Machine Consciousness: A Computational Model

Janusz A. Starzyk¹ and Dilip K. Prasad²

¹School of Electrical Engineering and Computer Science, Ohio University, Athens, OH 45701 USA e-mail - starzyk@bobcat.ent.ohiou.edu

²School of Computer Engineering, Nanyang Technological University, Singapore – 639798 e-mail: dilipprasad@pmail.ntu.edu.sg

Abstract Despite many efforts, there are no computational models of consciousness that can be used to design conscious intelligent machines. This is mainly attributed to available definitions of consciousness being human centered, vague, and incomplete. Most researchers give up attempts of defining consciousness in physical terms, saying that consciousness is a metaphysical phenomenon. In this paper, we explain why it is important to define consciousness in physical terms. Through a biological analysis of consciousness and concept of machine intelligence, we propose a physical definition of consciousness with the hope to model it in intelligent machines.

Introduction

Understanding consciousness and implementing it in manmade machines has interested researchers for a long time. However, since last two decades, research towards making machines conscious has gained great momentum and a lot of work is being done towards this end by various researchers [1-17]. Despite the large amount of research efforts, research regarding computational modeling of machine consciousness is very limited. Among various technical, philosophical and computational difficulties, the primary reason that stands out is the difficulty in understanding consciousness and related abstract notions like thought, attention, awareness, etc. This paper addresses some of these issues as related to computational models of intelligence, and related modeling requirements for consciousness, attention and awareness.

Many researchers (philosophers, cognitive neuroscientists, psychologists, artificial intelligence researchers, etc.) tried to define or characterize consciousness [2, 10,

12, 13, 16, 17]. Many have discussed consciousness as causal (or non-causal) [12], accessible (or inaccessible) [13, 16], stateless (or having physical state) [13, 16], representational (or non-representational) [3, 5] and so on. However, none of the approaches provides a complete theory of consciousness that may become a foundation for modeling requirements to design conscious machines. Specifically, there is a lack of a physical description of consciousness that could become a foundation for building computational models.

In what follows, we justify the need for yet another definition of consciousness. There are several obvious reasons why such definition may be useful, such as follows:

- To introduce a unifying definition of consciousness with minimum behavioral, design, and structural requirements that can be used as test of consciousness.
- To move from meta-physical to physical description of consciousness and obtain its computational model.
- To describe underlying mechanism that can result in consciousness.
- To formulate a constructive approach towards the implementation of conscious machines.
- To describe consciousness in the context of emerging phenomenon in the process of perception, learning, and building associative memories.

We present our view of consciousness in relation to embodied intelligence ability to build stable sensory representations, and predict results of its actions in the environment. Self-organizing mechanism of emerging motivations and other signals competing for attention will be used to design models of conscious machines. In the next two sections, we briefly discuss the scientific and philosophical view of consciousness. The discussion presented is not exhaustive. It is rather representative of some important works that served as our inspiration. In the subsequent section, a discussion on emergence of consciousness is presented. This discussion helps us to understand consciousness from biological perspective. The section after this takes up the main task of defining consciousness. In this section, we first identify requirements for a conscious machine. Then, we build upon these requirements to define consciousness in physical terms. After this, we propose a computational model of machine consciousness based on our definition. Finally we conclude the paper.

Scientific view of consciousness

John Searle [33] said that "Studying the brain without studying consciousness would be like studying the stomach without studying digestion, or studying genetics without studying the inheritance of traits".

Marvin Minsky discusses consciousness in his book "The Emotion Machine" [18]. He analyzes consciousness from the point of view of common sense, as well as presents views of other thinkers, philosophers, neuropsychologists, researchers in artificial intelligence and cognitive science. Their views of consciousness are different, from everything that makes human spiritual experiences, mystical links between sensing and highest levels of mind, to statements that "nobody has a slightest idea of how anything material can be conscious" [19]. According to William Calvin and George Ojeman [20], consciousness refers to focusing attention, mental rehearsal, thinking, decision making, awareness, alerted state of mind, voluntary actions and subliminal priming, concept of self and internal talk. Sloman [21] suggests that it may be pointless trying to define consciousness, its evolution or function as they may have many different interpretations, similar to other big words like perception, learning, knowledge, attention, etc. Minsky points out that philosophers do not help in understanding consciousness, nor give recipe on how to test one.

Jeff Hawkins [22] suggests that consciousness is a combination of self awareness and qualia (feelings associated with sensations but not related to sensory input). He also points out that consciousness is associated with declarative memory; the moment this memory is erased our conscious experience disappears. He is certain that memory and prediction play crucial roles in creating consciousness, whatever way one defines it.

Susan Greenfield's [23] concept, 'continuum of consciousness,' says that "Consciousness is a dynamic process and it changes with development of brain. Further, at macro-level there is no consciousness centre and at micro-level there are no committed neurons or genes dedicated to consciousness."

Philosophical view of Consciousness and Awareness

Higher-order theory supported by Rosenthal [8] postulates the existence of a pair of distinct mental states: a first-order quasi-perceptual state, and a higher-order thought or perception representing the presence of that first-order state. In higher order theory "phenomenally conscious states are those states that possess fine-grained intentional contents of which the subject is aware, being the target or potential target of some sort of higher-order representation" [8].

Baars [2] says that "Consciousness is accomplished by a distributed society of specialists that is equipped with working memory, called a global workspace, whose contents can be broadcast to the system as a whole". Further Baars says [34] that "only one consistent content can be dominant at any given moment." The content of the memory is decided by the consciousness. Dennett [7] suggests that there is no single central place where conscious experience occurs; instead there are "various events of content-fixation occurring in various places at various times in the brain". When "content-fixation" takes place in one of these places, its effects may propagate so that it leads to the utterance of one of the sentences that make up the story in which the central character is one's "self". Dennett believes that consciousness is a serial account for the brain's underlying parallelism.

Until now we had ignored an important question: Are consciousness and awareness the same thing? This question is important because many researchers often confuse these terms and use them interchangeably. In order to differentiate between consciousness and awareness, let's explore them from philosophical perspective.

Though most people perceive these two words as meaning basically the same thing, philosopher Nisargadatta [24] points to two very different meanings of consciousness and awareness. When he uses the term "consciousness", he seems to equate that term with the knowledge of "I Am". On the other hand, when he talks about "awareness", he points to the absolute, something altogether beyond consciousness, which exists non-dualistically irrespective of the presence or absence of consciousness. Thus, according to him, awareness comes first and it exists always. Consciousness can appear and disappear, but awareness always remains.

Our interpretation of his theory is that contrary to the general belief, awareness is not a part (subset) of consciousness; in fact, he suggests that awareness is the superset of consciousness. In some sense he is relating awareness to something similar to a central executive, which remains whether we are conscious or not, and takes care of all the biological activities without actually making us conscious of them happening. We adopt his approach in our model.

Emergence of consciousness

To further understand consciousness, we also need to understand - when does human consciousness emerge? What makes conscious beings aware of the fact that they are conscious? In the following, we discuss these questions from two viewpoints: developmental stages of human fetus and evolution of brain.

In human fetus, the first simple movement can be interpreted as the first indication of consciousness. Lou [25] argues against it. He argues that 8 weeks fetus' nervous system is in nascent state and its movement is the result of simple reflexes similar to the movement of headless chicken.

Development stages of brain indicate that consciousness can be a feature of cortex. In human fetus, cortex develops over many months and in stages. Spreen et. al. [26] say that cortical cells come at the correct position in the 6th week after gestation. 20th week onwards, cortical region is insulated with myelin sheath and from 25th week, the development of local connections between neurons takes place. 30th week onwards, fetus' brain generates electrical wave patterns. These developments are gradual and incomplete until after birth. This makes it difficult to determine the exact moment of emergence of consciousness. On the other hand, we might conclude that this may be the reason for limited consciousness exhibited by the fetus, and that consciousness emerges gradually with the development of brain. Our analysis of relation between consciousness and the evolution of brain (based on [23, 27]) is summarized in table 1. In this table, we also indicate the current ability of machines to model and mimic related evolutionary traits. This study is in agreement with Gerry Edelman [28] who suggested that lower animals are not conscious. However, there are some interesting exceptions. For example, Young [29] observed that octopus, though an invertebrate, possesses sophisticated learning and memory skills. We also observe that most animals, including the animals with simpler brains, such as insects, exhibit some kind of circadian sleep wake cycle [30]. This, in our opinion, is a crude form of consciousness i.e., consciousness is not yet developed to treat them as conscious being according to our definition of consciousness.

From the table, we conclude that some important traits necessary for consciousness include the presence of cortex, cross-modal representation, associative memory, and learning units.

		Living Being ¹	Evolutionary traits	Analogous feasibility in machines
↑	Conscious	Human Beings	Fully developed cross-modal representation Sensory capabilities: auditory, taste, touch, vi- sion, etc. Bi-frontal cortex: planning, thought, motivation	Impossible at present
		Hedgehog (earli- est mammals)	Cross-modal representation Sensory capabilities: auditory, touch, vision (less developed), etc. Small frontal cortex	Impossible at present
		Birds	Primitive cross-modal representation Sensory capabilities: auditory, touch, vision, ol- factory. Primitive associative memory	Associative memories
	*	Reptiles ²	Olfactory system Primitive vision	Computer vi- sion (nascent)
	Not Conscious	Hagfish (early vertebrate)	Primitive olfactory system Primitive nervous system	Artificial neu- ral networks
		Lower level ani- mals (hydra, sponge, etc.)	Sensory motor units Point to point nervous system	Mechanical and/or elec- tronic control systems
¹ Kingdom Animalia ; ² * inconclusive\consciousness in transition				

Table 1: Evolution and consciousness.

Harth [31] has related consciousness to connections between the lateral geniculate nucleus (LGN) of the thalamus and the corresponding visual cortex. Thus we can conclude that the presence of cortex is important for emergence of consciousness, though consciousness may not be located specifically in cortex.

Defining Consciousness

For designing models of conscious intelligent machines, we need to define consciousness in the context of such machines. We will adopt an approach similar to the one we took providing definition of intelligence [32], where our aim was not to remove ambiguity from philosophers' discussion about intelligence or various types of intelligence, but to describe mechanisms and the minimum requirements for the machine to be considered intelligent. In a similar effort, we will try to define machine consciousness in functional terms, such that once a machine satisfies this definition, it is conscious, disregarding the level or form of consciousness it may possess. Consciousness will be very much a function of embodied form of intelligence that a machine will possess. It will be an emerging property of machine development and its final form will depend on perception, memory, motor skills, motivations, thoughts, plans, etc.

In our prior work [32] on motivation in machine learning, we defined embodied intelligence as a mechanism that learns how to survive in a hostile environment. Thus, learning ability is a critical aspect of intelligent machine and knowledge is a result of learning.

Since our aim is to design conscious machines, which exist and interact in an environment, we will use the most successful paradigm of building intelligent machines based on embodiment. Embodied intelligence uses sensors and actuators within its embodiment to perceive and act on the environment. In this interaction with environment, it learns to recognize objects and learns effects of its actions. By learning limitations of its embodiment and predicting how its own embodiment may affect the environment around it, machine learns to be aware of itself in a given environment. Thus in order to be aware of "who I am?", to predict the results of own actions, to anticipate, etc., a *mechanism to acquire and represent knowledge* about the environment is required.

In order to be aware of "who I am?", attention to self is required. Similarly, for being conscious about other events/objects, attention needs to be given to those events/objects. Thus, *a mechanism for attention and attention switching* is required.

There are few interesting questions regarding attention and attention switching. When we are conscious of an object, our attention is focused on that object. During a thought process our attention switches from one object/fact/event to another one. What happens during this attention switch or what is the mechanism for attention switching during a thought process? Are we conscious during the attention switch? This should not be confused with the D. Rosenthal's statement that "higher-order thoughts are themselves seldom conscious; so we are typically unaware of them" [8] as he focuses on a conscious thought not the underlying mechanism which creates a conscious thought. In our model there is no unconscious thought.

One possible explanation of the *mechanism of attention switching* is that it is a result of competition between *functional units* of brain. Attention switches to the semantic relation (object/fact/event) corresponding to the unit that wins. Sometimes, we think of an object/fact/event, but our attention drifts to something else, which was not directly related to this object/fact/event. This can be easily explained by the competition between numerous semantic relations as our *brain explores in parallel* while reaching a target goal or externally driven unconscious stimuli that switch our attention.

Now, we present our views on the state of consciousness during attention and attention switching. For this purpose, we make use of the definitions of attention and attention switching.

Attention is a selective process of cognitive perception, action or other cognitive experiences. This selective process of attention results from attention switching (needed to have cognitive experience).

Attention switching is a dynamic process resulting from competition between representations related to motivations, sensory inputs and internal thoughts including spurious signals (like noise). Thus attention switching may be a result of deliberate cognitive experience (and thus fully conscious signal) or it may result from subconscious process (stimulated by internal or external signals). Thus, while paying attention is a conscious experience, switching attention does not have to be.

Note, that in this respect the major mechanism responsible for conscious thoughts and their dynamics combines both top-down (conscious) and bottom-up (unconscious) signals. This is different from Rosenthal HOT theory [8] or Baar's Global Workspace [34] which are entirely driven by conscious thoughts. Our approach is closer to Dennett's "frame in the brain" idea [7] where coalitions of neurons compete for the frame with winners becoming a conscious experience.

Another requirement for consciousness is an act of cognitive perception, actions related to it, and/or other cognitive experiences like thoughts or dreams. Cognitive perception or other cognitive experiences are related to *semantic memories* resulting from *knowledge building*. In addition, *associative memory* is required to relate perception to knowledge.

Even as we discuss various requirements for consciousness, we need to note that brain can't be easily compartmentalized, i.e., there is no point-to-point connection between senses and neurons in brain. For example, visual system activates at least 30 different areas of brain. Similarly, no single brain region (in agreement with Susan Greenfield [23] and D. C. Dennett [7]) or no single chemical process in the

brain is responsible for consciousness. We identify that consciousness is a result of interactions between interconnected modules.

As discussed, to be conscious, the ability to define "Who I am?" is essential. In functional terms, it means that the system can perceive, act, and predict results of its actions, including the effect of its own embodiment on the environment (that may include limitations resulting from this embodiment). Thus, self-awareness is a synonym for consciousness at its minimalistic functional meaning. Following Nisargadatta, we accept general awareness as a prerequisite for consciousness. A plant is aware of light and cold, yet it is not conscious. Consequently, in our definition, consciousness requires cognitive awareness.

A central executive, which operates no matter whether we are conscious or not, is required as the platform for the emergence, control, and manifestation of consciousness. In human, central executive takes care of all the biological activities without making us aware of what is happening, as well as of all cognitive perceptions, thoughts and plans. In machine, central executive will control its conscious and subconscious processes, driven by its learning mechanism, creation and selection of motivations and goals. Thus, central executive, using cognitive perception and cognitive understanding of motivations, thoughts or plans will be responsible for self-awareness and create conscious state of mind.

We define machine consciousness as follows:

A machine is conscious if besides the required components for perception, action, and associative memory, it has a central executive that controls all the processes (conscious or subconscious) of the machine; the central executive is driven by the machine's motivation and goal selection, attention switching, learning mechanism, etc. and uses cognitive perception and cognitive understanding of motivations, thoughts, or plans. Thus, central executive, by relating cognitive experience to internal motivations and plans, creates self-awareness and conscious state of mind.

Computational model of consciousness

In this section, we propose a computational model that integrates functionalities of biological systems in a virtual machine. We are taking functional inspiration from biological systems to model consciousness and realize it in machines. It should be noted that we are not advocating to mimic biological systems exactly in machines, but rather to model and use their functional organization of conscious processing.

Our model consists of three main functional blocks, viz., Sensory-motor, Episodic Memory and Learning, and Central executive. A detailed block diagram of the model is presented in Figure 1.

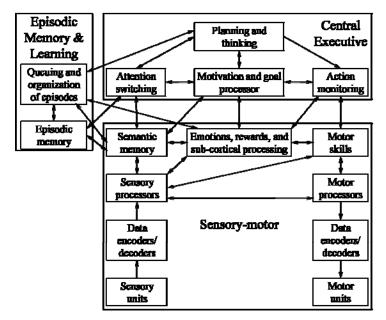


Fig. 1: Computational model of a conscious machine.

Sensory-motor block

Sensory-motor block is composed of three parts: sensory processors integrated with semantic memory, motor processors integrated with motor skills, and subcortical processor integrated with emotions and rewards.

The sensory processors are connected to the sensors through encoders/decoders. They receive the sensory data and are responsible for concept formation in selforganized hierarchical structure [22]. Through interaction with central executive, episodic memory and sub-cortical processor sensory processors build and activate semantic memory that represents knowledge about the environment. Semantic memory blocks activated by various sensory inputs are interconnected with each other making the cross-modal representation in such systems possible.

Similarly, the motor processors are connected with the motor units through encoders/decoders. The processors actuate the motors and receive feedback through sensory processors. Through interaction with central executive and sub-cortical processor they build a hierarchy of motor skills.

The processors in the emotions, rewards, and sub-cortical processing are used for generation of emotional and reward signals that govern learning and serve as an interface to other units. Specifically they cue episodic and semantic memories,

switch attention, provide motivations and help to select goals and interact with action monitoring.

Multiple processors in the sub-cortical processor block execute their programs in parallel and generate individual outputs. These outputs may compete among themselves (at a subconscious level) or may be used by the central executive unit (at a conscious level) to make a selection. Such a mechanism shall be helpful for attention, goal selection, motivation, etc.

Episodic memory and learning block

Episodic memory and learning block is composed of two parts, episodic memory unit and learning unit. Episodic memory unit is a collection of smaller functional blocks, each dedicated to capture a spatio-temporal sequence of semantic relationships like relations between objects observed in an episodic experience with their significance derived from emotional context.

Cueing and organization of episodes unit is able to recognize the novel events/patterns in various processes and help to build semantic relationships. For doing so, it shall collect and process data from all the units including motivations and interpretations of cognitive experiences from the central executive. Subsequently, it stores and manages episodes and also initiates learning about specific events/patterns if directed by the central executive.

Central executive block

Central executive block is responsible for coordination and selective control of all the other units. This block interacts with other units for performing its tasks, gathering data and giving directions to other units. Its tasks include cognitive perception, attention, attention switching, motivation (based on goal creation [32] and winner-take-all), goal creation and selection, thoughts, planning, learning, etc. For this purpose, it needs the capability to dynamically select and direct execution of programs that govern attention, cueing, episodic memory and action monitoring. In addition, central executive can activate semantic memory and control emotions.

Central executive directs cognitive aspects of machine experiences but its operation is influenced by competing signals representing motivations, desires, and attention switching that are not necessarily cognitive or consciously realized. Central executive does not have any clearly identified decision making center. Instead, its decisions are result of competition between signals that represent motivations, pains and desires. At any moment, competition between these signals can be interrupted by attention switching signal. Such signals constantly vary in intensity as a result of internal stimuli (e.g., hunger) or externally presented and observed opportunities. Thus, the fundamental mechanism that directs machine in its action is physically distributed as competing signals are generated in various parts of ma-

10

chine's mind. Further, it is not fully cognitive, since, before a winner is selected, machine does not interpret the meaning of competing signals.

Cognitive aspect of the central executive mechanism is predominantly sequential, as a winner of the internal competition is identified and serves as an instantaneous *director* of the cognitive thought process, before it is replaced by another winner.

Once a winner of internal competition is established, central executive provides cognitive interpretation of the result, providing top down activation for perception, planning, internal thought or motor functions. It is this cognitive realization of internal processes that results in central executive's *decision* of what is observed, *planning* how to respond, internal *talk* of what to do, that we associate with a conscious experience and a continuous train of such experiences constitutes consciousness.

It should be noted that though the sensory and motor units are not the part of the brain as defined above, they are essential for embodiment of the complete system.

Conclusion

Opposed to the metaphysical interpretation of consciousness, we present a physical definition of consciousness based upon the biological study of consciousness and our model of embodied intelligence [32]. Though, our definition of consciousness is based on biological perspective, the proposed definition clearly encompasses various possible phenomenological characteristics of consciousness.

Our proposed organization of conscious machine model is based on two important observations. First, biological evolution as well as development of human brain indicates that a functional unit similar to pre-frontal cortex is strongly related to the emergence of consciousness. Second, a central executive which controls and coordinates all processes, whether conscious or subconscious, and which can perform some of its tasks (like memory search) using concurrent dynamic programming, is necessary for developing consciousness.

The proposed computational model of consciousness mimics the biological systems functionally and retains a well-defined architecture necessary for implementing consciousness in machines. It might be neither complete or foolproof nor practically feasible. However, it should provide guidance towards building models of conscious embodied machines.

References:

- [1] M. L. Anderson and T. Oates, "A review of recent research in metareasoning and metalearning," AI Magazine, vol. 28, pp. 7-16, 2007.
- [2] B. J. Baars "A cognitive theory of consciousness," Cambridge University Press, 1998.
- [3] R. Chrisley, "Embodied artificial intelligence," Artificial Intelligence, vol. 149 (1), pp. 131-150, 2003.

- [4] R. Clowes, S. Torrance, and R. Chrisley, "Machine consciousness Embodiment and imagination," Journal of Consciousness Studies, vol. 14 (7), pp. 7-14, 2007.
- [5] P. O. A. Haikonen, "Essential issues of conscious machines," Journal of Consciousness Studies, vol. 14 (7), pp. 72-84, 2007.
- [6] E. T. Rolls, "A computational neuroscience approach to consciousness," Neural Networks, vol. 20 (9), pp. 962-982, 2007.
- [7] D. C. Dennett, Consciousness Explained, Penguin Press, 1993.
- [8] D. M. Rosenthal, The nature of Mind, Oxford University Press, 1991.
- [9] A. Sloman and R. Chrisley, "Virtual machines and consciousness," Journal of Consciousness Studies, vol. 10 (4-5), pp. 133-172, 2003.
- [10]R. Sun, "Learning, action and consciousness: A hybrid approach toward modelling consciousness," Neural Networks, vol. 10 (7), pp. 1317-1331, 1997.
- [11]J. G. Taylor, "CODAM: A neural network model of consciousness," Neural Networks, vol. 20 (9), pp. 983-992, 2007.
- [12]M. Velmans, "Making sense of causal interactions between consciousness and brain," Journal of Consciousness Studies, vol. 9 (11), pp. 69-95, 2002.
- [13]M. Velmans, "How to define consciousness: And how not to define consciousness," Journal of Consciousness Studies, vol. 16 (5), pp. 139-156, 2009.
- [14]S. Densmore and D.C. Dennett, "The virtues of virtual machines," Philosophy and Phenomenological Research, vol. 59 (3), pp. 747-761, 1999.
- [15]D. Gamez, "Progress in machine consciousness," Consciousness and Cognition, vol. 17 (3), pp. 887-910, 2008.
- [16]D.C. Dennett, "Are we explaining consciousness yet?," Cognition, vol. 79 (1-2), pp. 221-237, 2001.
- [17]S. Blackmore, "There is no stream of consciousness," Journal of Consciousness Studies, vol. 9 (5-6), pp. 17-28, 2002.
- [18]M. Minsky, The emotion machine. New York: Simon & Schuster Paperbacks, 2006.
- [19]J. A. Fodor, "The big idea: can there be science of the mind," Times Literary Supplement, pp. 5-7, July 1992.
- [20]W. H. Calvin and G. A. Ojemann, Conversation with Neil's brain: the neural nature of thought and language: Addison-Wesley, 1994.
- [21]A. Sloman, "Developing concept of consciousness," Behavioral and Brain Sciences, vol. 14 (4), pp. 694-695, Dec 1991.
- [22]J. Hawkins and S. Blakeslee, On intelligence. New York: Henry Holt & Company, LLC., 2004.
- [23]S. Greenfield, The private life of the brain. New York: John Wiley & Sons, Inc., 2000.
- [24]Nisargadatta, I am that. Bombay: Chetana Publishing, 1973.
- [25]H. C. Lou, Developmental neurology. New York: Raven Press, 1982.
- [26]O. Spreen, A. T. Risser, and D. Edgell, Developmental psychology. New York: Oxford University Press, 1995.
- [27]G. Lynch and R. Granger, Big Brain: the origins and future of human intelligence. New York: Palgrave Macmillan, 2008.
- [28]G. Edelman, Bright air, bright fire: on the matter of the mind. New York: Penguin, 1992.
- [29]J. Z. Young, Programs of the brain. Oxford: Oxford University Press, 1978.
- [30]W. Kaiser and J. Steiner-Kaiser, "Neuronal correlates of sleep, wakefulness, and arousal in a diurnal insect," Nature, vol. 301, pp. 707-709, 1983.
- [31]E. Harth, The creative loop: How the brain makes a mind. New York: Addison-Wesley, 1993.
- [32]J. A. Starzyk, "Motivation in Embodied Intelligence," in Frontiers in Robotics, Automation and Control: I-Tech Education and Publishing, pp. 83-110, 2008.
- [33] J. R. Searle, "Consciousness," Annual Review of Neuroscience, vol. 23, pp. 557-578, 2000.
- [34]B. J. Baars, "The conscious access hypothesis: Origins and recent evidence," in Trends Cogn Science, vol. 6 (1), pp. 47–52, 2002.

12