Consciousness and Cognition 45 (2016) 210–225

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Review article

Artificial consciousness and the consciousness-attention dissociation

Harry Haroutioun Haladjian [a,*], Carlos Montemayor [b]

[a] Laboratoire Psychologie de la Perception, CNRS (UMR 8242), Université Paris Descartes, Centre Biomédical des Saints-Pères, 45 rue des Saints-Pères, 75006 Paris, France

[b] San Francisco State University, Philosophy Department, 1600 Holloway Avenue, San Francisco, CA 94132 USA

ARTICLE INFO

Article history:
Received 6 July 2016
Accepted 12 August 2016

Keywords:
Artificial intelligence
Artificial consciousness
Consciousness
Visual attention
Phenomenology
Emotions
Empathy

ABSTRACT

Artificial Intelligence is at a turning point, with a substantial increase in projects aiming to implement sophisticated forms of human intelligence in machines. This research attempts to model specific forms of intelligence through brute-force search heuristics and also reproduce features of human perception and cognition, including emotions. Such goals have implications for artificial consciousness, with some arguing that it will be achievable once we overcome short-term engineering challenges. We believe, however, that phenomenal consciousness cannot be implemented in machines. This becomes clear when considering emotions and examining the dissociation between consciousness and attention in humans. While we may be able to program ethical behavior based on rules and machine learning, we will never be able to reproduce emotions or empathy by programming such control systems—these will be merely simulations. Arguments in favor of this claim include considerations about evolution, the neuropsychological aspects of emotions, and the dissociation between attention and consciousness found in humans. Ultimately, we are far from achieving artificial consciousness.

Contents

* Corresponding author.
E-mail addresses: haroutioun@gmail.com (H.H. Haladjian), cmontema@sfsu.edu (C. Montemayor).
URLs: http://www.haroutioun.com (H.H. Haladjian), http://www.carlosmontemayor.org (C. Montemayor).

http://dx.doi.org/10.1016/j.concog.2016.08.011
1053-8100/

1. Introduction

One of the most compelling topics currently debated is how it may be possible to develop consciousness in machines. While related questions have been discussed academically in the cognitive sciences for some time now, the idea of artificial consciousness has received more attention in popular culture recently. There is a growing number of articles in magazines and newspapers that discuss related advances in Artificial Intelligence (AI), from self-driving cars to the "internet of things" where common household objects can be intelligently connected through a centralized control system. Most recently, Google's DeepMind group developed an artificial agent capable of winning the game of Go against humans, which is considered to be a huge accomplishment in AI since it goes beyond brute-force search heuristics and uses deep learning models (Silver et al., 2016). This advance promises more impressive innovations in the future.

Along with these advancements is a growing fear that we may be creating intelligent systems that will harm us (Rinesi, 2015). This topic has been addressed in many settings, from international conferences to popular books (e.g., Bostrom, 2014; Brooks, Gupta, McAfee, & Thompson, 2015). Films and television shows, such as *Ex Machina*, *Her*, and *Battlestar Galactica*, present scenarios where AI systems go rogue and threaten humans. While these fictional accounts remain unachievable with today's technology, they are beginning to feel more and more possible given how fast computers are "evolving".

The increase of these discussions in the mainstream media is telling, and one could say that Artificial Intelligence is truly at a turning point. Some think that the so-called 'singularity' (the moment in which AI surpasses human intelligence) is near. Others say there is now a Cambrian explosion in robotics (Pratt, 2015). Indeed, there is a surge in AI research across the board, looking for breakthroughs to model not only specific forms of intelligence through brute-force search heuristics but by truly reproducing human intelligent features, including the capacity for emotional intelligence and learning. Graziano (2015), for example, has recently claimed that artificial consciousness may simply be an engineering problem—once we overcome some technical challenges we will be able to see consciousness in AI. Even without accomplishing this lofty goal of machine sentience, it still is easy to see many examples of human-like rational intelligence implemented in computers with the aim of completing tasks more efficiently.

What we will argue, however, is that *phenomenal consciousness*, which is associated with the first-person perspective and subjectivity, cannot be reproduced in machines, especially in relation to emotions. This presents a serious challenge to AI's recent ambitions because of the deep relation between emotion and cognition in human intelligence, particularly social intelligence. While we may be able to program AI with aspects of human conscious cognitive abilities, such as forms of ethical reasoning (e.g., "do not cause bodily harm", "do not steal", "do not deceive"), we will not be able to create actual emotions by programming certain monitoring and control systems—at best, these will always be *merely simulations*. Since human moral reasoning is based on emotional intelligence and empathy, this is a substantial obstacle to AI that has not been discussed thoroughly.

Before proceeding, however, it is crucial to distinguish the present criticism of AI from a very influential one made by Searle (Searle, 1980, 1998). Searle has famously criticized AI because of its incapacity to account for *intentionality* (i.e., the feature of the mental states that makes them about something, essentially relating them to semantic contents), which he takes to be exclusively a characteristic of conscious beings. He also argues that a consequence of this criticism is that phenomenal consciousness (i.e., what it is like to have an experience) is necessarily a biological phenomenon. Searle, therefore, takes the limitations of AI to be principled ones that will not change, regardless of how much scientific progress there might be.

Critics have argued that the intuition that only biological beings can have intentional minds may be defeated (e.g., see Block, 1995a) and that cyborg systems or an adequate account of how the brain computes information could refute the Chinese room thought experiment (Churchland & Churchland, 1990; Pylyshyn, 1980). These criticisms have merit, and we largely agree with them, but only with respect to the kind of consciousness that Block (1995b) calls 'access consciousness'. Thus, we believe there is a very important ambiguity in this debate. While we agree with Searle that phenomenal consciousness is essentially a biological process and that AI is severely limited with respect to simulating it, we agree with his critics when they claim that AI may be capable of simulating and truly achieving *access consciousness*. This is why the consciousness and attention dissociation is crucial for our purposes, because it states that attention is essentially related to accessing information (see Montemayor & Haladjian, 2015).

Our criticism of AI, therefore, is more nuanced than Searle's in three important respects. First, we limit our criticism exclusively to the type of consciousness that is characteristic of feelings and emotions, independently of how they are related to semantic contents or conceptual categories (i.e., phenomenal consciousness). Second, the limitations of AI with respect to simulating phenomenal consciousness will be independent of considerations about understanding the meaning of sentences. The limitations of AI that we outline will extend to other species, which do not manifest the capacity for language but which very likely have phenomenal consciousness. Thus, our criticism of AI is more truly based on biological considerations than Searle's. Third, and quite importantly, we grant that AI may simulate intelligence, rationality, and linguistic behavior success-