PSYCHE

Précis of *Ways of Seeing, the Scope and Limits of Visual Cognition*

Pierre Jacob
Institut Jean Nicod
UMR 8129
Pavillon Jardin
Ecole Normale Supérieure
29, rue d'Ulm
F-75005 Paris, France
Pierre.Jacob@ehess.fr

Marc Jeannerod
Institut des Sciences Cognitives
67, boulevard Pinel
69675 Bron cedex, France
jeannerod@isc.cnrs.fr

PSYCHE 13/2, April 2007

KEY WORDS: two-visual systems, ventral vs. dorsal pathways, semantic vs. pragmatic processing, visual illusions, visual experience

PRECIS OF: P. Jacob & M. Jeannerod. (2003). *Ways of Seeing: The Scope and Limits of Visual Cognition.* Oxford: Oxford University Press, + 287pp. ISBN 0198509200.

Human vision raises a number of puzzles. Among them are the puzzles of visual experience: how to provide a scientific understanding of the phenomenal character of the visual experiences of the shapes, textures, colors, orientations and motion of perceived objects? How can a purely subjective visual experience be the basis of so much objective knowledge of the world? Visually guided actions raise a different (almost complementary) puzzle: how can actions directed towards a target be so accurate in the absence of the agent's awareness of many of the target's visual attributes? *Ways of Seeing* (WoS) has three related goals, the first of which is to make the case for a broadly representational approach to the above set of puzzles. The second goal of WoS is to argue that the version of the 'two-visual systems' model of human vision best supported by the current empirical evidence has the resources to solve the puzzle of visually guided actions, which has been at the center of much recent work in the cognitive neuroscience of vision and action. The third goal of WoS is to draw attention to some of the tensions between acceptance of the two-visual systems model of human vision and some influential views about the nature and function of the content of visual experience espoused by philosophers in response to the puzzles raised by visual experience.

1. The anatomical segregation

The origin of the two-visual systems model is the discovery by Ungerleider and Mishkin (1982) (henceforth U&M) of an important segregation in the visual cortex of primates between two anatomical pathways: the ventral pathway and the dorsal pathway. The former projects the primary visual areas onto the inferotemporal areas. The latter projects the primary visual areas onto the parietal cortex, which is a relay towards the pre-motor and the motor cortex. Although there are many important unsettled issues about the cross-talk between the two visual pathways, one important question has been the basic functional significance of the anatomical bifurcation. On U&M's view, the ventral pathway is primarily the What-stream or object-channel (specialized in object-identification) and the dorsal pathway is primarily the Where-stream or space-channel (specialized in the spatial localization of objects). Milner and Goodale (1995) (henceforth M&G) emphasized the distinction between vision-for-perception and vision-for-action.

How does WoS stand with respect to the differences between U&M's and M&G's models of the division between the two visual streams? As we point out in the Epilogue, the main shortcoming of U&M's model is that it assumes that the chief, if not the only, function of the primate visual cortex is visual perception: the so-called 'visuomotor transformation', i.e., the conversion of visual information into motor commands of hand actions, is relegated to non-cortical connections. The great merit of M&G's model is to incorporate the experimental tradition initiated by Mountcastle et al. (1975), which showed that the visuomotor transformation fully belongs to the primate visual cortex. The main shortcoming of M&G's model is that it underestimates the contribution of part of the human parietal lobe to conscious visual perception. As we emphasize, neuropsychological observations of neglect patients show that the right inferior parietal lobe contributes to the visual perception of spatial relations among objects in a visual scene, which is arguably a pre-condition for full visual awareness of objects.

2. Semantic vs. pragmatic processing

On our view, the basic insight of the two-visual systems model is that many visual stimuli can give rise to two distinct kinds of visual processing in human cognition, according to the task. Thanks to the dexterity of their hands, humans (and to a lesser extent, non-human primates) can reach, grasp and manipulate some of the objects (e.g., a hammer or a tea-cup) that they can see. No doubt, humans can also see a wide variety of things that cannot be manipulated because either they are not individuals (e.g., substances and gases) or they are too large and/or too far (e.g., volcanoes and planets). According to the two-visual systems model, objects that can be manually reached and grasped can be submitted to two kinds of visual processing, which, borrowing from the study of language, we call respectively 'semantic' and 'pragmatic' processing. The former is at the service of the identification and recognition of visually presented objects. The latter is at the service of actions directed upon visually presented objects.

On our view, the main claim of the two-visual systems model is that each kind of processing can occur independently of the other. Support for this claim is provided by double dissociations found in both neuropsychological patients (chapter 3) and healthy subjects (chapter 4) as well. Neuropsychological patients, who are impaired in the visual control of object-oriented actions, turn out to exemplify surprising residual visuomotor capacities. Neuropsychological patients, who are impaired in the visual control of object-oriented actions, turn out to be able to recognize visually presented objects. In healthy subjects, psychophysical experiments show interesting dissociations between perceptual and visuomotor responses to illusory stimuli.

Perceptual and visuomotor tasks make sharply different requirements on the visual system. Since perceptual re-identification of objects must be achieved from many different spatial perspectives, with respect to different orientations of objects, in different occlusion conditions and in different lighting conditions, it requires the encoding of enduring visual properties of objects, i.e., visual information that can be matched onto conceptual information stored in memory. By contrast, the visual control of an object-oriented hand action requires encoding the target's visual attributes relevant for prehension, such as its current orientation, absolute size and shape and its distance from the agent. In chapters 2-4 of WoS, we review the empirical evidence for the two-visual systems model of human vision from single cell recordings in monkeys, the examination of brain-lesioned human patients and psychophysical experiments in healthy human subjects.

If semantically processed, an object that can be manually manipulated gives rise to a visual percept. If pragmatically processed, the same object gives rise to a visuomotor representation. The job of a visual percept is to provide visual information for the benefit of thought or what philosophers call the 'belief-box', i.e., to memory and reasoning systems. The job of a visuomotor representation is to provide visual information for the benefit of an agent's intention. Whereas beliefs have a mind-to-world direction of fit, intentions have a world-to-mind direction of fit (see chapter 1). In chapter 6, we argue that visuomotor representations are, like Millikan's (1996) pushmi-pullyu representations, hybrid representations: they represent facts in a format suitable, not for forming beliefs (and acquiring knowledge) about the world, but for informing an agent's motor intention to act on a target.

3. What is special about visuomotor representations?

In chapter 6, we set ourselves two tasks. The first, which is based on evidence reported in detail in chapter 4, is to show that visuomotor representations are genuine mental representations. In chapter 4, we examine in detail a series of psychophysical experiments performed on healthy human subjects that provide evidence for dissociations between perceptual responses and visuomotor responses to Titchener-like illusory stimuli, in which a central two-dimensional circle surrounded by an annulus of circles (either greater or smaller than it) has been replaced by a graspable three-dimensional disk. Although some of the methodological issues raised by these experiments are unsettled (see e.g., Franz et al., 2001), one interesting experiment performed by Haffenden et al. (2001) suggests that the calibration between index and thumb may be sensitive to different parameters whether it is performed in a perceptual task or in a visuomotor task. In the perceptual task, it depends on the comparison between the diameter of the central disk and the diameter of the circles in the surrounding annulus. In the visuomotor task, it depends on the computation of the gap between the central disk and the surrounding annulus. The visuomotor system would seem to treat the surrounding annulus, which is a 2D feature of the visual display, as a 3D obstacle. If so, then this experiment shows that visual illusions are not restricted to perception (or semantic processing). Special features of the visual display can also fool the visuomotor system. Thus, on our view, the importance of this finding lies in the fact that it provides evidence for the view that the visuomotor system generates misrepresentations, hence genuine representations, of the display.

The second task we set ourselves in chapter 6 is to specify the systematic differences between a visual percept and a visuomotor representation of one and the same object. The main difference, we argue, lies in the way visual percepts and visuomotor representations code the spatial position of the represented object. In a visuomotor representation, what matters is the representation of the actual shape and absolute size of the target to be grasped. In order to guide an action of prehension, a visuomotor representation must specify the position of the target in some egocentric frame of reference centered on the agent's body. In a visual percept, what matters is the representation of the perspectival shape and the relative size of objects contained in a visual scene. In fact, we argue that it is of the essence of a visual percept that it offers the basis for comparative judgments about the relative shape, orientation, size, color, texture and motion of objects contained in a visual array. As a result, the spatial position of a perceived object must be specified in some allocentric frame of reference centered on some other constituent of the visual scene. In a nutshell, we argue that the visuomotor representation of a target of prehension enables apperceptive visual form agnosic patient DF to grasp accurately objects whose shapes and sizes she is not visually aware of. Nor are healthy subjects aware of any tension created by the dissociations between their visuomotor responses and their perceptual responses to illusory stimuli (e.g., Titchener disks surrounded by an illusory annulus). We thus conclude that, unlike the non-conceptual content of visual percepts, the non-conceptual content of visuomotor representations does not contribute to the phenomenal character of the conscious visual experience of objects.

4. Actions performed and actions perceived

Much work in visual science has emphasized the complexity and hierarchical structure of visual perception (or as we call it, the semantic processing of visual inputs). By contrast, the complexity of visually guided actions has been relatively neglected. As the neuropsychological evidence (mostly discussed in chapter 3) shows, the semantic processing of visual inputs can be disrupted at several different levels. Following a lesion in the primary visual cortex, blindsight patients are deprived of the visual experience of objects altogether. Lesions at various steps in the ventral pathway produce perceptual impairments of different degrees or severity. Patients with associative visual form agnosia have been said to form a normal percept "stripped of its meaning". In other words, the information contained in the percept cannot be matched onto information stored in memory for recognition to occur. Patients with apperceptive visual form agnosia (e.g., DF) are unable to assemble an object's local shapes into its global contour. This clinical distinction fits with the conceptual distinction made by some philosophers between two levels of normal visual awareness, which Dretske (1969) calls respectively "non-epistemic perception" and "epistemic perception". Whereas the former does not require that the information picked up by the visual system be matched against information stored in memory, the latter does and, as a result, it can ground perceptual beliefs (see chapter 5).

The hierarchical structure of the semantic processing is paralleled by the hierarchical structure of the pragmatic processing of visual inputs. Indeed, as the neuropsychological evidence (discussed in both chapters 3 and 7) also shows, the pragmatic processing can be disrupted at several distinct levels too. Lesions on either side of the superior parietal lobe result in optic ataxia. In optic ataxic patients, the visuomotor transformation is impaired: they cannot reach objects with the hand contra-lateral to the lesion site and/or calibrate the grip between the index and thumb of their contra-lateral hand to the physical size of their target. The visuomotor transformation is but the lowest level of the pragmatic processing of visual inputs. Lesions in the left inferior parietal lobe result in apraxic syndromes, whereby the pragmatic processing of visual inputs is disrupted at a higher-level: apraxic patients are impaired in the use and recognition of complex tools, in the ability to mime the use of absent tools and in the ability to recognize pantomime of actions with absent tools. In apraxic patients, however, the visuomotor transformation may be intact. Whereas M&G tend to restrict vision-for-action to the visuomotor transformation, we emphasize the fact that the contribution of vision to human actions can occur at higher-levels of pragmatic processing of visual inputs, including the recognition of complex cultural tools and the perception of actions involving the use of tools.

In fact, actions are not merely things human agents perform; they are also things humans visually perceive. In chapter 7, we review the evidence for the view that the human brain contains two complementary networks that respond to the perception of respectively object-oriented actions and actions directed towards conspecifics, and thus provide two complementary entries to the human mindreading system. The former involves connections between parts of STS, the inferior parietal lobe and the pre-motor cortex (in which mirror neurons have been recorded). The latter involves parts of STS, the amygdala and the orbito-frontal cortex.

5. The two-visual systems model and visual awareness of objects

Whereas U&M assumed that both the so-called 'object-channel' and 'space-channel' of the primate visual system are at the service of visual perception, M&G assumed that, unlike the ventral stream, which delivers visual percepts, the dorsal stream delivers visuomotor representations of targets of action. On our view, both semantic processing, which depends on the activity of brain areas in the ventral stream, and pragmatic processing, which depends on the activity of brain areas in the dorsal stream, give rise to representations of visual stimuli at different levels. Neither can visually guided actions in humans be restricted to reaching and grasping, nor should presumably the dorsal stream be restricted to the superior parietal areas (involved in the visuomotor transformation) and deprived of e.g., the left inferior parietal lobe whose activity is required for executing and pantomiming skilled actions involving the use of tools.

Furthermore, M&G seriously underestimated the contribution of areas in the dorsal stream to the conscious perception of objects. As the neuropsychological evidence shows, after a lesion in the right inferior parietal lobe, neglect patients lose the visual awareness of such visual attributes as the shape, size, color, texture and orientation of objects presented in the field contra-lateral to the site of their lesion. However, in neglect patients, areas in the ventral stream involved in the semantic processing of the shapes, sizes, colors, textures and orientations of visual objects are active. Why are neglect patients not aware of the visual attributes presented in their contra-lateral field? What neglect patients seem to be deprived of by the lesion in their right inferior parietal lobe is the ability to represent in allocentric coordinates the spatial locations of objects that are presented in their neglected hemi-space. In other words, they have been deprived of the ability to represent the spatial relations of objects in their neglected hemi-space (see e.g., Driver and Vuilleumier, 2001). Thus, evidence from neglect patients shows that loss of awareness of the spatial relations between objects causes loss of awareness of other visual attributes of objects. In other words, awareness of the visual attributes of objects asymmetrically depends on the awareness of spatial relations between objects. If so (and if the right inferior parietal lobe is part of the dorsal stream), then activity in the dorsal stream does contribute to visual awareness of objects.

6. Some philosophical implications of the two-visual systems model

As we said at the outset, visual experience is puzzling in several respects: why does visual processing give rise to phenomenal experience (rather than not)? How can subjective experience be the basis of objective knowledge of the world? Most twentieth century philosophers have approached the puzzles of visual experience via the following questions that seem more manageable: does the phenomenal character of visual experience consist in visually representing the world? Is the latter necessary and sufficient for the former? Representationalists say ‘Yes’. Non-representationalists say ‘No’. Representationalists further divide into conceptualists and non-conceptualists. According to conceptualists, there is no relevant difference between the content of visual perception and the conceptual content of thoughts, judgments and beliefs. According to non-conceptualists, the non-conceptual content of visual experience differs from the conceptual content of thoughts and judgments in that the former is informationally richer and more fine-grained than the latter.

As we argue in chapter 1, we endorse a broadly representationalist framework for dealing with the puzzles of visual experience. In chapter 5, we address the question how visual experience can be the basis of objective knowledge of the world. We accept the arguments in favor of the distinction between the conceptual content of thoughts and judgments and the non-conceptual content of visual experience. We assume that the non-conceptual content of visuomotor representations is distinct from the conceptual content of thoughts and judgments. On the basis of the evidence for the two-visual systems model of human vision, we further argue that the non-conceptual content of visuomotor representations cannot be identical to the non-conceptual content of visual percepts. In other words, we conclude that non-conceptual content ought to be bifurcated between the content of respectively visuomotor representations and visual percepts.

Whichever view of the content of visual experience they espouse, most twentieth century philosophers who have dealt with the puzzles of visual experience would probably identify vision with sight (or visual experience). From the standpoint of the two-visual systems model, this is a highly questionable assumption. As we pointed out, many philosophers disagree about whether to accept the distinction between the conceptual content of thoughts and judgments and the non-conceptual content of visual experience. Whether they accept it or not, they all tend to accept what Clark (2001) calls the assumption of experience-based control (EBC), i.e., the idea of a constitutive link between the content of visual experience and the fine-tuning of visually guided actions directed towards objects. This assumption is clearly part of O'Regan and Noë's (2001) enactive theory, according to which an agent's visual experience of an object consists in his or her skillful actions with respect to the object or in his or her tacit knowledge of the sensory consequences of his or her actions. This assumption, however, is not compatible with the evidence in favor of the two-visual systems model of human vision, according to which neither is all vision geared towards visual experience, nor is perceptual processing at the service of object-oriented actions. We reject EBC. Instead, we accept what Clark (2001) dubs the assumption of experience-based selection (EBS), according to which the content of visual experience may contribute to the selection of a target of an action of prehension within a set of distractors and competitors. Once, however, the perceptual selection is performed, then the control and monitoring of the bodily movements are achieved by the visuomotor representation of the target.

References

Clark, A. (2001). Visual experience and motor action: are the bonds too tight? *Philosophical Review*, 110, 495-519.

Dretske, F. (1969) *Seeing and Knowing*. Chicago: Chicago University Press.

Driver, J. & Vuilleumier (2001) Perceptual awareness and its loss to unilateral neglect and extinction. In Dehaene, S. & Naccache, L. (eds.) *The Cognitive Neuroscience of Consciousness*, Cambridge, Mass.: MIT Press.

Franz, V.H., Fahle, M., Bülthoff, H.H. & Gegenfurtner, K.R. (2001). Effects of visual illusion on grasping. *Journal of Experimental Psychology: Human Perception and Performance*. 27, 1124-1144.

Haffenden, A.M., Schiff, K.C. & Goodale, M.A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: non-illusory effects of pictorial cues on grasp. *Current Biology*, 11, 177-181.

Millikan R.G. (1996). Pushmi-pullyu representations. In Tomberlin, J. (ed.) *Philosophical Perspectives,* vol. IX, Atascadero CA: Ridgeview Publishing.

Milner, A.D. & Goodale, M.A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Mountcastle, V.B., Lynch, J.C., Georgopoulos, A. Sakata, H. and Acuna, C. (1975). Posterior parietal association cortex of the monkey: command functions for operations within extra-personal space. *Journal of Neurophysiology*, 38, 871-908.

O'Regan, J.K. & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939-1031.

Ungerleider, L. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D.J. Goodale, M.A. and Mansfield, R.J.W. (eds.) *Analysis of visual behavior*. Cambridge, Mass.: MIT Press, pp. 549-586.