Artificial Intelligence in Medicine (2008) **44**, 79—82

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

GUEST EDITORIAL

Artificial consciousness: Theoretical and practical issues

Morpheus: at some point in the early twenty-first century all of mankind was united in celebration. We marveled at our own magnificence as we gave birth to AI

Neo: AI - you mean artificial intelligence?

Morpheus: A singular consciousness that spawned an entire race of machines. (from The Matrix script, 1999)

Introduction

When the digital computer was invented more than half a century ago, many felt that the essence of thinking, had been find [1]. All of a sudden, it seemed possible to simulate and reproduce thinking, problem solving, even natural language—namely, the mind. Yet, consciousness has such was purposefully left out. Subsequently, the field spawned several partially overlapping yet connected fields as robotics, control systems, data mining, logistics, speech recognition, facial recognition, and many others, but something was still missing. Besides, artificial intelligent agents fall short in many areas with respect to biological agents. Could consciousness be the missing ingredient?

In recent years, several researchers ventured the hypothesis of designing and implementing a model for artificial consciousness [2—4]. On one hand there is hope of being able to design a conscious machine [5,6], on the other hand such models could be helpful for understanding human consciousness [7,8]. According to Aleksander, "machine consciousness refers to attempts by those who design and analyse informational machines to apply their methods to various ways of understanding consciousness and to examine the possible role of consciousness in informational machines" [9].

Since the beginning of the century, there has been an upsurge of interest in artificial consciousness as shown by several dedicated workshops [10—12], several dedicated volumes and special journal issues [4,13,14]. It is timely to check the status of theoretical and practical issues concerning artificial consciousness. In the AI there has been a considerable interest towards consciousness. It is now customary to read papers that suggest to design conscious machines as a way to achieve a better understanding of consciousness: "contemplating how to build such a machine will inevitably shed light on scientists' understanding of our own consciousness" [15].

In this special issue of *Artificial Intelligence in Medicine*, we focus on theoretical and practical aspects of artificial consciousness. It has to be admitted that the field is still rather fuzzy and lacking an overarching sets of accepted assumptions. The views expressed by different scholars vary widely and it is still unclear whether there is any true possibility to reproduce consciousness in a machine [16]. It is also to be estimated whether artificial consciousness is really adding something to the field of AI [17]. Some sceptics pointed out that "prefacing rather standard engineering processes with loaded philosophical terms does not suffice to bestow upon the artefacts the properties associated with these terms, and most roboticists will simply ignore these terms anyway" [14].

Artificial consciousness: a brief history

Since the heydays of cybernetics, the goal of many scientists had been the understanding of the elusive phenomenon of the mind. Yet, due to its proble-

0933-3657/$ — see front matter
doi:10.1016/j.artmed.2008.08.001

matic nature, it was considered wiser to focus on apparently more tractable aspects of the mind such as behaviour, problem solving, syntactical manipulation, and sensory-motor coordination.

The first official mention of the name "artificial consciousness" is to be traced to a book by Nemes, *Kibernetikai gépek* [18], later translated in English [19]. A part of that paragraph is given below [19]. The very idea of making a conscious machine was seldom mentioned because the very notion of consciousness was considered highly suspicious. The reasons for this long lasting scientific banishment are articulated at length in many excellent books [20,21]. However, since the beginning of the last decade of the last century, a new scientific interest for consciousness arose [22—24] leading to current widespread approaches in neuroscience [25—27]. Such increased acceptance of the topic allowed many researches in robotics and AI to reconsider the possibility of modelling and implementing a conscious machine.

Aleksander was one of the first serious scholars that explicitly mentioned the field of artificial consciousness [3] although similar claims and questions were put forward in other contexts. In 2001, at the Cold Spring Harbour Laboratories sponsored by the Swartz Foundation, it was held a seminar on the topic whether machines could ever be conscious. It was concluded that "there is no known law of nature that forbids the existence of subjective feelings in artefacts designed or evolved by humans".

The field slowly gained new contributions and more momentum. Following a robotic experiment [28], the editorialist Adami explicitly addressed the issue of artificial consciousness in a seminal paper on *Science* [27] suggesting that "robots [...] could play an interesting role in our quest to understand the nature of consciousness". Moreover, in 2007 the AAAI accepted to host a symposium on the issue of "AI and Consciousness: Theoretical Foundations and Current Approaches" [10].

As Miller wrote "until fairly recently, tackling the subject of consciousness was a dubious career move for any scientist without tenure (and perhaps a Nobel Prize already in the bag). Fortunately, more young researchers are now joining the fray" [24]. Will the same take place for artificial consciousness?

Excellent overviews of the field as well as more detailed narration of its recent development can also be found elsewhere [9,29]

Overview of the special issue

Since artificial consciousness is still seeking a methodological and theoretical accepted framework, each of the manuscripts is balanced between a more general introduction and some more implantation details. This is to be expected in a field whose foundations are hindered by one of the most difficult problem of science—namely, the 'hard' problem.

Clowes and Seth [30] stresses the deep methodological issues underlying artificial consciousness. The study of consciousness itself is still in a preparadigmatic stage whereas a commonly accepted theoretical and conceptual framework is still missing. Hence they dwell on the distinction between strong and weak artificial consciousness. They analyse several classic approaches like Damasio's dynamic core, Baars' global workspace, and Aleksander's axiomatic approach to conclude that it is premature to make any claims about the true nature of consciousness. Their hope is the progressive development of more advanced models will tend on the instantiation of systems that might be considered conscious.

Manzotti and Tagliasco tackle heads down on the hard problem of consciousness [31]. After sketching the origin of artificial consciousness and showing the AI roots of the field, they stress the importance of explaining the phenomenological aspect of consciousness. How can a physical system feel something? According to them, artificial consciousness has to provide an answer to this fundamental question. A possible solution could be found in externalist approaches to the mind, since such approaches do not rely on problematic notion of emergence or of internal representations, but rather single out conscious experience as some kind of causal interaction between agents and their environment.

In his passionate paper [32], Chrisley argues for a philosophical foundation of artificial consciousness. The widespread apprehensions about many issues of consciousness are likely to be misplaced. They are based on misunderstandings of AI, and a lack of awareness of the possible roles AI might play in accounting for or reproducing consciousness. Moreover AI has always tried to emulate mental activities and thus the separation between AI and artificial consciousness has been overestimated. Eventually, he reviews several classic objections to artificial consciousness in the light of three key concepts in his work: interactive empiricism, synthetic phenomenology, and ontologically conservative heterophenomenology.

Buttazzo [33] describes pros and cons of current approaches to artificial consciousness by asking several hazardous questions each pinpointing a key issue. He is positive about the chances of success. After all, human beings are biological