Consciousness and Cognition 48 (2017) 40–54

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Review article

Semantic and pragmatic integration in vision for action

Silvano Zipoli Caiani [a,*], Gabriele Ferretti [b]

[a] Department of Humanities, University of Florence, via Bolognese, 52, 50139 Florence, Italy
[b] Department of Pure and Applied Science, University of Urbino Carlo Bo, Italy

ARTICLE INFO

Article history:
Received 21 July 2016
Revised 10 October 2016
Accepted 23 October 2016
Available online 4 November 2016

ABSTRACT

According to an influential view, the detection of action possibilities and the selection of a plan for action are two segregated steps throughout the processing of visual information. This classical approach is committed with the assumption that two independent types of processing underlie visual perception: the *semantic* one, which is at the service of the identification of visually presented objects, and the *pragmatic* one which serves the execution of actions directed to specific parts of the same objects. However, as our knowledge of vision has improved over the years, this established view has turned out to be only an approximation. This paper sets out the details of a non-modularist approach to visual perception of action possibilities and explains how to resist the lure of cognitive segregation.

Contents

1. Introduction

The environment presents cognitive agents at all times with two fundamental tasks: the *detection* of action possibilities and the *selection* of one possible action among them. In the former case, the agent must be able to spot the available motor

* Corresponding author.
E-mail addresses: silvano.zipolicaiani@unifi.it (S. Zipoli Caiani), gabriele.ferretti88@gmail.com (G. Ferretti).

http://dx.doi.org/10.1016/j.concog.2016.10.009
1053-8100/

interaction patterns in the environment, whereas in the latter case the agent must be able to achieve the right motor plan according to his or her actual skills and goals. These tasks may be performed, in part, by relying on visual information about the environment and, in part, by exploiting information about the agent's semantic identification of the target and his/her motor intentions.

According to an influential paradigm in the cognitive science of vision, the agent approaches the *detection* and *selection* tasks in serial mode, namely, finding what to do in the environment before specifying how to do it motorically (Davidson, 1980; Searle, 1983). In this view, the cognitive system addresses these issues through an ordered and hierarchical series of steps: the agent first constructs a spatial representation of the environment, then identifies the presence of motor possibilities therein, and finally select and executes the appropriate motor plan for action. To accomplish these tasks, the system initially collects sensory information to build a representation of the external world; this information is subsequently compared with the representations of previously acquired knowledge and current intentions to decide upon the right course of action. After the external and internal information have been combined at the *final stage* of the perceptual process, a decision is made, and the resulting motor plan runs to generate a desired action (e.g., Marr, 1982; Newell & Simon, 1972).

This classical conception assumes that two *segregated* types of processing contribute to visual perception for action: the *pragmatic* one, which serves the execution of actions on specific parts of visually presented objects, and the *semantic* one, which is at the service of the identification and recognition of the same objects (e.g., Jacob & Jeannerod, 2003; Jacob & Jeannerod, 2007). According to this distinction, pragmatic processing subserves the detection task of computing patterns of visual information that are related to possibilities of action, whereas semantic processing is involved in the selection tasks and provides identity information concerning, for example, the *customary* way to interact and use visually presented items. According to this view, the agent first collects action-related visual information *independently* of motor intentions by means of pragmatic processing, then uses this information to make decisions based on higher-order information concerning object identity and intentional plans for action by means of semantic processing.[1] Consequently, the *pragmatic detection* of action possibilities and the *semantic selection* of motor plans for action appear as two segregated steps in a chain of visual processes.

This view is also reflected in the traditional way of thinking about the role of the motor system. For decades the function of motor areas has been reduced to executive tasks. This framework found sensory areas in occipital lobe and motor areas in posterior part of frontal lobe, while, between them, associative (temporo-parietal) areas put together information from sensory areas and send percepts to motor areas to organize movement: the idea was that associative areas were committed to higher cognitive functions, while motor areas dealt only whit motor execution (Rizzolatti & Sinigaglia, 2008 for a critical analysis).

However, as our knowledge of visual processing has increased over the years, this classical view has turned out to be only an approximation. Recent proposals support an alternative view according to which cognitive systems detect sensorimotor patterns by means of a continuous *matching* between the perceptual stimulus and the agent's vocabulary of motor acts (Rizzolatti et al., 1988). Consequently, the detection of possible motor actions and the selection from among possible motor plans is made *concurrently* on the basis of contextual cues and intentions (e.g., Borghi & Riggio, 2015; Cisek, 2007). According to such an approach, the visual detection of the possibilities of action not only depends on spatial and structural properties of the environment but also relies on a variety of factors, such as the target's usability and the observer's motor capacities, plans and goals, as well as on the visuomotor memory s/he has built in past experience of motor interactions. This conception supports the view that *detecting* and *selecting* possibilities of action are not two hierarchical and segregated steps during visual processing and emphasizes the role of a *motor format of representation* shared by both sub-processes. Accordingly, the traditional way of thinking about the motor system and its role among the other cortical areas is no longer able to explain how visual information is translated in movement. Contemporary evidence, indeed, suggests that while the motor system is not only involved in executive functions, the motor cortex influences the perceptual side of the parietal lobe which is strongly involved in motor activities so to be considered as part of the motor brain (Fogassi & Luppino, 2005; Mountcastle, Lynch, Georgopoulos, Sakata, & Acuna, 1975; Rizzolatti & Sinigaglia, 2008).

The notion of *motor representation* holds a special place in this non-classical view, serving the function of shared vehicle for *perceptual information* and *motor intentions,* that is, for translating perceptual information in motor information suitable to satisfy motor intentions. Motor representations are usually conceived as representations of actions, which are involved in determining the series of bodily movements that an agent needs to perform to reach a certain goal (Butterfill & Sinigaglia, 2014; Ferretti, 2016b; Jeannerod, 1994; Jeannerod, 2006; Pacherie, 2000). The integration of visuo-spatial and visuomotor parameters in relation to the intentional contents delivered by such a class of representations has been made explicit by much experimental evidence. It is well known, for example, that viewing tools and images of tools activates the observer's premotor cortex independently of his/her intention to act (e.g., Chao & Martin, 2000; Grafton, Fadiga, Arbib, & Rizzolatti, 1997). This suggests that the visual representation of a tool activates the motor program used to interact with that tool. According to this hypothesis, Handy, Grafton, Shroff, Ketay, and Gazzaniga (2003) found that the recognition of action-related attributes in the visual field could bias object competition, driving spatial attention toward the location of graspable objects. In particular, authors examined the event-related potentials (ERPs) elicited by images of tools and non-tools, finding that the observer's spatial attention was drawn to tools in the hemifields that are dominant for visuomotor processing

[1] When talking about intentions, we mean those linked to representations concerning the level of grain at which action selection occurs during motor processing. The selection of an action triggered by an intention might be viewed most of the time as automatic and unconscious - this is uncontroversial in the literature (see Jacob & Jeannerod, 2003: esp. Sec. 8.1; Jeannerod, 1994; Pacherie, 2000).