Biologically Inspired Cognitive Architectures (2016) xxx, xxx–xxx

Available at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bica

RESEARCH ARTICLE

What are the computational correlates of consciousness?

James A. Reggia [a,b,*], Garrett Katz [a], Di-Wei Huang [a]

[a] Department of Computer Science, University of Maryland, United States
[b] UMIACS, University of Maryland, United States

Received 5 June 2016; received in revised form 28 July 2016; accepted 28 July 2016

KEYWORDS
Cognitive phenomenology;
Computational explanatory gap;
Phenomenal consciousness;
Machine consciousness;
Artificial consciousness

Abstract
Cognitive phenomenology refers to the idea that our subjective experiences include deliberative thought processes and high-level cognition. The recent ascendance of cognitive phenomenology in philosophy has important implications for biologically-inspired cognitive architectures and the role that these models can play in understanding the fundamental nature of consciousness. To the extent that cognitive phenomenology occurs, it provides a new route to a deeper understanding of consciousness via neurocomputational studies of cognition. This route involves identifying computational correlates of consciousness in neurocomputational models of high-level cognitive functions that are associated with subjective mental states. Here we develop this idea and compile a summary of potential neurocomputational correlates of consciousness that have been proposed/recognized during the last several years based on biologically-inspired cognitive architectures. We conclude that the identification and study of computational correlates of consciousness will lead to a better understanding of phenomenal consciousness, a framework for creating a conscious machine, and a better understanding of the mind-brain problem in general.

Introduction

The term *phenomenal consciousness* refers to the subjective qualities (qualia) of sensory phenomena, such as the redness of an object or the pain from a skinned knee, that we experience when awake (Block, 1995). Phenomenal consciousness is very poorly understood at present. Here we describe a framework for biologically-inspired computa-

* Corresponding author at: Department of Computer Science, University of Maryland, United States.
E-mail addresses: reggia@cs.umd.edu (J.A. Reggia), garrett.katz@gmail.com (G. Katz), dwh@cs.umd.edu (D.-W. Huang).

http://dx.doi.org/10.1016/j.bica.2016.07.009
2212-683X/

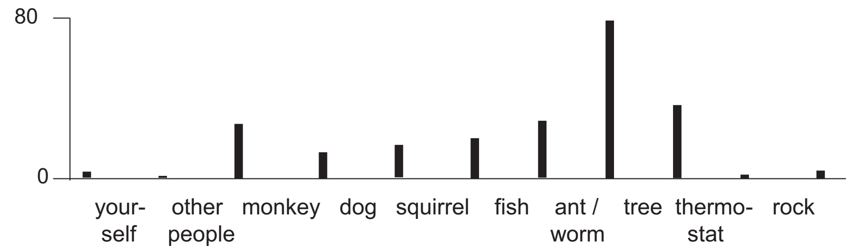

Fig. 1 Number of responses at each location when 228 university students were asked to indicate the dividing line between conscious and non-conscious entities on a horizontal scale (Reggia et al., 2015). More than half of the responses indicated that all animal life is conscious. There was no statistically significant difference in how subjects responded based on whether they identified themselves as dualists, idealists, materialists, or otherwise. Such results are consistent with those found in other related studies (Arico, Fiala, Goldberg, & Nichols, 2011; McDermott, 2007).

tional investigations that we believe could ultimately shed substantial light on the fundamental nature of phenomenal consciousness, and perhaps even lead to a conscious machine experiencing qualia and subjective experiences. Such a claim may initially seem improbable given current beliefs about the nature of consciousness and the many controversies that surround the field of consciousness studies. For example, there appears to be fairly widespread agreement that existing computers are *not* phenomenally conscious. If one asks university students whether or not they believe that contemporary electronic computers are conscious, 82% say no, 3% say yes, and 15% are undecided (Reggia, Huang, & Katz, 2015). This result is particularly striking in that people in general are usually quite generous in attributing consciousness to others, frequently identifying animals, including ants and worms, and in some cases even plants,[1] as being conscious (see Fig. 1). Some philosophers would go even further, arguing that phenomenal machine consciousness will *never* be possible for a variety of reasons: the non-organic nature of machines (Schlagel, 1999), it would imply panpsychism (Bishop, 2009), the absence of a formal definition of consciousness (Bringsjord, 2007), or the insufficiency of computation to underpin consciousness (Manzotti, 2012; Piper, 2012). More generally, it has been argued that the objective methods of science cannot shed light on phenomenal consciousness due to its subjective nature (McGinn, 2004), making computational investigations irrelevant.

Fortunately, such negative sentiments have not completely deterred the study of artificial consciousness over the last two decades, and this work has led to substantial progress and informative results (Reggia, 2013). For example, neurocomputational models have been studied that support global workspace theories of consciousness (Dehaene, Kerszberg, & Changeux, 1998), that allow expectation-driven robots to effectively pass the well-known mirror test used to identify self-recognition in animals (Takeno, 2013), and that match behavioral data from human blindsight subjects during post-decision wagering tasks (Pasquali, Timmermans, & Cleeremans, 2010). However, this progress in artificial consciousness has mainly involved computationally simulating neural/behavioral/cognitive aspects of consciousness, much as is done in using computers to simulate other natural processes (e.g., models of weather/climate). There is nothing particularly mysterious about such work: Just as one would not expect that a computer used to simulate a nuclear reactor would become radioactive, one would not expect that a computer used to model some aspect of conscious information processing would become phenomenally conscious. There is no claim that phenomenal consciousness is actually present in a machine in this type of work. While computational modeling has become a widely accepted tool for studying consciousness, there is currently no existing computational approach to artificial consciousness that has yet presented a compelling demonstration or design of phenomenal consciousness in a machine, or even clear evidence that it will eventually be possible.

In the following we describe our framework for computational investigations that we believe will ultimately provide insight into the nature of phenomenal consciousness, and perhaps even lead to a conscious machine experiencing qualia and subjective experiences. We begin with brief background information about *cognitive phenomenology*, which asserts that our subjective experiences are not restricted to just traditional qualia but also encompass deliberative thought processes and high-level cognition (Bayne & Montague, 2011b). We then ask, from a purely computational/engineering viewpoint: What is the main practical barrier to further progress on creating phenomenal machine consciousness? The answer suggested here is that it is a *computational explanatory gap*, our current lack of understanding of how high-level cognitive computations can be captured in low-level neural computations (Reggia, Monner, & Sylvester, 2014). The significance of this gap derives from the claims of cognitive phenomenology, in that bridging it may be a critical step in gaining a better understanding of phenomenal consciousness. In other words, we are suggesting that a key implication of cognitive phenomenology is that it provides a new route to a deeper understanding of consciousness via computational studies. This route would focus on identifying possible *computa-*

[1] While the notion that plant life is conscious may strike some as surprising, there is growing scientific evidence that plants do ecologically-relevant information processing and communicate with one another via chemical signals, e.g., some plant species can "hear", discriminate and respond appropriately to sounds that indicate a threat (Appel & Cocroft, 2014). This has contributed to a substantial literature, online discussions, and media coverage about plant consciousness and intelligence that may have influenced such views (Cvrckova, Lipavska, & Zarsky, 2009; Marder, 2012; Nagel, 1997).

Please cite this article in press as: Reggia, JA et al., What are the computational correlates of consciousness? ..., *Biologically Inspired Cognitive Architectures* (2016), http://dx.doi.org/10.1016/j.bica.2016.07.009