New Ideas in Psychology 31 (2013) 151–165

Contents lists available at SciVerse ScienceDirect

New Ideas in Psychology

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

Setting the bar for cognitive agency: Or, how minimally autonomous can an autonomous agent be?

Itay Shani*

Department of Philosophy, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, 130-701 Seoul, South Korea

Keywords:
Autonomous agents
Embodied cognition
Minimal autonomy
The adaptive-behavior approach
Recursive self-reliance
The systemic-constructivist approach

A B S T R A C T

Embodied cognition comes in many variants. Yet, in most if not all of these variants the notion of agency, or more precisely autonomous agency, occupies center stage. However, whereas in its original context of application autonomy applies strictly to the human sphere, cognitive theory must needs generalize this concept so as to render it applicable to a much wider range of entities and processes. Theorists of embodied cognition must therefore strive to articulate a valid sense of *minimal autonomy* applicable to animals and, arguably, to artificial agents as well. In this paper I discuss two major attempts to articulate a conception of minimal autonomy which I describe, respectively, as the *adaptive-behavior approach* (ABA) and the *systemic-constructivist approach* (SCA). The differences between these two leading views on minimal autonomy reveal fundamental disagreements not only with respect to the nature of autonomous agency but also with respect to embodiment and the relevance of biological life to the making of mental life.

1. Introduction

> "Only living things have needs and act on needs". Hans Jonas, *the Phenomenon of Life*

Embodied cognition is a centripetal force in contemporary cognitive science around which a plurality of theories and research agendas assemble giving rise to myriad interpretations, concerns, and emphases which often differ from each other considerably (see, e.g., Borghi & Pecher, 2011; Goldman & de Vignemont, 2009; Wilson, 2002). Yet, despite the multiplicity of variants falling under the banner of embodied cognition and the lack of any clear-cut definition of the subject, I think it is not too difficult to point to the core idea around which everything else revolves. It is what we may call the *embodiment thesis*:

Embodiment: The body plays a central role in determining the character and qualities of one's mental life. Mental processes are constrained, shaped, and illuminated in significant ways by the details of one's embodiment.

Whether one believes that the body plays a constitutive role in fashioning thought (Gallese & Lakoff, 2005) or merely a causal-facilitative role (Rupert, 2004); whether one believes that the contribution of the body to the making of mental life is irreplaceable (Dempsey and Shani, in press), or maintains that it might be countered, or canceled out, by other contributing factors (Clark, 2008); whether one's focus lies on perception (Noë, 2004), or concepts (Lakoff & Johnson, 1999), or language (Glenberg, Havas, Becker, & Rinck, 2005), or memory (Yang, Gallo, & Beilock, 2009), or emotions (Damasio, 1999), or intentional action (Jeannerod, 2001); whether one affirms the reality of mental representations and their explanatory significance (Pezzulo, 2011) or is prone to representational skepticism (Chemero, 2009); in each and every one of these instances one cannot be considered an advocate of embodied cognition *without* upholding *Embodiment*.

* Tel.: +82 70 44056967; cell: +82 10 22056967.
E-mail address: ishani479@hotmail.com.

0732-118X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.newideapsych.2012.09.003

Sure enough, that *Embodiment* is presupposed by so many views which are otherwise distinct from each other in important ways attests to its highly general and somewhat inexact character yet this fact by no means implies that the thesis is vacuous or lacking heuristic power. Clearly, it serves to mark a breaking point from a tradition which viewed mental processing as an abstract affair in which formalized symbol structures drive cogitations along inferential paths which have little, if anything, to do with the specifics of the bodies in which they are lodged, or through which they are executed.

Two other theses are strongly connected to Embodiment and are almost always evoked in connection with it. They are the theses of *action-centeredness* and of *situatedness* (cf. Anderson, 2003):

Action-centeredness: Cognitive agents are active beings geared-up for skillful adaptive manipulations of their respective environments. Cognitive abilities and operations do not merely serve this basic practical orientation but are also, to a significant degree, its products. That is, the imperative to act shapes the nature of cognition, and may even constitute its focal point.

Situatedness: Cognition reflects the fact that cognitive agents are situated in real world environments. Cognitive processes are enacted from a particular perspective, embedded in a particular context, respond to particular circumstances, and are structured, and often greatly facilitated, by the particulars of the context within which they are embedded.

The affiliation between Action-centeredness and Embodiment is evinced in the fact that the contribution of the body to the structuring of one's mental life is effected through the actions and internal activities it enables, sustains, motivates, constrains, or inhibits. In this sense, embodiment implies a shift toward greater emphasis on action. In turn, the notion of an embodied cognizer engaged in adaptive interactions presupposes, as a dual-component, an environmental context encountered from a specific viewpoint, under specific circumstances, posing specific challenges and opportunities, etc. (see Turvey, Shaw, Reed, & Mace, 1981). In short, the route from Embodiment to Situatedness seems equally robust.[1] It is therefore not too steep to suggest that Action-centeredness, and Situatedness complement Embodiment to form a conceptual triad which constitutes the kernel of the embodied cognition framework.[2]

When looked through such lenses, one fact about embodied cognition which stands out is that it reinstates *agency* into center stage. It is agents which are embodied, active, and situated, a fact that is duly reflected in the literature. There is ample talk about agents in artificial intelligence and robotics (e.g., Gigliotta & Nolfi, 2012), enacted cognition (Varela, Thompson, & Rosch, 1991), ecological psychology (Gibson & Pick, 2000), and the study of the somatic and neural bases of cognition (Garbarini & Adenzato, 2004), self-knowledge (De Vignemont & Fourneret, 2004), the problem of other minds (Colle, Beccio, & Bara, 2008), and so on. Strictly speaking, however, the concept of agency is too broad to be of much help. An agent, in the widest sense of the word, is any force or substance that causes a change. As such, being an agent bears no particular connection to cognition as is evident from the fact that we speak of chemical agents, infectious agents, and so on. An overflowing river, or an advancing crack on the wall, may be agents of destruction yet their agency has nothing to do with psychological agency. What sets psychological agents apart from such simple causal agents is that they are autonomous, a fact which is often recognized explicitly (as, for example, in the literature on autonomous agent robotics) but which must, at any rate, be born in mind whenever we speak of agents in a psychological sense.

Clearly, then, it is not agency as such which is central to the embodied cognition framework but rather *autonomous agency*. Hence, it is on the adjective 'autonomous' which we must focus if we wish to understand the precise sense in which agency is crucial for the embodied cognition project. When thinking of autonomy in this context, one problem which immediately emerges is that the concept has its roots in *human* personal and collective behavior while its application in cognitive theories of agency is much broader. Clearly, some features which are deemed essential for human-level manifestations of autonomy – in particular those which have to do with complex forms of rational and moral competence – cannot possibly be considered necessary for the manifestation of autonomy at the sub-human realm of animals and machines. As a result, advocates of embodied cognition must strike a fine balance between two opposing constraints. On the one hand, they must develop a notion of *minimal* autonomy – one which is broad enough to accommodate the whole gamut of intelligent creatures, and not merely humans. On the other hand, minimal autonomy must still be *autonomy enough*, that is, it must retain non-trivial resemblance to our default conception of autonomy which, to repeat, is quintessentially human. I discuss this challenge in more detail in Section 2 below.

Negotiating a coherent conception of minimal autonomy is therefore a desideratum of contemporary embodied cognitive science. Yet, apart from the sheer challenge of identifying a robust sense of minimal autonomy the significance of this task has also to do with the fact that different *solutions* to the problem of minimal autonomy have far reaching ramifications for our understanding of the nature of cognition, for the methodologies of the cognitive sciences, and for the explanatory aspirations and implementation strategies of artificial intelligence. Naturally, there are far more allusions to autonomy in the literature than thorough analyses of the concept. Yet, as if to simplify

[1] It is sometimes argued that situatedness implies reliance on an *online* streaming of sensory-motor flow so that off-line cognitive activities such as daydreaming or memorizing are interpreted as un-situated problem cases (see, e.g., Wilson, 2002). My own understanding of situatedness is more relaxed: it emphasizes the perspectival and contextualized character of cognition without making the stronger claim that this implies constant coupling to online sensory-motor processing. This more relaxed interpretation is in line with the general expositions of the meaning and roots of situated cognition offered by Clancey (2009) and Gallagher (2009).

[2] Notice that, just as in the case of Embodiment, a commitment to Action-centeredness and to Situatedness can be shared by researchers which disagree on many other related issues such as the status of representation, the exact manner in which the body influences cognition, or the ambitious claims of extended cognition.