Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 49 (2018) 157–177

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Computational model of enactive visuospatial mental imagery using saccadic perceptual actions

Action editor: Khan Iftekharuddin

Jan Jug[a,1], Tine Kolenik[a,1], André Ofner[b,1], Igor Farkaš[c,*]

[a] University of Ljubljana, Slovenia
[b] University of Vienna, Austria
[c] Comenius University in Bratislava, Slovakia

Received 11 August 2017; received in revised form 22 December 2017; accepted 30 January 2018
Available online 22 February 2018

Abstract

From the onset of cognitive revolution, the concept of mental imagery has been given different, many times opposing, theoretical accounts. Mental imagery appears to be a ubiquitous, yet wholly individual, easy to explain experience on the one hand, being hard to deal with scientifically on the other hand. The focus of this research is on an enactive approach to visuospatial mental imagery, inspired by Sima's perceptual instantiation theory. We designed a hybrid computational model, composed of a forward model, an inverse model, both implemented as neural networks, and a memory/controller module, that grounds simple mental concepts, such as a triangle and a square, in perceptual actions, and is able to reimagine these objects by performing the necessary perceptual actions in a simulated humanoid robot. We tested the model on three tasks – salience-based object recognition, imagination-based object recognition and object imagination – and achieved very good results showing, as a proof of concept, that perceptual actions are a viable candidate for grounding the visuospatial mental concepts as well as the credible substrate of visuospatial mental imagery.

Keywords: Enaction; Mental imagery; Visuospatial cognition; Saccades; Cognitive robotics

1. Introduction

Mental imagery (MI) is a phenomenon that has been given multiple (many times opposing) theoretical accounts from the start of the cognitive revolution, being tackled by such prominent figures as Pylyshyn (1973, 2002), Fodor (1975), Block (1981), Kosslyn (1980, 1994) and Barsalou (1999). The plethora of research on the topic is grounded in the fact of MI being an ubiquitous, yet wholly individual experience on the one hand, and easy to explain, yet hard to deal with scientifically on the other hand. A textbook definition (Eysenck, 2012) paints MI as the representation in a person's mind of the physical world outside of that person. It is characterized as a quasi-perceptual experience, as it occurs in the absence of what is perceived to be the appropriate stimuli from the outside. Aside from representing such a rich element in our mental lives, it is thought to be central to many cognitive abilities, such as memory (Paivio, 1986) and motivation (McMahon, 1973), but its foremost role is its involvement in visuospatial reasoning (Sima, 2014) and creative thought (Palmiero et al., 2016). The former is the focus of our own research.

* Corresponding author.
E-mail address: farkas@fmph.uniba.sk (I. Farkaš).
[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.cogsys.2018.01.005
1389-0417/

There are many approaches to researching visuospatial MI, both theoretical and methodological. There are three prevailing theories: the pictorial theory (Kosslyn, 1994), the descriptive theory (Pylyshyn, 2002) and the enactive theory (Thomas, 1999). The pictorial theory claims that MI is the processing of the mental image in the visual buffer using processes of visual perception. This visual buffer is supposedly used in a parallel way during visual perception in order to create a mental representation of what is perceived. The descriptive theory claims that MI is the processing of amodal descriptions, which constitute the mental image. These descriptions are not a part of, or processed by, sensorimotor-related mechanisms. The enactive theory claims that MI emerges with the use of the same schemata that are used for perceiving the external world, e.g., certain schemas of eye movements. For instance, the well known Soar symbolic cognitive architecture, extended with a spatial visual system and a mental imagery module (Lathrop & Laird, 2009) has features of pictorial and descriptive theories, but not the enactive theory.

The enactive theory will be described more in-depth, as it serves as a paradigm for this research. Methodologically, analytic and synthetic approaches to science (Mirolli & Parisi, 2009) are both valid when researching MI (Sima, 2014). The analytic approach to science constitutes researching a phenomenon through observation and experiment. Cognitive psychology (Chambers & Reisberg, 1985), cognitive neuroscience (Bartolomeo & Chokron, 2002) and phenomenology (Thompson, 2007) have dealt with MI in this way. The synthetic approach to science tries to understand phenomena by making computer or robot models. The approach tries to apply principles, used and learned from successful implementations of computer models, to explain real phenomena. It sees models as possible explanations of reality. More specifically, one of the most common methods in modeling cognitive phenomena is the use of artificial neural networks (ANNs), which serve as a bridge between behavior and biology (O'Reilly & Munakata, 2000). ANNs were used in this research as well.

The paper is organized as follows. Section 1 provides of an overview of enactive approaches to mental imagery, including perceptual instantiation theory (Sima, 2014), that serves as the main conceptual source for our work. Section 2 presents the architecture of our model. Section 3 presents the simulations of the proposed model on three specified tasks. Section 4 describes the results of simulations. Section 5 provides the discussion of the model performance and the potential extensions. Section 6 summarizes the paper.

1.1. Enactive approaches to vision

The fundamental movement that spawned enactive sensorimotor approaches was the ecological cognition movement. One of the most important concepts from it is Neisser's (1976) schema, conceptualized to account for his idea of cognition, especially perception. According to Neisser, organisms don't just pick up information from the environment, they actively search for the information they need from the environment. Schemata serve to explain how organisms extract needed information. Organisms use participatory schemata to select information by constructing anticipations of information and waiting for the information to occur in the environment. Only then can information be acquired. Neisser's notion summarizes this: "We can see only what we know how to look for" (Neisser, 1976, p. 20). Therefore, there is a direct relation between perception and action. Schemata are a part of the perception-action cycle: schemata direct action to information, which is picked up by action and goe to schemata, modifying it in the process.

Neisser's account is somewhat consistent with the well-known ecological approach to visual perception (Gibson, 1986). It similarly focuses on researching how an active agent extracts information from the environment. Gibson also rejects the idea that sensory inputs are simply transformed into perceptions by some processes in the mind, and strongly advocates that perception can only be explained in terms of active observers, especially observers that move (or, more accurately, perform a motor action). Perception is therefore by definition not passive. The most relevant concept from Gibson's approach for the means of this research is the idea of affordances. Simply stated, an affordance is what environment affords or offers the agent. In more applicable terms, it is especially connected to categorization. By taking affordances seriously, categories can be defined by actions affording the perceptions of a specific category.

Arbib (1981) relies on Gibsonian ecological psychology and Neisser's concepts to offer his account on the phenomena, heavily shaped by cybernetics and control theory. He unambiguously characterizes perception "as potential action" (Arbib, 1981, p. 1459) through the concept of action-perception cycles, saying: "The subject's exploration of the visual world is directed by anticipatory schemas, which Neisser defines as plans for perceptual action as well as readiness for particular kinds of optical structure. The information picked up modifies the perceiver's anticipations of certain kinds of information that, thus modified, direct further exploration and prepare the perceiver for more information" (Arbib, 1981, p. 1458).

These approaches were most prominently followed by a more contemporary enactive, sensorimotor theory of perceptual consciousness (O'Regan's and Noë's, 2001). A similar idea emerges as before – that sensory stimulation depends on an active agent, on a perceiver in action. However, O'Regan and Noë attribute more power to action, as they don't believe that acting is only for retrieving sensory information – it equally contributes to perception itself as a whole, as experience.

Another aspect, not directly present in enactive visual perception accounts, yet clearly related, is the construction of our personal visual world and the role of saccades in this