Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 11 (2010) 343–356
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Extended cognition and the priority of cognitive systems

Action editor: Leslie Marsh

Robert D. Rupert

Department of Philosophy, University of Colorado at Boulder, Boulder, CO, USA

Received 2 March 2010; accepted 2 April 2010
Available online 23 July 2010

Abstract

This essay begins by addressing the role of the so-called Parity Principle in arguments for extended cognition. It is concluded that the Parity Principle does not, by itself, demarcate cognition and that another mark of the cognitive must be sought. The second section of the paper advances two arguments against the extended view of cognition, one of which – the conservatism-or-simplicity argument – appeals to principles of theory selection, and the other of which – the argument from demarcation – draws on a systems-based theory of cognition. The final section contests the claim, made by Andy Clark, that empirical work done by Wayne Gray and colleagues supports the extended view.

Keywords: Extended cognition; Extended mind; Distributed cognition; Situated cognition; Embedded cognition; Parity Principle

1. Introduction

According to the hypothesis of extended cognition (HEC), human cognitive processing extends beyond the boundary of the human organism.[1] As I understand HEC, it is a claim in the philosophy of cognitive science; the extended view is meant to capture a central fact about human beings' core cognitive faculties and, in doing so, reshape the study of human cognition. So understood, HEC cannot be vindicated by oddball cases, science-fiction thought experiments, or the development of assistive technologies. Such results—even if they constitute individual cases of extended cognition or entail the conceptual possibility of it—fall far short of the revolution promised by HEC's proponents.[2] These results are germane only if they help to reveal what actually are (or should be) the properties of central causal-explanatory interest in cognitive science, that is, the properties appearing in cognitive science's best account of everyday, paradigmatically cognitive capacities—language-acquisition, inference, categorization, planning, theory-construction, perception, and reading—and the associated behavior.

Many criticisms of HEC have made their way into print (Adams & Aizawa, 2001, 2007; Block, 2005; Butler, 1998; Grush, 2003; Rupert, 2004, 2008, 2009, 2010; Segal, 1997; Sterelny, 2004; Weiskopf, 2008; Wilson, 2002). The present essay adds to this critical literature but does so in a constructive way: by presenting an independently motivated account of cognition. Given the empirical facts about human beings, this conception of cognition likely countenances very little of

E-mail address: robert.rupert@colorado.edu

[1] An enormous amount has been written in support of HEC, from a philosophical standpoint: see Clark (1997, 2001, 2003, 2007, 2008), Clark and Chalmers (1998), Dennett (1996), Haugeland (1995), Hurley (1998), Millikan (1993, essays 7–9), Rowlands (1999), van Gelder (1995), Wheeler (2005), and Wilson (2004). For a sampling of the empirical work that has inspired HEC, see Brooks (1986), Dawkins (1982), Donald (1991), Hutchins (1995), and Thelen and Smith (1994). Also see the references provided in note 2.

[2] The language of paradigm shift, revolution, and the reorientation of cognitive science can be found in Brooks (1999), Gibbs (2006), Haugeland (1995), Lakoff and Johnson (1999), Spivey, Richardson, and Zednik (2010), Thelen and Smith (1994), van Gelder (1991, 1995), Varela, Thompson, and Rosch (1991), and Wheeler (2005).

1389-0417/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2010.04.002

the extended sort—far too little to support HEC. Here is the view in a nutshell: x is a cognitive state[3] if and only if x is the state of a component of a cognitive system, where a cognitive system is characterized as an integrated set of mechanisms that contribute distinctively to the production of cognitive science's *explananda*. This emphasis on integration reflects a basic fact about cognitive science: that the fundamental theoretical construct of virtually all successful cognitive science – whether computationalist, connectionist, or dynamicist—is that of a persisting architecture interacting with an ever-changing cast of external materials to produce intelligent behavior. If there is any theoretically interesting divide between what is distinctively cognitive and what merely causally contributes to intelligent behavior, it is to be found in the persisting, integrated nature of cognitive architectures.

Not all of the arguments offered herein depend on the details of this positive conception of cognition; nevertheless, the idea of a relatively integrated, persisting system plays a central role throughout the essay, even in the more critical sections. For example, the advocate for HEC must make room for architectures and their distinctive contributions to the production of intelligent forms of behavior, and she can help herself to these; but in doing so, she renders her view otiose. She recapitulates, within an extended framework, all of the distinctions, constructs, and steps found in orthodox explanations, now gratuitously relabeled or reconceptualized.

The paper proceeds as follows. Section 2 criticizes an argument for HEC based on what is known as the Parity Principle; this discussion is of independent interest, but it also helps to motivate my emphasis on cognitive systems. Section 3 presents a pair of arguments against HEC and develops the positive conception of cognition outlined above. The final section addresses empirical work that might appear to undercut the arguments of Section 3.

2. Cognitive systems and the Parity Principle

In "The Extended Mind," Andy Clark and David Chalmers (C&C, hereafter) make the following claim:

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process (1998, p. 8).

This has come to be known as the Parity Principle (PP), and C&C's appeal to it in support of HEC has been at the center of much debate.[4] For that reason alone, it is worth asking whether PP can ground an argument for HEC. In the remainder of this section, I argue that PP provides no sound basis for HEC – and for reasons that frame much of the discussion to come.

The Parity Principle expresses (at least) the correct intuition that cognition is cognition, wherever it occurs. On a weak reading of PP, then, it is merely about location: it warns us against thinking that being in the "wrong" location precludes something's being cognitive. It will be instructive to consider a problem with PP, read in this weak way, before taking up a more sophisticated interpretation of it. In what follows, I give two examples of cases that have nothing directly to do with cognition; these examples are meant to provide only counterexamples to a general reading of the PP, one that implies that, in general, location is irrelevant to the instantiation of theoretical properties.

For comparison only (this has nothing directly to do with cognition), consider a lone neuron in preparation. It may be that, were the neuron in my head, it would have the property *being a contributor to large-scale neural processing*, and we would recognize it as such. Nevertheless, in the dish, it has no such property. As an organ, the brain is a physically and functionally integrated system, and something can have the capacity to be part of the neural system without actually being part of it. The antecedent of this instance of a neural Parity Principle (of the weak, merely location-oriented sort) is plausibly true: if the neuron were in my skull, we would regard it as a participant in large-scale neural processing. Absent further information concerning how the neuron might have gotten into my skull, the most plausible chain of events is one in which the neuron has been physically and functionally integrated into my brain. Why else would it be there? At the same time, the consequent is false: the lone neuron in preparation is not part of my brain.

This phenomenon – the non-location-neutrality of theoretical properties – is not limited to neural contexts. For example, the act of firing a gun can instantiate military properties, or not, depending on the social and political context. Whether a given shooting has military properties depends on whether it occurs as part of certain kind of social-political system: a war.

Thus, a generalized location-oriented PP should be rejected outright. It is simply not true that, in general, a thing recognized as having property P were it in one location therefore has P regardless of its location. Location is irrelevant *only if P is the sort of property that survives a change in location*. Lots of properties, however, do not, as a general rule, survive such changes. The location-oriented PP is thus deeply uninformative. It warns us not to be biased by unexpected location, but this is sound advice only if location makes no difference to something's cognitive status; the location makes no difference, though, only if change in location is not correlated with change in the status of the item or process in question. Thus, when applying the location-oriented PP in a given domain, we can have confidence in PP's verdict only if we already know

[3] I move freely between talk of states and processes. I take the processes in question to be causally connected series of states, and because of this deep connection, everything I mean to say about states can be translated, without loss of argumentative force, into talk of processes, and *vice versa*.

[4] It has also been called the "parity claim," the "parity thesis," and the "fair-treatment principle." For discussion, see Adams and Aizawa (2001, 2007), Clark (2005, 2007, 2008, 2010), Levy (2007), Menary (2006), Sprevak (2009), Rowlands (2009), Rupert (2009), and Wilson and Clark (2008).