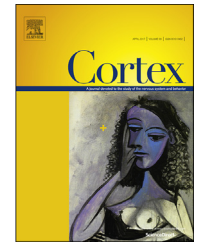
Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex



Research report

Action perception as hypothesis testing

Francesco Donnarumma [a], Marcello Costantini [b,c,d], Ettore Ambrosini [e,c,d], Karl Friston [f] and Giovanni Pezzulo [a,*]

[a] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
[b] Centre for Brain Science, Department of Psychology, University of Essex, Colchester, UK
[c] Laboratory of Neuropsychology and Cognitive Neuroscience, Department of Neuroscience and Imaging, University G. d'Annunzio, Chieti, Italy
[d] Institute for Advanced Biomedical Technologies – ITAB, Foundation University G. d'Annunzio, Chieti, Italy
[e] Department of Neuroscience, University of Padua, Padua, Italy
[f] The Wellcome Trust Centre for Neuroimaging, UCL, London, UK

ARTICLE INFO

Article history:
Received 30 May 2016
Reviewed 12 September 2016
Revised 21 November 2016
Accepted 18 January 2017
Action editor Laurel Buxbaum
Published online 31 January 2017

Keywords:
Active inference
Action observation
Hypothesis testing
Active perception
Motor prediction

ABSTRACT

We present a novel computational model that describes action perception as an *active* inferential process that combines *motor prediction* (the reuse of our own motor system to predict perceived movements) and *hypothesis testing* (the use of eye movements to disambiguate amongst hypotheses). The system uses a generative model of how (arm and hand) actions are performed to generate hypothesis-specific visual predictions, and directs saccades to the most informative places of the visual scene to test these predictions – and underlying hypotheses. We test the model using eye movement data from a human action observation study. In both the human study and our model, saccades are proactive whenever context affords accurate action prediction; but uncertainty induces a more reactive gaze strategy, via tracking the observed movements. Our model offers a novel perspective on action observation that highlights its *active* nature based on prediction dynamics and hypothesis testing.

1. Introduction

The ability to recognize the actions of others and understand their underlying intentions is essential for adaptive success in social environments – and we humans excel in this ability. It has long been known that brain areas such as superior temporal sulcus (STS) are particularly sensitive to the kinematic and dynamical signatures of biological movement that permit its fast recognition (Giese & Poggio, 2003; Puce & Perrett, 2003). However, the neuronal and computational mechanisms linking the visual analysis of movement kinematics and the recognition of the underlying action goals are more contentious.

* *Corresponding author*. Institute of Cognitive Sciences and Technologies, National Research Council, Via S. Martino della Battaglia, 44, 00185 Rome, Italy.
E-mail address: giovanni.pezzulo@istc.cnr.it (G. Pezzulo).
http://dx.doi.org/10.1016/j.cortex.2017.01.016

In principle, the recognition of action goals might be implemented in perceptual and associative brain areas, similar to the way other events such as visual scenes are (believed to be) recognized, predicted and understood semantically. However, two decades of research on action perception and mirror neurons have shown that parts of the motor system deputed to specific actions are also selectively active during the observation of the same actions when others perform them. Based on this body of evidence, several researchers have proposed that the motor system might support – partially or totally – action understanding and other functions in social cognition (Kilner & Lemon, 2013; Rizzolatti & Craighero, 2004). Some theories propose an automatic mechanism of *motor resonance*, according to which the action goals of the performer are "mirrored" in the motor system of the perceiver, thus permitting an automatic understanding (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Other theories highlight the importance of (motor) prediction and the covert reuse of our own motor repertoire and internal models in this process. For example, one influential proposal is that STS, premotor and parietal areas are arranged hierarchically (in a so-called *predictive coding* architectural scheme) and form an *internal generative model* that predicts action patterns (at the lowest hierarchical level) as well as understanding action goals (at the higher hierarchical level). These hierarchical processes interact continuously through reciprocal top-down and bottom-up exchanges between hierarchical levels, so that action understanding can be variously influenced by action dynamics as well as various forms of prior knowledge; such as the context in which the action occurs (Friston, Mattout, & Kilner, 2011; Kilner, Friston, & Frith, 2007). Numerous other theories point to the importance of different mechanisms besides mirroring and motor prediction, such as Hebbian plasticity or visual recognition (Fleischer, Caggiano, Thier, & Giese, 2013; Heyes, 2010; Keysers & Perrett, 2004), see Giese and Rizzolatti (2015) for a recent review. However, these theories implicitly or explicitly consider action observation as a rather passive task, disregarding its enactive aspects, such as the role of *active information sampling* and *proactive eye movements*.

In everyday activities involving goal-directed arm movements, perception is an active and not a passive task (Ahissar & Assa, 2016; Bajcsy, Aloimonos, & Tsotsos, 2016; O'Regan & Noe, 2001); and eye movements are proactive, foraging for information required in the near future. Indeed, eyes typically shift toward objects that will be eventually acted upon, while being rarely attracted to action irrelevant objects (Land, 2006; Land, Mennie, & Rusted, 1999; Rothkopf, Ballard, & Hayhoe, 2007). A seminal study (Flanagan & Johansson, 2003) showed that when people observe object-related manual actions (e.g., block-stacking actions), the coordination between their gaze and the actor's hand is very similar to the gaze-hand coordination when they perform those actions themselves. In both cases, people proactively shift their gaze to the target sites, thus anticipating the outcome of the actions. These findings suggest that oculomotor plans that support action performance can be reused for action observation (Flanagan & Johansson, 2003) and might also support learning and causal understanding of these tasks (Gredebäck & Falck-Ytter, 2015; Sailer, Flanagan, & Johansson, 2005).

Here we describe and test a novel computational model of action understanding and accompanying eye movements. The model elaborates the predictive coding framework of action observation (Friston et al., 2011; Kilner et al., 2007) but significantly extends it by considering the specific role of active information sampling. The model incorporates two main hypotheses. First, while most studies implicitly describe action observation as a passive task, we cast it as an active, *hypothesis testing* process that uses a generative model of how different actions are performed to generate hypothesis-specific predictions, and directs saccades to the most informative (i.e., salient) parts of the visual scene – in order to test these predictions and in turn disambiguate among the competing hypotheses (Friston, Adams, Perrinet, & Breakspear, 2012). Second, the generative model that drives oculomotor plans across action performance and observation is the same, which implies that the motor system drives predictive eye movements in ways that are coherent with the unfolding of goal-directed action plans (Costantini, Ambrosini, Cardellicchio, & Sinigaglia, 2014; Elsner, D'Ausilio, Gredebäck, Falck-Ytter, & Fadiga, 2013).

We tested our computational model against human data on eye movement dynamics during an action observation task (Ambrosini, Costantini, & Sinigaglia, 2011). In the action observation study, participants' eye movements were recorded while they viewed videos of an actor performing an unpredictable goal-directed hand movement toward one of two objects (targets) mandating two different kinds of grip (i.e., a small object requiring a precision grip or a big object requiring a power grasp). To counterbalance the hand trajectories and ensure hand position was not informative about the actor's goal, actions were recorded from the side using four different target layouts. Before the hand movement, lasting 1000 msec, the videos showed the actor's hand resting on a table (immediately in front of his torso) with a fixation cross superimposed on the hand (1000 msec). Participants were asked to fixate the cross and to simply watch the videos without further instructions. In half of the videos, the actor preformed a reach-to-grasp action during which the preshaping of the hand (either a precision or a power grasp, depending on the target) was clearly visible as soon as the movement started (preshape condition), whereas in the remaining half, the actor merely reached for – and touched – one of the objects with a closed fist; that is, without preshaping his hand according to the target features (no shape condition). Therefore, there were four movement types, corresponding to the four conditions of a two factor design (pre-shape and target size); namely, no shape–big target, no shape–small target, pre-shape–big target and pre-shape–small target. The four conditions were presented in random order so that the actor's movement and goal could not be anticipated. The main result of this study was that participants' gaze proactively reached the target object significantly earlier when motor cues (i.e., the preshaping hand) were available. In what follows, we offer a formal explanation of this anticipatory visual foraging in terms of active inference.