Consciousness and the Quest for Sentient Robots

Pentti O.A. Haikonen

Department of Philosophy
University of Illinois at Springfield
One University Plaza
Springfield, IL, 62703, USA
pentti.haikonen@pp.inet.fi

Abstract. Existing technology allows us to build robots that mimic human cognition quite successfully, but would this make these robots conscious? Would these robots really feel something and experience their existence in the world in the style of the human conscious experience? Most probably not. In order to create true conscious and sentient robots we must first consider carefully what consciousness really is; what exactly would constitute the phenomenal conscious experience. This leads to the investigation of the explanatory gap and the hard problem of consciousness and also the problem of qualia. This investigation leads to the essential requirements for conscious artifacts and these are: 1.) The realization of a perception process with qualia and percept location externalization, 2.) The realization of the introspection of the mental content, 3.) The reporting allowed by seamless integration of the various modules and 4.) A grounded self-concept with the equivalent of a somatosensory system. Cognitive architectures that are based on perception/response feedback loops and associative sub-symbolic/symbolic neural processing would seem to satisfy these requirements.

1 Introduction

Are we any nearer to sentient robots? Can we create conscious robots by designing systems that emulate cognitive functions and integrate information, maybe within the frameworks of different cognitive architectures? Is this only a matter of system complexity? We may think so and the large number of proposed cognitive architectures [23] would also indicate a trend towards that proposition.

Indeed, the recent surge of research has brought forward new insights and ideas that should not be overlooked. The philosophical studies of Baars (e.g. [4]), Boltuc (e.g. [6]), Block (e.g. [5]), Dennett (e.g. [9]), Harnad (e.g. [16]), Sloman (e.g. [26]) and others have influenced many practical approaches and the important works of Aleksander & Morton (e.g. [2]), Hesslow (e.g. [17]), Kinouchi (e.g. [20]), Manzotti (e.g. [22]), Shanahan (e.g. [25]), to name a few, are well-known. From the engineering point of view the empirical study of consciousness is an exercise in embodied cognitive robotics. Here Chella (e.g. [8]), Holland (e.g. [18]), Kawamura (e.g. [19]), Sanz (e.g. [24]), Takeno (e.g. [27]) and others have done important research.

A. Chella et al. (Eds.): Biologically Inspired Cognitive Architectures 2012, AISC 196, pp. 19–27.
springerlink.com

Therefore, we should be confident that progress is being made and conscious cognitive robots should be just around the corner. However, there may be a catch. Things may not be that easy. We may build robots that are able to execute a large number of cognitive functions and may in this way be able to mimic human behaviour successfully, but is something still missing? Do our robots really have "somebody inside" or are they only cleverly tuned automata? Do these robots really feel something and experience their existence in the world in the style of human conscious experience? And, what exactly would constitute this phenomenal conscious experience, what is consciousness actually? This is a fundamental question and a difficult one that is easily pushed aside for that very reason. However, there will be no true conscious machines unless this issue is satisfactorily solved.

2 The Problem of Consciousness

It has been said that we all know, what consciousness is; it is the state in which we are, when we are not unconscious. Apart from that consciousness is easily assumed to be a phenomenon that nobody can really explain.

On the other hand, it has also been proposed that consciousness were computational [10] and were related to "information integration" [28, 25]. This is not necessarily so. Rather than seeing the problem of consciousness as a problem of cognitive functions, skills and their integration, it should be seen as a problem of phenomenal experience.

It is a fact that the brain is a neural network and our thoughts and feelings are based on the activity of neurons, synapses and glia. On the other hand, it is also fact that we do not perceive the activity of the brain in this way, as the activity patterns of neurons. Instead, these activity patterns appear to us as our sensory percepts, thoughts, sensations and feelings directly. However, most of the activity of the brain does not have any internal appearance at all.

A typical example of this process is vision. The lens of the eye projects an image of the outside world on the retina. Light-sensitive receptor cells at each point of the retina transduce the light intensity into corresponding neural signal patterns. However, we do not "see" these neural signal patterns as such, instead our direct and immediate mental impression is that of an external world out there and its visual qualities. This constitutes our conscious awareness of the visual environment.

More generally, a perception process involves the transduction of the sensed physical phenomenon into neural signal patterns. These patterns have the internal appearance of apparent physical qualities as such, externalized to the outside world or to a body part. Visually seen objects are out there, sound sources are out there and if we hurt our finger or some other part of the body, the pain appears to be in the hurt body part, not inside the brain. The related neural activity is in the brain, but the actual appearance that is related to the sensed physical phenomenon, appears to be out there. We experience the world with its apparent qualities to be around us, not inside our brain. This appearance is most useful as it allows direct and easy interaction with the world.

How does the neural activity of the brain give rise to this kind of subjective experience and appearance? This is the hard problem of consciousness [7]. We can inspect the neural processes of the brain with various instruments and we can study and understand the physics of these processes to a great detail. On the other hand, we can study cognition and understand its functions, also to a great detail. We may also find the correspondence between the neural functions and cognitive functions; the neural correlates of cognitive functions.

3 Qualia and the Internal Experience

But then, there is the question of the internal appearance of the neural activity and the subjective experience. Why do our percepts appear as they are? Why does blue appear as blue, why does sweet taste like sweet, why does pain hurt? The qualitive appearances of our percepts are called qualia and no percepts seem to be without qualia. Again, we can find neural correlates of qualia, but we have not been able to explain, why these neural processes would give rise or be perceived as some subjective experience or qualia. This problem is called the explanatory gap [21]. This explanatory gap appears in the philosophy of the mind as the hard problem [7] and the mind-body problem, which has resulted in the nowadays not so popular dualistic theories of mind.

Could it be possible that the subjective experience would be an inherent property of biological neural activity patterns? That would appear to be an easy answer. In that case we might study cell biology and find the details of the process there, maybe. But then, this situation would exclude the artificial production of consciousness in non-biological systems.

However, if indeed, the subjective experience were an inherent property of biological neural signal patterns, then why does most of the activity of the brain remain subconscious, without any subjective experience at all? This observation would seem to show that the biological basis alone were not a sufficient or maybe even a necessary condition for the subjective experience and qualia. Instead, there must be something special in those neural signal patterns that appear as subjective experience. And there is. Introspection shows that the content of our conscious subjective experience consists of percepts with qualia. Consequently, the neural activity patterns perceived in this way are products of sensory perception. The problem of conscious perception would thus be reduced into the solving of the conditions for neural representations that could carry the qualities of the sensed entities in the form of qualia.

Qualia are direct, they need no interpretation or any additional information in order to be understood. Red is red and pain is pain directly, there is no need to evoke some learned meaning for those. The directness of qualia gives also an indication of the nature of the neural transmission that carries qualia. The transmission must be direct and transparent so that only the carried information conveys the effect, while the carrying neural machinery remains hidden. The directness of qualia also excludes the symbolic representation of sensory information. A symbolic representation would be a description, while qualia are the experience, the experienced qualities of the sensed entities.

In the brain there are no sensors that could perceive neurons and their firings as such. Therefore, only the effects of the neural signals are transmitted. The neural activity cannot appear or be perceived as neural firings, yet it has effects. In sensory perception circuits these effects are related to the qualities of the perceived entities. Therefore, it would seem that on the system level, "in the mind", the only possibility is the appearance of sensory signals as externally grounded qualia. All inner neural activity does not have this direct relationship to perception and without this it remains "sub-conscious".

4 Introspection

However, not all of our conscious mental content is related to direct sensory perception. Our inner speech is inside, our imagery is inside and our feelings are inside. Yet, there are no sensors inside the brain that could sense the internal neural activity patterns. If consciousness were related only to sensory perception, then how could we become aware of our own thoughts? This is the problem of introspection.

Therefore, is the hypothesis about consciousness as a sensory perception–related phenomenon wrong? Not necessarily. The problem of introspection can be solved by associative feedback loops that return the products of the internal processes into virtual percepts so that the mental content can be observed in terms of sensory percepts. In this way the perception process facilitates also the awareness of mental content. Introspection shows that we perceive our mental content in the form of virtual sensory qualities; we perceive and hear our verbal thoughts in the form of silent inner speech and perceive our imaginations as vague imagery. Without hearing the inner speech we would not know what we think. These kinds of feedback loops have been proposed by the author [11-15] and others, e.g. Chella [8] and Hesslow [17], who has proposed that thinking is simulated perception.

5 Reportability and Information Integration

Conscious states are reportable states that can be remembered for a while. A report is a "message" or a "broadcast" that is sent to various parts of the brain or an artificial cognitive apparatus. The message must be received and have effects on the receiving part, such as the forming of associative memories. These effects may also manifest themselves as motor actions including motions. The motions may include actions like reaching out or the turning of the head. Voiced reactions and verbal comments are possible forms of reports, like "Ouch" and "I see a black cat". The report may be about perceived external entities and conditions or internal ones, like "I feel pain" or "I feel like drinking soda".

The verbal reports may also remain silent, as a part of the inner imagery or silent speech. A verbal or non-verbal internal report may affect the behaviour of the subject and may, for instance, lead to the planning of new actions. Reporting involves the activation of cross-connections between the various parts of the brain. This kind of cross-coupling is sometimes called information integration. Tononi [28] has proposed that the degree of information integration could be used as a measure of consciousness. Information aspects of consciousness have been treated also by Aleksander and Morton [2].

6 The Concept of Self

We perceive the world with its apparent qualities to be around us, not inside the brain, but what is inside? Inside is the mind and the *impression of I*, that appears as the conscious, perceiving, feeling and thinking self. We are aware of ourselves, we are self-conscious. We are the system that is aware of the perceived qualia.

Which comes first, the perceiving self or the qualia? They come together. It is proposed that the qualia are a property of certain kinds of embodied perceptual systems and the impression of self arises from the qualia of self-percepts and also from the introspected mental contents. The "I" is associated with the body and its sensations. The somatosensory system operates as fundamental grounding to the self-concept, because it provides continuous information about the state of the body. The body and its sensations are with us wherever we go, while the environment changes. Thus the body is a fixed point of reference that can be associated with all things that make us an individual; our personal memories, needs, desires, etc. But more directly; the body is the vantage point for our percepts. The somatosensory percepts about the body relate directly to the system-self. For instance, when we feel pain, it is us who feel the pain, because the pain itself creates that impression. Thus, the conscious, perceiving, feeling and thinking self is a rather simple product of an embodied perceptual system.

What is Consciousness?

According to the previous chapters, consciousness is seen as the condition of having reportable qualia-based (phenomenal) perception. As such, it is only an internal appearance, not an executing agent. The difference between a conscious and non-conscious subject is the presence of the internal qualia-based appearance of the neural activity, which manifests itself as the direct perception of the external world with its qualities, the perception of the body with its sensations and the introspective perception of mental content such as imagery, inner speech and feelings. A non-conscious agent may process similar information, but without the experience of the direct internal appearance. It does not perceive its internal states in terms of world qualities, it does not perceive them at all.

Consciousness is not directly related to cognitive abilities or intelligence. Most probably even the simplest animals experience their consciousness in the same way as humans, as the direct qualia-based appearance, but the scope, content and fine structure of their conscious experience may be extremely narrow. However, a system that supports consciousness, the phenomenal internal appearance of its internal states, may be better suited for the realization of genuine general intelligence.

Consciousness and qualia go together. Without experienced qualia there is no consciousness. Qualia are sub-symbolic and therefore cannot be created by symbolic means. Therefore there are no algorithms that could create consciousness. Programmed artifacts cannot be conscious. This statement does not exclude conscious awareness of symbolic representations in otherwise conscious agents.

7 Sub-symbolic and Symbolic Processing

The creation of artificial conscious robots involves some system-technical issues beyond the fundamental issues of perception with internal appearances and qualia.

Human cognition utilizes sub-symbolic qualia, but is also based on symbolic information processing. There could not be any language or mathematics without the ability to use symbols. Yet, qualia are sub-symbolic and the very preconditions for qualia would seem to exclude symbolic processing. This problem has been evident in the traditional artificial neural networks. These networks operate in sub-symbolic ways and are unable to run programs. On the other hand, digital computers run programs, but are not able to process sub-symbolic representations directly. Thus, a gap between sub-symbolic and symbolic systems would seem to exist. The brain is able to bridge this gap, but how should this bridging be executed in artificial systems?

There have been some hybrid approaches, combinations of neural networks and program-based digital computers. Hybrid solutions are easily clumsy and too often combine the shortcomings of the component systems. The brain is definitely not a hybrid system.

An elegant non-hybrid solution to the sub-symbolic/symbolic problem is offered by the use of associative neural networks with distributed representations [13, 15]. These operate with sub-symbolic neural signals and signal patterns, which can be associated with each other. Via this association sub-symbolic signal patterns may become to represent entities that they do not directly depict; they become symbols for these entities. In this way, for instance, neural signals that are caused by heard sound patterns can be used as words that have associated meanings that are not naturally related to the sounds themselves.

8 Criteria for Consciousness

According to the aforesaid, the first and necessary criterion for consciousness is the presence of the phenomenal, qualia-based internal appearance of the neural activity. Unfortunately this internal appearance is subjective and cannot be directly measured unless some ingenious methods are invented. Therefore some indirect criteria for consciousness must be used. Such criteria have been developed by various researchers and include, for instance, the Aleksander's axioms [1] that try to define prerequisites for conscious systems and the ConsScale of Arrabales [3] that tries to determine the scope of consciousness by evaluating the presence of a number of behavioural and cognitive functions.

9 Requirements for Cognitive Architectures

The afore presented issues lead to the outlines and requirements for a cognitive architecture for artificial brains. The essential requirements relate to 1.) The realization of a perception process with qualia and percept location externalization, 2.) The realization of the introspection of the mental content, 3.) The reporting allowed by seamless integration of the various modules and 4.) A grounded self-concept with the equivalent of a somatosensory system. In addition, the architecture should facilitate the transition from sub-symbolic to symbolic processing.

It is proposed, and there is also some experimental proof that an architecture based on perception/response feedback loops and associative sub-symbolic/symbolic neural processing would satisfy these requirements. The author's "Haikonen Cognitive Architecture" (HCA) is an architecture realized in this way [13, 15].

10 An Example: The Robot XCR-1

It was argued afore that qualia cannot be simulated and consciousness cannot be created by programs. Therefore hardware experiments with embodied robots are necessary.

The author's Experimental Cognitive Robot XCR-1 is designed as a simple test bed for machine cognition experiments that involve direct perception and sensorimotor integration with motor action generation and control using the HCA architecture and associative neural processing [14, 15].

The robot XCR-1 is an autonomous, small three-wheel robot with gripper arms and hands, and simple visual, auditory, touch, shock and petting sensors. The XCR-1 has a limited natural language self-talk that reflects its instantaneous percepts and actions. It recognizes some spoken words and has a limited vocabulary self-talk and this allows limited verbal learning.

The sub-symbolic/symbolic processing properties of the HCA cannot be fully produced by computer programs. The HCA is also a parallel processing system and this is another benefit that would be lost if microprocessors were used. Therefore the robot XCR-1 utilizes hardwired neural circuits instead of the more common program-driven microprocessors and possible links to a master computer, which would execute the more demanding computations. Thus, the XCR-1 is not program-driven and represents a completely different approach to autonomous self-controlled robots.

The XCR-1 has a set of motor action routines and hard-wired reactions to stimuli. These routines and the hard-wired reactions can combine in various ways depending on the situation. Cognitive control may override reactions and modify the robot's behavior.

The XCR-1 utilizes the author's System Reactions Theory of Emotions (SRTE) [12, 13] for emotional learning and motivation. Corporal reward and punishment can be used to evoke pleasure and pain equivalent states. These can be associated with ongoing situations and in this way the robot will learn to seek or avoid the associated situations.

It is not claimed that the robot XCR-1 were conscious. The main purposes of this project have been the verification of the feasibility of associative information processing and the main principles of the HCA. However, XCR-1 would seem to satisfy certain requirements for consciousness, at least to a minimal degree, including:

- Direct sub-symbolic perception process
- The externalization of non-contact percepts
- Sensorimotor integration and information integration
- Attention
- Introspection of mental content in terms of sensory features via feedback loops
- Responses and reports including self-talk

- Emotional learning, control and motivation
- Somatosensory system for the grounding of the self-concept
- The transition from sub-symbolic to symbolic processing

11 Conclusions

It is argued here that consciousness is qualia-based perception. Qualia are the way in which the perception-related neural activity appears to us; not as the firing patterns of neurons, but as the apparent, externalized qualities of the world and the body. Therefore we experience us as being a vantage point in the world. All the other proposed aspects of consciousness build on the qualia-based perception. Feedback loops allow introspection, the awareness of mental content. Associative cross-connections, also known as information integration, allow internal and external reporting, memory making and recall as well as symbolic processing. Artificial realization of these call for direct, embodied perception and associative neural processing with a special cognitive architecture. All these requirements are implementable with available technology. Consequently, true conscious robots could be built in the foreseeable future, provided that the real hard problem of machine consciousness can be solved, namely money; the financing of the conscious robot design and development projects.

References

1. Aleksander, I., Dunmall, B.: Axioms and Tests for the Presence of Minimal Consciousness in Agents. In: Holland, O. (ed.) Machine Consciousness. Imprint Academic, UK (2003)
2. Aleksander, I., Morton, H.: Aristotle's Laptop. World Scientific, Singapore (2012)
3. Arrabales, R., Ledezma, A., Sanchis, A.: The Cognitive Development of Machine Consciousness Implementations. IJMC 2(2), 213–225 (2010)
4. Baars, B.J.: In the Theater of Consciousness. Oxford University Press, Oxford (1997)
5. Block, N.: On a Confusion about a Function of Consciousness. BBS 18(2), 227–287 (1995)
6. Boltuc, P.: The Philosophical Issue in Machine Consciousness. IJMC 1(1), 155–176 (2009)
7. Chalmers, D.J.: Facing Up to the Problem of Consciousness. JCS 2(3), 200–219 (1995)
8. Chella, A.: Perception Loop and Machine Consciousness. APA Newsletter on Philosophy and Computers 8(1), 7–9 (2008)
9. Dennett, D.: Are we explaining consciousness yet? Cognition 79, 221–237 (2001)
10. Edelman, S.: Computing the mind; how the mind really works. Oxford University Press, Oxford (2008)
11. Haikonen, P.O.: An Artificial Cognitive Neural System Based on a Novel Neuron Structure and a Reentrant Modular Architecture with implications to Machine Consciousness. Dissertation, Helsinki University of Technology (1999)
12. Haikonen, P.O.: The Cognitive Approach to Conscious Machines. Imprint Academic, UK (2003)
13. Haikonen, P.O.: Robot Brains. John Wiley & Sons, UK (2007)
14. Haikonen, P.O.: XCR-1: An Experimental Cognitive Robot Based on an Associative Neural Architecture. Cognitive Computation 3(2), 360–366 (2011)

15. Haikonen, P.O.: Consciousness and Robot Sentience. World Scientific, Singapore (2012)
16. Harnad, S.: The Turing Test Is Not a Trick: Turing Indistinguishability Is a Scientific Criterion. SIGART Bulletin 3(4), 9–10 (1992)
17. Hesslow, G.: Conscious thought as simulation of behaviour and perception. Trends in Cognitive Sciences 6(6), 242–247 (2002)
18. Holland, O., Knight, R., Newcombe, R.: A Robot-Based Approach to Machine Consciousness. In: Chella, A., Manzotti, R. (eds.) Artificial Consciousness. Imprint Academic, UK (2007)
19. Kawamura, K., Gordon, S.: From intelligent control to cognitive control. In: Proc. 11th International Symposium on Robotics and Applications, ISORA (2006)
20. Kinouchi, Y.: A Logical Model of Consciousness on an Autonomously Adaptive System. IJMC 1(2), 235–242 (2009)
21. Levine, J.: Materialism and qualia: the explanatory gap. Pacific Philosophical Quarterly 64, 354–361 (1983)
22. Manzotti, R., Tagliasco, V.: An Externalist Process-Oriented Framework for Artificial Consciousness. In: Proc. AI and Consciousness: Theoretical Foundations and Current Approaches, AAAI Fall Symposium 2007. AAAI Press, Menlo Park (2007)
23. Samsonovich, A.V.: Toward a Unified Catalog of Implemented Cognitive Architectures. In: Samsonovich, A.V., Johannsdottir, K.R., Chella, A., Goertzel, B. (eds.) Biologically Inspired Cognitive Architectures 2010. IOS Press, Amsterdam (2010)
24. Sanz, R., López, I., Bermejo-Alonso, J.: A Rationale and Vision for Machine Consciousness. In: Chella, A., Manzotti, R. (eds.) Artificial Consciousness. Imprint Academic, UK (2007)
25. Shanahan, M.: Embodiment and the Inner Life. Oxford University Press, Oxford (2010)
26. Sloman, A.: An Alternative to Working on Machine Consciousness. IJMC 2(1), 1–18 (2010)
27. Takeno, J., Inaba, K., Suzuki, T.: Experiments and examination of mirror image cognition using a small robot. In: Proc. 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2005), pp. 493–498 (2005)
28. Tononi, G.: Consciousness as Integrated Information: A provisional Manifesto. Biological Bulletin 215(3), 216–242 (2008)