdisciplinary Journal of Complex Mental Processes*. Retrieved from http://cogprints.org/2557/ on 14.12.2005

Morin A. (2003). Let's Face It. A review of The Face in the Mirror: The Search for the Origins of Consciousness by Julian Paul Keenan with Gordon C. Gallup Jr. and Dean Falk. *Evolutionary Psychology*, 1:161-171

Morin A. (2005). Possible links between self-awareness and inner speech: Theoretical background, underlying mechanisms and empirical evidence. In *Journal of Consciousness Studies*. Volume 12, No. 4-5, April-May 2005

Morin A., Everett, J. (1990). Inner speech as a mediator of self-awareness, self-consciousness, and self-knowledge: an hypothesis. In *New Ideas in Psychol*. Vol 8. (1990) No. 3, pp. 337 - 356

Nairne J. S. (1997). *The Adaptive Mind.* USA: Brooks/Cole Publishing Company

Newell A, Simon H.A. (1975). Computer Science as Empirical Inquiry: Symbols and Search. 1975 ACM Turing Award Lecture. *Communications of the ACM*. March 1976, Vol. 19 No. 3. pp. 113 – 126

Nilsson N. J. (1998). *Artificial Intelligence: A new Synthesis*. San Francisco: Morgan Kaufmann

Penrose R. (1989). *The Emperor's New Mind*. Oxford University Press.

Putnam H. (1961). Brains and Behavior. Originally read as part of the program of the American Association for the Advancement of Science, Section L (History and Philosophy of Science), December 27, 1961

Schneider J. F. (2002). Relations among self-talk, self-consciousness, and self-knowledge. *Psychological Reports*, 91: 807-812

Searle J. R. (1980). Minds, Brains, Programs. *The Behavioral and Brain Sciences*, number 3, 1980, pp. 417 - 427, Cambridge University Press

Searle J. R. (1984). *Minds, Brains & Science*. London England: Penguin Books Ltd

Searle J. R. (1997). *The Mystery of Consciousness*. London: Granta Books;

Shirakura Y., Suzuki T., Takeno J. (2006). A Conscious Robot with Emotions. *3rd International Conference on Autonomous Robots and Agents* (ICARA 2006) 12 – 14 December 2006, Palmerston North, New Zealand

Siegrist M. (1995). Inner speech as a cognitive process mediating self-consciousness and inhibiting self-deception. *Psychological Reports,* 76, pp. 259-265

Sloman A. (2000a). From intelligent organisms to intelligent social systems: how evolution of meta-management supports social/cultural advances. *Proceedings of the AISB'00 symposium on how to design a functioning mind*. UK: University of Birmingham; pp. 130 – 133

Sloman A. (2000b). Introduction: Models of Models of Mind. *Proceedings of the AISB'00 Symposium on How to design a functioning mind*. UK: University of Birmingham; pp. 1 – 9

Steels L. (2003a). Language Re-Entrance and the "Inner Voice". In O. Holland (Ed.), *Machine Consciousness*, pp. 173 – 185, UK: Imprint Academic

Steels L. (2003b). Evolving grounded communication for robots, In *Trends in Cognitive Science*, 7(7), July 2003, pp. 308 – 312

Wegner D. M. (2003). The mind's best trick: how we experience conscious will. *TRENDS in Cognitive Sciences*, Vol. 7 No 2 February 2003 pp 65 – 69