DO ARTIFICIAL GENERAL INTELLIGENT SYSTEMS REALLY NEED TO BE CONSCIOUS?

J. Ignacio Serrano and M. Dolores del Castillo
Bioengineering Group, CAR - CSIC, Ctra. Campo Real km 0.200, La Poveda 28500, Arganda del Rey, Spain

Keywords: Artificial general intelligent systems, Autonomous systems, Machine consciousness.

Abstract: Consciousness has been studied for long time from heterogeneous perspectives and knowledge fields. In spite of the great numbers of debates and the huge amount of work, the matter is still full of questions and even enigmas at different levels. In addition to well known issues such as the evolutionary utility of consciousness, the existence (or not) of the "hard problem" and the definition itself (to mention just a few), consciousness has also been brought into the mind/body problem. This controversy has encouraged many researchers to tackle the simulation and implementation of consciousness, thus giving rise to the so-called Machine Consciousness (also Artificial Consciousness), which in turn motivates the inclusion of consciousness in Artificial General Intelligent (AGI) Systems. However, do an AGI system need consciousness in order to be (general) intelligent? This paper poses a humble reflection on this subject with the only aim of making the readers think about it before starting working.

1 INTRODUCTION

Consciousness has been the object of study of many research fields for decades (Seager, 2007). The abstraction level of such a concept, the philosophical implications and the unavailability of empirical measures have made the research as appealing as difficult (Damasio, 2010). As for many other phenomena in science, there is not yet a commonly accepted notion of what consciousness is. In 1874, Franz Brentano stated that"...there is no question of there being a commonly accepted, exclusive sense of the term (consciousness)... I prefer to use it as synonymous with 'mental phenomenon' or 'mental act' " (Brentano, 1874). A century afterwards, R. J. Joynt claimed that "Consciousness is like the Trinity; if it is explained so that you understand it, it hasn't been explained correctly" (Joynt, 1981). Although many other researchers have dared to finely delimit the boundaries of the term (Rosenthal, 2009)(Antony, 2002) , there is no common agreement and the object of what is wanted to know is blurred.

From this perspective, intuitively, it does not seem scientifically orthodox the attempt to physically (or empirically) explain a phenomena from, let us say, a partially arbitrary hypothesis. Nevertheless, is this feature the one that determine the human nature? Is there any conscious animal apart from the human beings? Unfortunately, the answer will depend on what is wanted to mean with being conscious.

In spite of the broad disagreement about the definition of consciousness (Velmans, 2009), everyone would accept that human beings are somehow conscious. Consequently, the design and construction of an artificial system that is intended to be generally intelligent (AGI), that is, to the same extent as any single human subject is, should include (artificial/machine) consciousness (Aleksander, 2009). However, is consciousness coupled with human-like intelligence? Is it an inherent requirement for the humans to carry out all the different tasks? Is consciousness mandatory for the exhibition of human behavior? Actually, there are other factors that seem to avoid the realization of truly human-like intelligent systems (McClelland, 2009).

The inverse case, out of the criticism, is the study of consciousness itself by computational/robotic tools (Artificial/Machine Consciousness) (Seth, 2009). In such a case, the research target is the conscious phenomenon and the artificial systems are the methodological tools (Taylor, 2010). Even the use of an AGI system for the study of consciousness is justified under the hypothesis that consciousness emerges from the immersion of the subject within a real environ-

Ignacio Serrano J. and Dolores del Castillo M..
DO ARTIFICIAL GENERAL INTELLIGENT SYSTEMS REALLY NEED TO BE CONSCIOUS?.
DOI: 10.5220/0003194906740676
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 674-676
ISBN: 978-989-8425-40-9

ment that requires adaptation, multi-tasking and interaction with other subjects. In such a case, an AGI system is a useful tool that provides the closest-to-ideal conditions for the study of consciousness. This latter approach, far from tangle, could effectively contribute to the development of the knowledge about consciousness.

In summary, the big question is whether it is really necessary to try to figure out a complex issue (such as consciousness) by means of a technological patch[1] just in order for a machine to seem to behave as a human being (in Turing terms) (Harnad, 2006). Next, some of the motivations for a positive answer to the question are exposed and refuted.

2 BECAUSE OF THE EVOLUTIONARY ADAPTIVE VALUE

One might consequently think that if the 'conscious' human has successfully survived and remained until nowadays, consciousness must have taken a critical part in such an achievement. Following this line, an AGI system should reasonably include a consciousness mechanism in order for the system to collect the intrinsic human nature accumulated through the evolutionary history. This way, the AGI system would be closer to the human referent and, in turn, to the final aim.

According to Harnad (Harnad, 2001), the consequences of adaptation are the consequences of the function. Therefore, a different feature (variation in Darwin's terms) that does not produce a different function is not a profitable feature for adaptation. Besides, there is no reason to deny that human beings could have reached the same level of adaptation by means of the same consciousless internal mechanism (Humphrey, 2006). That is, there is no functional difference between doing a task consciously or doing the same task without consciousness (imagine a plane in auto pilot or human pilot). Therefore, it might be enough for an AGI system to be provided with the capability to carry out the same functions that have made the human beings finely adapt and proliferate, regardless of consciousness.

3 BECAUSE OF THE HUMAN PERFORMANCE APPEARANCE

It is commonly intended for an AGI system to interact with and be immersed in the environment. This environment include, of course, other human subjects. Thus, in order for an AGI system to success it has to be accepted (or recognized) by the other subjects as one of them. Consciousness might seem the key for this question if human nature is supposed to be exhibited by conscious processes. However, consciousness is not required to behave as a human being. For instance, an artificial (consciousless) system that pases the Turing Test (Harnad, 2006) is considered to be human by other human judge. In addition, a human ('conscious') subject might behave 'unconsciously', as showed by the Chinese Room argument (Damper, 2004). In conclusion, behavior does not necessarily imply consciousness, and behavior (as well as physical appearance) is what make subjects suppose the human nature of other subjects. After this assignment, consciousness is attributed by default either actually present or not.

4 BECAUSE OF THE HUMAN SINGULARITY

Under the assumption that consciousness is the quality that distinguishes humans from other animals, an AGI must implement consciousness for reaching human-level intelligence. However, consciousness is not only present in human beings. Many studies have pointed out the existence of the notion of consciousness in mammals such as elephants, dolphins and many kinds of primates, mainly derived from the presence of shared brain structures that are thought to produce the conscious behavior both in humans and animals (Baars, 2005). Moreover, other simpler brains such as birds have also shown conscious processing (Beshkar, 2008). Therefore, the implementation of consciousness in an artificial system does not guarantee a human-level performance, but just a vertebrate-level performance.

[1]Note that this is not a criticism against technological solutions or methodologies.

5 BECAUSE OF THE GENERALITY AND UNIVERSALITY

Since an AGI system should be given with all the capabilities that the intelligent human beings actually own, and every subject would accept that humans are conscious, the AGI system must include such a feature. However, while most human subjects (except impaired or people with disorders) can accomplish the same tasks with similar performance (in the absence of training), consciousness is different and particular for each individual. Consciousness is not innate[2]. It emerges during development so it is affected by the context and the stimuli for each individual (Zelazo, 2003). Besides, consciousness is not universal, it differs among cultures and even civilizations (Earley, 2002). So, when an AGI systems incorporates a consciousness mechanism this should be a particular one different from other artificial systems. This mechanism should also evolve so that it can be modulated by context and culture. Consciousness is not therefore a general mechanism (in functional terms).

6 CONCLUSIONS

The problem of consciousness can be considered among the most controversial within the "Hard" Artificial Intelligence (AI). It implies a lot of considerations from different perspectives and knowledge fields (Damasio, 2010), to the extent that its own existence is even questioned. A lot of hard and rigorous work must be carefully carried out to get closer and closer to the solution of the problem. Rush and naïve approaches might introduce noise and blur the matter even more than it actually is. The position exposed in this paper claims for avoiding such attempts when there is no reasonable need for it. This might be the case of the implementation of consciousness mechanisms in AGI systems, which do not require consciousness as mandatory to achieve their goals, that is, to be generally intelligent and more human-like (Harnad, 2003)(McClelland, 2009).

REFERENCES

Aleksander, I. L. (2009). The potential impact of machine consciousness in science and engineering. *International Journal of Machine Consciousness*, 1(1):1–9.

Antony, M. V. (2002). Concepts of consciousness, kinds of consciousness, meanings of 'consciousness'. *Philosophical Studies*, 109(1):1–16.

Baars, B. J. (2005). Subjective experience is probably not limited to humans: The evidence from neurobiology and behavior. *Consciousness and Cognition*, 14(1):7–21.

Beshkar, M. (2008). Animal consciousness. *Journal of Consciousness Studies*, 15(3):5–33.

Brentano, F. (1874). *Psychology from an Empirical Standpoint*. International Library of Philosophy. Routledge, 2nd (1995) edition.

Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon, 1st edition.

Damper, R. I. (2004). The chinese room argument–dead but not yet buried. *Journal of Consciousness Studies*, 11(5-6):159–169.

Earley, J. E. (2002). The social evolution of consciousness. *Journal of Humanistic Psychology*, 42(1):107–132.

Harnad, S. (2001). *Turing Indistinguishability and the Blind Watchmaker*. John Benjamins, Amsterdam.

Harnad, S. (2003). Can a machine be conscious? how? *Journal of Consciousness Studies*, 10(4):67–75.

Harnad, S. (2006). *The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence*. Kluwer.

Humphrey, N. (2006). *Consciousness: The Achilles Heel of Darwinism? Thank God, Not Quite*. Vintage.

Joynt, R. J. (1981). Are two heads better than one? *Behavioral and Brain Sciences*, 1(4):108–109.

McClelland, J. (2009). Is a machine realization of truly human-like intelligence achievable? *Cognitive Computation*, 1(1):17–21.

Rosenthal, D. ((forthcoming) 2009). *Concepts and Definitions of Consciousness*. Elsevier.

Seager, W. E. (2007). *A Brief History of the Philosophical Problem of Consciousness*. Cambridge University Press.

Seth, A. K. (2009). The strength of weak artificial consciousness. *International Journal of Machine Consciousness*, 1(1):71–82.

Taylor, J. G. (2010). On artificial brains. *Neurocomputing*, page (In press).

Velmans, M. (2009). How to define consciousness: And how not to define consciousness. *Journal of Consciousness Studies*, 5(18):139–158.

Zelazo, P. D. (2003). The development of conscious control in childhood. *Trends in Cognitive Sciences*, 1(8):12–17.

[2]It refers to complete and mature consciousness, not to its the capacity of development