

Acting on What You are Perceiving: The Two-Visual-Systems Hypothesis Revisited

Bin Zhao

Abstract

The two-visual-systems hypothesis proposed by Goodale and Milner is a radical one. If it were to be true, then our common sense such as we are acting on what we are perceiving should be completely abandoned. In this paper, I argue that the hypothesis over-generalizes what happens in simple tasks to what happens in complex tasks. By contrast, I demonstrate that what happens in complex tasks is compatible with our common sense. In a word, though what we are acting on may come apart from what we are perceiving in some cases, that is not the whole story.

Key Words: action, complex tasks, perception, representation, the two-visual-systems hypothesis

DOI: 10.5281/zenodo.10874499

49

Introduction

Since the discovery of the existence of two visual systems by Schneider (1969), the distinction between two separate visual systems has become more and more influential among neuroscientists. Both systems originate from the primary visual cortex. The so-called “dorsal stream” goes forward to the parietal lobe and the so-called “ventral stream” goes forward to the temporal lobe. It is also widely accepted that there is a functional difference between the two streams though it is relatively unclear what the difference is. Mishkin & Ungerleider (1982) might be the first attempt to characterize the difference as they characterize it as the where vs what distinction. In more detail, they propose that the dorsal stream and the ventral stream process spatial and visual features respectively.

Unlike their precursors who understand the functional difference in terms of their visual inputs, i.e., spatial and visual features, Goodale and Milner understand the difference in terms of the output systems

Corresponding author: Bin Zhao

Address: Institute of Foreign Philosophy, Department of Philosophy and Religious Studies, Peking University, 5 Yiheyuan Rd., Haidian Dist., Beijing 100871, China

e-mail ✉ zhaobin@pku.edu.cn

the two streams serve. They think that both streams process spatial and visual features sent from the primary visual cortex. What makes them different from each other is that they process and transform the visual inputs in different ways. The ventral stream transforms visual inputs into perceptual representations that encode the detailed features of objects. The perceptual representations can be then used for cognitive operations such as object recognition and identification. In contrast, the dorsal stream constructs body-centered representations that directly guide action. In this sense, the ventral stream serves “vision for perception” and the dorsal stream serves “vision for action” (Goodale & Milner, 1992; Milner & Goodale, 2008).² The two-visual-systems hypothesis seems very promising. After all, “perceiving the environment and acting on it each present unique problems for the brain to solve. The successful execution of their separate functions demands that the ventral and dorsal streams utilize radically different kinds of processing” (Foley *et al.*, 2015).

Given the independence of the two streams proposed by the two-visual-systems hypothesis, the ventral stream and the dorsal stream construct different representations for different tasks. As Goodale and Milner suggest, “the dorsal stream does not use the high-level perceptual representations of the object constructed by the ventral stream but instead relies on current bottom-up information from the retina to specify the required movement parameters such as the trajectory of the reach and the required grip aperture needed to grasp the target object” (Milner & Goodale, 2008). In a word, the representation in perception is not the representation in action. That is to say, we are not acting on what we are perceiving. For the sake of simplicity, I shall call this idea “the two-visual-representations hypothesis” in this paper.

This idea seems very difficult to accept. For instance, when I’m reaching out to the coffee mug on my desk and trying to grasp it, it sounds plausible to say that I’m using my visual experience to guide my on-line visuomotor control. It will be odd if it turns out that my visual experience has nothing to do with my on-line visuomotor control. This paper targets the two-visual-representations hypothesis and will assume the following structure: I first sketch Goodale and Milner’s alleged double dissociation between vision for perception and vision for action. I shall argue that the empirical evidence does not support the double dissociation. I then argue that a single representation serving both perception and action in complex tasks should be posited. Lastly, I attempt to locate the single representation in the ventral and dorsal streams.

² There are many competing characterizations in the literature. For instance, Jacob and Jeannerod characterize it as the pragmatic processing vs. semantic processing distinction (Jacob and Jeannerod, 2003; Jeannerod and Jacob, 2005); Schenk (2006) and De Wit *et al.* (2012) characterize it as the egocentric vs. allocentric distinction; while Pisella *et al.* (2006) characterize it as the central vision vs. peripheral vision distinction.

The Double Dissociation Between Vision for Perception and Vision for Action

To demonstrate that vision for perception and vision for action are independent of each other, a double dissociation should be established. Basically, we need to show that it is possible for one group to be impaired in vision for perception but not vision for action and another group to be impaired in vision for action but not vision for perception.

Goodale and Milner propose that the dissociation between visual form agnosia and optic ataxia is the dissociation between vision for perception and vision for action. Visual form agnosia results from a lesion in the ventral stream. Patients are impaired in the recognition of the size, shape, and orientation of visually presented objects. However, their preserved visuomotor transformation enables them to reach and grasp visual targets. In contrast, optic ataxia results from a lesion in the dorsal stream. Patients are still able to recognize the size, shape, and orientation of visually presented targets, but are impaired in reaching and grasping them. In what follows, I will go through more detail to see if the empirical evidence establishes the double dissociation between vision for perception and vision for action. Since Goodale and Milner mostly focus on visual form agnosia, I shall leave optic ataxia aside here.

D.F. is a visual form agnosia patient who suffered from carbon monoxide intoxication because of a leaky propane gas heater which caused severe damage to her ventral stream.³ Goodale and Milner designed two tasks to assess D.F.'s ability of vision for perception and vision for action respectively. D.F. was asked to view a cylinder into which a slot had been cut. The orientation of the slot could be varied by rotating the cylinder. She was holding a card in her hand. In the perception task, she was asked to match the orientation of the card to that of the slot. In the action task, she was asked to insert the card into the slot. D.F. failed the perception task but her performance on the action task was relatively normal. Based on the above finding as well as the finding that optic ataxia patients who can recognize objects but cannot use visual information to guide their actions, Goodale and Milner claim that there is a double dissociation between vision for perception and vision for action (Goodale *et al.*, 1991; Goodale & Milner, 1992; Milner & Goodale, 2008).

Another piece of evidence supporting the alleged dissociation comes from the different effects pictorial illusions have on vision for perception and vision for action respectively. On the one hand, vision for perception is sensitive to pictorial illusions. On the other hand, vision for action is insensitive to the pictorial illusions. For instance, it has been shown that “the scaling of grip aperture in-flight was

³ Recent studies demonstrate that she also has damage to her dorsal stream. This weakens Goodale and Milner's conclusion. See Hesse et al. (2012).

remarkably insensitive to the Ebbinghaus illusion, in which a target disc surrounded by smaller circles appears to be larger than the same disc surrounded by larger circles (Aglioti *et al.*, 1995). In short, maximum grip aperture was scaled to the real not the apparent size of the target disc” (Milner & Goodale, 2008). This sort of evidence further supports the alleged dissociation.

Is There a Double Dissociation?

Goodale and Milner cited some evidence for their alleged double dissociation between vision for perception and vision for action. In this section, I shall argue that their evidence is not strong enough to establish the double dissociation.

As mentioned above, a support for the double dissociation is that the visual form agnosia patient D.F.’s visuomotor control is relatively intact. But is D.F.’s visuomotor control really intact? Goodale and Milner think so because her performance on the action task in their experiment was relatively normal. However, I think that is not strong enough to establish the conclusion. The action task at most shows that her visuomotor control is relatively intact in simple tasks such as reaching out to and grasping objects. What about these complex tasks which involve complicated coordination of the body in space? Is her performance normal in these tasks?

Let’s focus on one such complex task. In the three-hole task, subjects were required to open their eyes, and immediately reach out to a circular disk that had three holes for fingers, and correctly place their fingers inside the holes. This task is more complicated than tasks that only involve reaching out to and grasping objects as it demands a finer-grained representation and complicated coordination of the body in space. The result shows that D.F.’s performance was inferior to controls. In a similar two-hole task, D.F.’s performance was much better. She adjusted her hand to the orientation of and the location of the holes almost as accurately as controls though she still placed her fingers in the wrong holes on many occasions and her grip aperture was insensitive to the distance between the holes (Dijkerman *et al.*, 1998; Briscoe & Schwenkler, 2015).

In a word, D.F.’s performance gets worse when the task gets complicated. Goodale and Milner can at most demonstrate that D.F. has relatively intact visuomotor control in simple tasks but not in complex tasks. On the one hand, visual form agnosia patient D.F. does not have intact visuomotor control in complex tasks. On the other hand, healthy subjects have intact visuomotor control in complex tasks as well as in simple tasks. Thus, what accounts for their difference in visuomotor control? Since the most significant difference between them lies in their ventral streams, it is reasonable to focus on the ventral stream to explain the difference in visuomotor control. This is also what Goodale and Milner attempt to do. As they write, “this

conscious monitoring of unpractised movements [in complex tasks] depends upon information provided by the perceptual networks in the ventral stream. As a consequence, ventral-stream processing can intrude into the visual guidance of these movements. Once the action is well-practised and becomes automatized, however, it seems that control of the constituent movements is passed to visuomotor networks in the dorsal stream, which then play the dominant visual role” (Milner & Goodale, 2008a).⁴

It seems that they are willing to attribute the ability of visuomotor control to the ventral stream in complex tasks. This directly contradicts their hypothesis which only attributes vision for perception to the ventral stream. Thus, they have revised their hypothesis as follows: the ventral stream is responsible for both vision for perception and vision for action in complex tasks. Once a task becomes well-practised and automatized, the dorsal stream will take over the vision for action though what contributes to vision for perception is still the ventral stream.

If that is the case, then could Goodale and Milner still maintain that the representation in perception is not the representation in action in complex tasks? They might claim that even though what is responsible for both vision for perception and vision for action is the ventral stream in complex tasks, there are two independent representations in the ventral stream which respectively contribute to vision for perception and vision for action. That is to say, we are still not acting on what we are perceiving in complex tasks. It seems that Goodale and Milner still have enough resources to maintain the two-visual-representations hypothesis even if the two-visual-systems hypothesis fails.

However, this way of maintaining the two-visual-representations hypothesis does not do enough justice to other evidence in the literature. One of Goodale and Milner’s evidence supporting the alleged dissociation between vision for perception and vision for action comes from the different effects pictorial illusions have on vision for perception and vision for action respectively. As they suggest, unlike vision for perception, vision for action is insensitive to optical illusions. The problem is that some experiments demonstrate that vision for action is also sensitive to pictorial illusions. Consider the following experiment,

“Participants were presented with a [regular two-tailed] version of a Müller-Lyer illusion [there are three configurations: hoop-in, hoop-out, and no hoop] and asked either to throw a beanbag to the end location of the corresponding line (i.e., shaft) or to provide a verbal estimate of the egocentric distance to that location. . . . In each trial participants stood at a distance of 1.5 m from the beginning of the shaft. When the participants stood at a distance

⁴ See Milner and Goodale (2008b) and van der Kamp et al. (2008) for related discussions.

from the illusion rather than on it, the tasks could also be performed by combining egocentric (i.e., distance from the participant's vantage point to the base of the Müller-Lyer shaft) and exocentric (i.e., distance from the base of the Müller-Lyer to the endpoint of the shaft) distances. The exocentric distance component makes additional allocentric information available on which participants can rely to base their responses. Potentially, this allows for a more powerful test of the robustness of vision for action with respect to the illusion... [P]articipants were asked either to perform a beanbag-throwing task or to provide verbal estimates while standing outside of the two-tailed illusion" (Cañal-Bruland *et al.*, 2013).

The result shows that distance was overestimated for the hoop-out relative to the hoop-in illusion. The illusion effect does not differ between the verbal, i.e., vision for perception, and motor tasks, i.e., vision for action. Therefore, contrary to Goodale and Milner's claim, vision for action sometimes is sensitive to optical illusions.

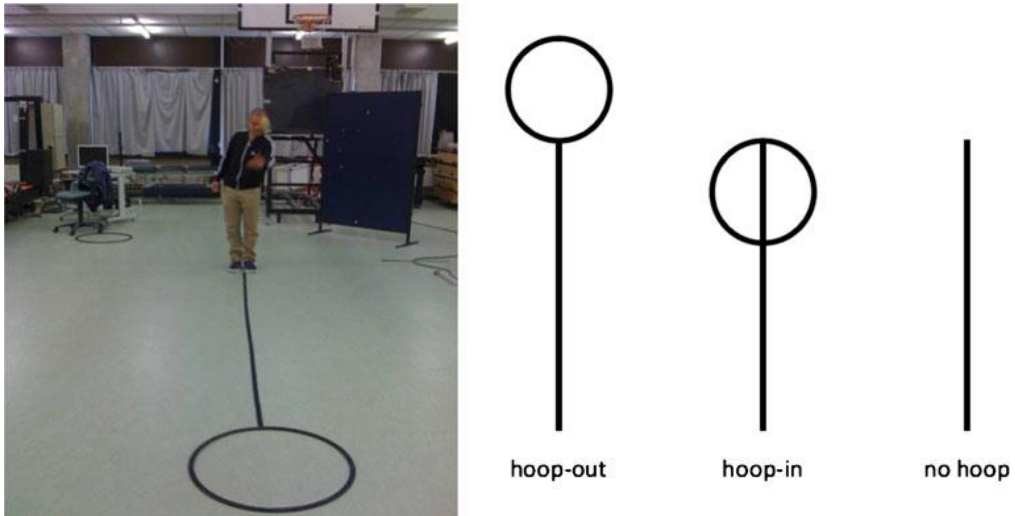


Figure 1. The setup of the experiment (Cañal-Bruland *et al.*, 2013).

It seems that some evidence is not on Goodale and Milner's side. Can they explain away the above result? As they mentioned, "not all experiments that appear to show an effect of perceptual illusions on action are truly doing so" (Milner & Goodale, 2008). If it turns out that the Ebbinghaus display does have some seeming perceptual illusions on vision for action, the seeming perceptual illusion might still be explained away. For instance, "the annulus of circles that surround the target disc in these experiments may influence the movements that are made for purely non-perceptual reasons. One important factor at work here is that the visuomotor system appears to treat these flanking stimuli as potential obstacles to the grasping movement" (Milner & Goodale, 2008). In a word, the seeming perceptual illusions

on vision for action might be explained away by the fact that the fingers are “trying” to avoid the annulus of circles which they treat as obstacles.

This is a plausible way of explaining away the seeming perceptual illusions on vision for action. However, it does not work in the experiment under discussion. Though the fins of the Müller-Lyer illusion might be regarded as obstacles to avoid in the grip test (Biegstraaten *et al.*, 2007), this explanation cannot accommodate the result of the current experiment. If the subject treats the hoop as an obstacle to avoid in the beanbag-throwing task, then distance should be underestimated for the hoop-out relative to the hoop-in illusion. However, the result shows the opposite. Therefore, this experiment demonstrates that both vision for perception and vision for action are vulnerable to pictorial illusions.

This experiment even puts the two visual-representations-hypothesis in danger. If there are two independent representations in the ventral stream which respectively contribute to vision for perception and vision for action, then why are perception and action both vulnerable to pictorial illusions? To locate vision for perception and vision for action in the same stream, e.g., the ventral stream, in complex tasks is not enough to accommodate all the evidence. On the contrary, a single representation should be posited to serve both perception and action.

The Single Visual Representation: Raveling the Golden Braid of Perception and Action

In the last section, some evidence in the literature that cannot be easily accommodated by the two-visual-systems hypothesis proposed by Goodale and Milner has been reviewed. I further argued that a single representation should be posited to serve both perception and action in complex tasks. In this section, I shall try to locate the single representation somewhere in the two streams.

It seems that Goodale and Milner are willing to attribute the ability of visuomotor control to the ventral stream in complex tasks. As a consequence, it is very tempting to locate the single representation in the ventral stream. However, I'm more inclined to locate the single representation in the dorsal stream.

First, if it were the case that the ventral stream is responsible for both vision for perception and vision for action in complex tasks. Then optic ataxia seems hard to explain. As mentioned above, optic ataxia results from a lesion in the dorsal stream. Patients are still able to recognize the size, shape, and orientation of visually presented targets, but are unable to guide the hand toward a specific object by using visual information. If we locate the single representation serving both perception and action in the ventral stream which is intact in optic

ataxia patients, then these patients should be able to perform well in complex tasks. As Goodale and Milner suggest, “ventral-stream processing can intrude into the visual guidance of these movements [in complex tasks]” (Miler & Goodale, 2008). However, the fact is that optic ataxia patients suffer from “misreaching in the contralesional visual field, difficulty preshaping the hand for grasping, and an inability to correct reaches online” (Andersen, 2014). They can neither perform well in simple tasks nor complex tasks. Therefore, the single representation serving both perception and action cannot be located in the ventral stream.

Second, neuroanatomical studies have established numerous connections between the ventral and dorsal streams. This indicates that the ventral and dorsal streams are not working independently. It is more likely that they interact with each other to perform a task. More specifically, “these physiological interconnections appear to be gradually more active as the precision demands of the grasp become higher” (van Polanen & Davare, 2015). Therefore, in complex tasks, it is more plausible that the dorsal stream retrieves detailed information about features of the object in the ventral stream to guide the visuomotor control. If the single representation serving both perception and action is located in the ventral stream, then there is no need for the dorsal stream to retrieve the information in the ventral stream. Given that the interconnections between the ventral and dorsal streams become more active when the task gets complex, I think the plausible picture is as follows: the ventral stream constructs a detailed representation of features of the object, and then the representation is sent to the dorsal stream when the task becomes complex. This representation then guides the visuomotor control in performing the relevant task. According to this picture, at least in some cases, the representation in perception is the representation in action. That is to say, we are acting on what we are perceiving.

Third, according to the above picture, the single representation serving both perception and action is located in the dorsal stream, i.e., the dorsal stream is responsible for vision for action as well as vision for perception. This contradicts the two-visual-systems hypothesis which only attributes vision for action to the dorsal stream. One might wonder if the dorsal stream can process a representation serving perception. This worry can be dismissed by a recent study carried out by Freud et al. (2018) in which a special technique which is called “continuous flash suppression (CFS)” was used. This technique abolishes activation in the ventral stream but still allows largely intact processing in the dorsal stream. The study shows that the activation in the dorsal stream has some priming effect on a later relative depth judgment task which does not involve any visuomotor control. The priming effect is possible only when the dorsal stream can derive a structural description of the target object. That is to say, the dorsal stream can process a representation that serves perception. Thus, the above worry for the picture is dismissed.

Conclusion

Drawing upon the evidence in the literature that cannot be accommodated by the two-visual-systems hypothesis proposed by Goodale and Milner, a rough picture of what happens in the ventral and dorsal streams in complex tasks has been offered in this paper. In this picture, the ventral stream constructs a detailed representation of the features of the object, then the representation is sent to the dorsal stream when the task becomes complex. This representation then guides the visuomotor control in performing the relevant task.⁵ One caution worth mentioning is that this picture does not completely contradict the two-visual-systems hypothesis because Goodale and Milner can still maintain that in simple tasks, the ventral stream constructs a representation that serves perception; whereas the dorsal stream constructs a representation that serves action.

Acknowledgements

Thanks to Alex Morgan, Anthony A. Wright, and an anonymous referee for helpful comments or discussions.

References

- Aglioti S, DeSouza JF & Goodale MA. Size Contrast Illusions Deceive the Eye but Not the Hand. *Curr. Biol* 1995; 5(6): 679-685.
- Andersen RA, Andersen KN, Hwang EJ & Hauschild M. Optic Ataxia: From Balint's Syndrome to the Parietal Reach Region. *Neuron* 2014; 81(5): 967-983.
- Biegstraaten M, de Grave DD, Brenner E & Smeets JB. Grasping the Müller-Lyer illusion: Not a Change in Perceived Length. *Experimental Brain Research* 2007; 176: 497-503.
- Briscoe R & Schwenkler J. Conscious Vision in Action. *Cognitive Science* 2015; 39(7): 1435-1467.
- Cañal-Bruland R, Voorwald F, Wielaard K & van der Kamp J. Dissociations Between Vision for Perception and Vision for Action Depend on the Relative Availability of Egocentric and Allocentric Information. *Attention, Perception, & Psychophysics* 2013; 75(6): 1206-1214.

⁵ The rough picture also has greater explanatory power than the two-visual-systems hypothesis when it comes to agnosia patient D.F. The reason why she can perform well in simple tasks is that there is a representation that serves basic actions in her dorsal stream; while the reason why she cannot perform well in complex tasks is that her damaged ventral stream is not able to process detailed information about features of the object. Therefore, there is no information being sent to the dorsal stream from the ventral stream to construct the representation serving complex actions.

- De Wit MM, Van der Kamp J & Masters RS. Distinct Task-Independent Visual Thresholds for Egocentric and Allocentric Information Pick Up. *Consciousness and Cognition* 2012; 21(3): 1410-1418.
- Dijkerman HC, Milner AD & Carey DP. Grasping Spatial Relationships: Failure to Demonstrate Allocentric Visual Coding in a Patient with Visual Form Agnosia. *Consciousness and Cognition* 1998; 7(3): 424-437.
- Freud E, Robinson AK & Behrmann M. More than Action: The Dorsal Pathway Contributes to the Perception of 3-D Structure. *Journal of Cognitive Neuroscience* 2018; 21: 1-12.
- Foley RT, Whitwell RL & Goodale MA. The Two-Visual-Systems Hypothesis and the Perspectival Features of Visual Experience. *Consciousness and Cognition* 2015; 35: 225-233.
- Goodale MA & Milner AD. Separate Visual Pathways for Perception and Action. *Trends Neurosci* 1992; 15(1): 20-25.
- Goodale MA, Milner AD, Jakobson LS & Carey DP. A Neurological Dissociation Between Perceiving Objects and Grasping Them. *Nature* 1991; 349(6305): 154-156.
- Hesse C, Ball K & Schenk T. Visuomotor Performance Based on Peripheral Vision is Impaired in the Visual Form Agnostic Patient DF. *Neuropsychologia* 2012; 50(1): 90-97.
- Jacob P & Jeannerod M. *Ways of Seeing, the Scope and Limits of Visual Cognition*. Oxford: Oxford University Press, 2003.
- Jeannerod M & Jacob P. Visual Cognition: A New Look at the Two-Visual Systems Model. *Neuropsychologia* 2005; 43: 301-312.
- McIntosh RD & Schenk T. Two Visual Streams for Perception and Action: Current Trends. *Neuropsychologia* 2009; 47(6): 1391-1396.
- Milner AD & Goodale MA. Two Visual Systems Re-viewed. *Neuropsychologia* 2008a; 46: 774-785.
- Milner AD & Goodale MA. The Two Visual Streams: In the right Ballpark? *International Journal of Sport Psychology* 2008b; 39: 131-135.
- Mishkin M & Ungerleider LG. Contribution of Striate Inputs to the Visuospatial Functions of Parieto-Preoccipital Cortex in Monkeys. *Behav Brain Res* 1982; 6(1): 57-77.
- Pisella L., Binkofski F, Lasek K, Toni I & Rossetti Y. No Double-Dissociation Between Optic Ataxia and Visual Agnosia: Multiple Sub-Streams for Multiple Visuo-Manual Integrations. *Neuropsychologia* 2006; 44: 2734-2748.
- Schenk T. An Allocentric Rather Than Perceptual Deficit in Patient D.F. *Nature Neuroscience*, 2006; 9: 1369-1370.
- Schenk T & McIntosh RD. Do we have Independent Visual Streams for Perception and Action? *Cognitive Neuroscience* 2010; 1(1): 52-78.
- Schneider GE. Two Visual Systems. *Science* 1969; 163(3870): 895-902.
- van der Kamp J, Rivas F, van Doorn H & Savelsbergh G. Ventral and Dorsal Contribution to Visual Anticipation in Fast Ball Sport. *International Journal of Sport Psychology* 2008; 39: 100-130.
- van Polanen V & Davare M. Interactions Between Dorsal and Ventral Streams for Controlling Skilled Grasp. *Neuropsychologia* 2015; 79: 186-191.
- Whitwell RL, Milner AD & Goodale MA. The two-visual-systems hypothesis: New Challenges and Insights from Visual form Agnostic Patient DF. *Frontiers in Neurology* 2014; 5: 255.

Authors hold copyright with no restrictions. Based on its copyright *Journal of NeuroPhilosophy* (JNphi) produces the final paper in JNphi's layout. This version is given to the public under the Creative Commons license (CC BY). For this reason authors may also publish the final paper in any repository or on any website with a complete citation of the paper.