Detecting and Selecting Action Possibilities An Integration Network

Silvano Zipoli Caiani[1]

Abstract. The environment presents cognitive agents all times with two fundamental tasks: the detection of action possibilities and the selection of one possible action. In the former case, the agent must be able to spot interaction patterns in the environment, while in the latter case the agent must be able to achieve the right motor plan according to her actual skills and goals. These tasks are performed, in part, by relying on information about the environment and, in part, by exploiting information about the agent's motor intentions. The aim of this paper is to illustrate how an integrated process of pattern detection and action selection may take place in natural cognitive systems. This amounts to show that the function of detecting and the function of selecting possibilities of action are not, respectively, the beginning and the end of the perceptual process, and that they are not subserved by reciprocally segregated flows of information. Contrary to the classical view, the processes of detecting and selecting occur concurrently and integrate each other, rather than being sequential and reciprocally encapsulated.

1 INTRODUCTION

The aim of this paper is to elucidate how the detection and the selection of visual patterns for action take place in natural cognitive systems. The paper shows that the function of detecting and the function of selecting possibilities of action are not, respectively, the initial step and the final step of a perceptual process, and that they are not grounded on two segregated streams of information. Contrary to what has been classically assumed, the processes of detecting and the process of selecting integrate each other, rather than being reciprocally encapsulated. The main point of this paper concerns the possibility to frame the detection and the selection of action opportunities in terms of information processing without relying on a massive modularity of perception, that is, without postulating modules dedicated to the perception of visual affordances. Indeed, there are several data which present some difficulty to be reconciled with the idea that the sensorimotor control is detached from cognition, and that detecting motor patterns in perception is a different process in respect of selecting motor plans for action. Contrary to the classical information processing approach to visual affordances, it is possible to assume that information integration occurs in different moments in the perceptual processing, allowing for a fine-grained selection of motor cues according to the agent's purposes. This means that a revised approach to visual affordance should abandon the old fashioned idea that the perception of action possibilities relies on a stimulus-driven process. Rather, a brand new approach to visual affordances does not require the postulation of dedicated modules.

Understanding how cognitive agents perceive possibilities of action in the environment might help to appreciate the role of the sensorimotor system in perception. Moreover, framing the perception of affordances within an information integration architecture, might be of relevance in the design of artificial agents suitable to an effective interaction with the environment.

2 THE CLASSICAL VIEW

Since Gibson has introduced the concept of affordance [1] the relation between action and perception is a largely discussed topic in the cognitive sciences. According to Gibson's influential view, the motor meaning of things in the environment is specified by the ambient array of light, and no internal information is required to complete the stimulus and to determine the subsequent action to execute. Indeed, since for Gibson the ambient is entirely specified in terms of external patterns of light, there is no need for internal representations and computations.

Today, many scholars supports Gibson's original claim that perception can be direct, focusing on the dynamical interactions between cognitive systems and environmental properties [2,3,4,5]. Nevertheless, this dynamical approach can be considered only a part of the story about the role of motor intentions in perception [6].

After decades of theoretical debate, several studies have exploited the notion of affordance as a conceptual tool in order to frame the role of action in perception. Despite the fact that Gibson's original purpose was to support a non-representational approach to cognition, many studies adopted the notion of affordance within a computational framework. According to a classical view, the agent forms an internal representation of the possibilities of action in the environment, and only subsequently integrates them with information concerning her goals and the identity of the target.

According to this view, David Marr [7] argued that the visual system does not defer on environmental information only, but that it internally constraints the variety of accepted inputs and possible outputs. In this way the visual system uses information incorporated into the processing concerning what mappings from distal to proximal stimulus are possible and most probable. Famously, also Jerry Fodor [8] has explicitly held the view that visual perception is based on a massive modular architecture, segregated from higher levels of cognition, and grounded on its own repository of "knowledge". As a consequence of this approach, since visual perception is segregated from the agent's

[1] Dept. of Letters and Philosophy, Univ. of Florence, Italy.
Email: silvano.zipolicaiani@unifi.it

intentions, the detection of affordances and the selection of motor actions take place autonomously.

It should be noted that, since the visual system is encapsulated, its operation should be considered as mainly "stimulus driven". In this case, indeed, only the stimulus itself may guide the visual process and determine what the agent perceives. In light of this, it might be tempting to postulate a visual module with the function of extracting motor cues from the stimulus, and thus framing the perception of affordances on the basis of a segregated processing [9]. Indeed, if visually perceiving the environment is considered as a modular process, the detection of targets for action might be possible without involving higher level resources, such as those concerning the identification of object's properties and agent's intentions.

Over the last few decades, such a view inspired a classical framework in cognitive neuroscience. Famously, Milner and Goodale [10] proposed a new version of the already known account of the two visual streams provided by Ungerleider and Mishkin [11], arguing that the ventral and the dorsal paths subserve two segregated flows of visual information. This classic view lies mainly in the discovery of a functional difference between the two streams. While the detection of possible actions in the environment involves the dorsal stream, [12], the ventral pathway assigns an identity to visual patterns encoding semantic information. Although both streams process information concerning objects and their spatial locations, they manipulate visual information in a quite different way: the dorsal stream has a role in mediating the visual detection of possibilities for action, while the ventral stream allows to recognize targets and to select intentional actions according to the agent's motor plans and goals.

On the basis of a modular approach to vision, Goodale and Milner [10, 13] have stated that the dorsal stream blends directly into the premotor areas and forms a segregated "parieto-frontal visuomotor system". In light of this, the dorsal stream does not exploit resources processed by the ventral stream, such as those concerning the identification of an object's property and the agent's goals, but rather it relies only on "stimulus-driven" information that directly specifies the possibilities of action available in the environment. In line with this classical view, also Raftopoulos [14] has argued that the selection of successful actions requires that the visual target is computed in an "absolute metric", and that the visual apparatus is able to pick up motor information directly from the visual stimulus, without relying on semantic cues and intentional information. Accordingly, since the processing taking place along the dorsal pathway is precluded to higher level processing, the resulting function of the dorsal flows is only that of computing motor information in a body-centred frame of reference.

Resuming, the classical computational approach to vision assumes that visual information is computed by two parallel streams of processing. One stream detects motor information by relying on a modular process driven by the perceptual stimulus, while the other stream complements the stimulus with semantic and intentional information. The interaction between such two flows occurs at the end of the process, where the blend of motor and intentional information gives rise to the selection of a plan for action.

3 THE REVISED VIEW

As an alternative to the classical computational approach to visual affordances, one may assume that the selection of motor plans is not confined to the very last part of the perceptual process, but occurs in parallel with the detection of visual patterns for action. In addition, it is also possible that the selection task actually affects the detection task, with the result that the perception of affordances in the environment is always shaped by the agent's motor intentions and goals. In particular, since this view avoids the introduction of ad hoc modular subsystems, it represents a dramatic change in respect of classical approach. In particular, while in the classical view the detection of motor patterns is based on the functioning of segregated computations, the revised view assumes that perceiving affordances is not entirely a stimulus-driven process. Rather, this hypothesis outlines an integration between perceptual and intentional information, thus setting out a different, non-modular architecture.

Interestingly, there is evidence that different interactions exist between the dorsal and ventral streams. Contrary to the classical hypothesis, several data suggest that the ventral stream processing biases the detection of motor patterns by means of functional interactions with different points of dorsal processing [15]. Famously, Jeannerod et al. [16] reported on a subject with a lesion in the dorsal stream who exhibited dysfunctional grasping movements when faced with uncommon objects, while he was accurate in grasping common objects of the same size. This evidence suggests that the ventral stream's encoding of semantic and functional information may compensate for dorsal processing deficit, when the agent is dealing with targets previously known. Moreover, evidence of the existence of an interaction between the two perceptual pathways has been seen in patients suffering from visual forms of optic ataxia and visual agnosia. Patients with an impairment of dorsal stream regions may suffer from optic ataxia, but nevertheless show intact performance when a delay is introduced between the perceptual stimulus and the behavioural response [17]. On the other hand, patients with an impairment of the ventral stream may develop a form of visual agnosia, but nevertheless show normal performance in recognizing objects when required to respond immediately [18]. Such evidences suggest that detection and recognition are not functions subserved by segregated flows of information. Indeed, the impairment of one of the two streams does not prevent agents from judging correctly and finding the right course of interaction with the environment. Rather, we witness a reallocation of functions between the dorsal and ventral paths that ensures the functions of both action detection and selection, revealing the existence of a network formed by the ventral and dorsal systems.

Several studies show that the dorsal stream branches into a number of sub-streams specialized in the detection of different kinds of purposeful interactions with the environment. It has been shown that the dorsal pathway is divided at least into two parallel routes: the dorso-dorsal and dorso-ventral sub-paths [19]. The first sub-path encodes information concerning the target's structure and controls the proximal arm movements for executing actions, whereas the second route encodes the target's role and controls the distal hand when executing movements [12]. This physiological and functional division supports the hypothesis that the function of the dorsal stream is to detect several kinds of visually guided opportunities of action. In line with this result, Baumann, et al. [20] provide evidence that the various affordances offered by a single object can evoke the activation of different areas in the anterior intra-parietal cortex. Since encoding multiple possibilities of action at the same time may increase the chances of interference, a supplementary mechanism is required to assess the value of the available patterns. Processes occurring along the ventral pathway can be seen as shaping the processes occurring along the dorsal path, influencing both the detection and the selection of action patterns. In particular, top-down biases from the ventral stream may help to select the possible patterns of interaction with the environmental in the light of the actual agent's knowledge and motor expertise. This includes the selection of a restricted class of bodily postures and behavioural gesture appropriate to a particular task at hand. Such information is thus made available in the course of dorsal processing, to program and control the series of movements needed to carry out a desired action and achieve a goal.

4 CONCLUSIONS

The abovementioned evidences and considerations suggest that the processes underlying the detection of motor affordances and the selection of related behaviours occur simultaneously and continuously. While the information encoded by the dorsal stream is used to specify the available patterns of action through a process of sensorimotor mapping, a different kind of information is concomitantly integrated to assess which of the available pattern will be executed according to the agent's intentions. The fact that the two visual streams are functionally specialized does not mean that their processes should be conceived also as reciprocally segregated. Rather, the integration of these two streams of information at different stages of visual processing makes for a refined view about the coordination between two basic tasks like the detection and the selection of motor opportunities in the environment.

It should be noted that this conclusion is not derived from the review of anatomical data only. The hypothesis that the perception of visual affordances is supported by information concerning the agent's motor intentions and goals is advanced on the basis of the adaptability of this process to deterioration in ventral stream functioning. Indeed, if the detection of sensorimotor possibilities of action were segregated in the dorsal stream, the impairment of the ventral pathway should leave the ability to detect and exploit patterns of action unaltered. However, the evidence shows that this ability is reduced in such cases, and that information from the ventral pathway is already involved at the level of motor processing. This, of course, does not mean that there are not perceptual modules at all, but only that, in the natural agents, modularity should not be sought within the integrated process of detecting and selecting possibilities of action in the environment. This finding should be carefully considered by those scholars involved in projects aiming at reproducing human basic cognitive abilities by means of artificial systems.

REFERENCES

[1] J. Gibson. Gibson, J.. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin (1979).

[2] L. Peper, et al. How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance*. 20:591-612.F (1994).

[3] E. Thelen, et al. The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences*, 24, 1-33 (2001).

[4] S. Lopresti-Goodman, et al. Negative hysteresis in the behavioral dynamics of the affordance "graspable". *Attention, perception and psychodynamics*, 75, 1075-1091 (2013).

[5] B. Mantel, Exploratory movement generates higher-order information that is sufficient for accurate perception of scaled egocentric distance. PLOS ONE, DOI: 10.1371/journal.pone.0120025.(2015).

[6] W. Bechtel, W. Representations and cognitive explanations. *Cognitive Science,* 22, 295-318 (1998).

[7] D. Marr, D. *Vision: A Computational Approach*. Freeman Co. (1982).

[8] J. Fodor, J. *The Modularity of Mind*. Cambridge, MA: MIT Press (1983).

[9] F. Dretske, Meaningful Perception. In *Visual Cognition*, ed. M. Kosslyn, D.N. Osherson. MIT Press, 331-352 (1995).

[10] A.D.Milner, M.A. Goodale *The visual brain in action*. Oxford: Oxford University Press (1995).

[11] L. Ungerleider, Mishkin, M. Two cortical visual systems. In *Analysis of Visual Behavior*, ed. D. Ingle, M. Goodale, and R. Mansfi eld. MIT Press (1984).

[12] G. Rizzolatti, M. Matelli, Two different streams form the dorsal visual system: anatomy and functions. *Exp Brain Res*, 153, 146–157 (2003).

[13] Z. Pylyshyn Is vision continuous with cognition? *Behavioral and Brain Sciences*, 22, 341–365 (1999).

[14] A. Raftoupolus, *Cognition and Perception. How Do Psychology and Neural Science*. Cambridge: MA: MIT Press (2009).

[15] S. Thill, S., et al. Theories and computational models of affordance and mirror systems: An integrative review. *Neuroscience and Biobehavioral Reviews*, 37, 491–521 (2013).

[16] M. Jeannerod, J. Decety, F. Michel, Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*, 32, 369-380 (1994).

[17] N. Cohen, et al, Ventral and dorsal stream contributions to the online control of immediate and delayed grasping: A TMS approach. *Neuropsychologia*, 47, 1553–1562 (2009).

[18] N. Rice, et al. Intact automatic avoidance of obstacles in patients with visual form agnosia. *Experimental Brain Research*, 174, 176–188 (2006).

[19] L. Buxbaum, S. Kalenine, Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Science*, 1191, 201–218 (2010).

[20] M. Baumann, M. Context-specific grasp movement representation in the macaque anterior intraparietal area. *Journal of Neuroscience*, 29, 6436-6448 (2009).