Précis of Ways of Seeing

Pierre Jacob, Marc Jeannerod

Book Symposium / Tribune du livre

Précis of *Ways of Seeing*[*]

PIERRE JACOB *Institut Jean Nicod*
MARC JEANNEROD *Institut des sciences cognitives (Lyon)*

According to common sense, and much philosophy of perception as well, human vision is a unitary psychological activity, whose single purpose is to yield a unified conscious picture of the visible features of the world. Many automatic behaviours (e.g., the control of human posture), however, depend on visual mechanisms, whose function is not to make an individual visually aware of the world. Furthermore, much empirical work in the cognitive neuroscience of vision in the last forty years has also cast doubt on the assumption that the single purpose of human vision is the construction of a unified picture of the visible world by supporting the so-called "two-visual-systems" model. According to this model, there exists

* This is a discussion of Pierre Jacob and Marc Jeannerod, *Ways of Seeing: The Scope and Limits of Visual Cognition* (Oxford: Oxford University Press, 2003). All otherwise unattributed page numbers are to this book.
Cette discussion porte sur le livre de Pierre Jacob et Marc Jeannerod, *Ways of Seeing: the Scope and Limits of Visual Cognition*, Oxford, Oxford University Press, 2003. Sauf mention contraire, les numéros de page entre parenthèses renvoient tous à cet ouvrage.

Dialogue **XLV (2006), 1-**

in the visual cortex of non-human and human primates an anatomical bifurcation between two dissociable pathways: the ventral pathway and the dorsal pathway. The former projects the primary visual areas onto the infero-temporal lobe. The latter projects the primary visual areas onto the parietal lobe (which further leads to the premotor and the motor cortex). The two-visual-systems model is based on empirical evidence, including single-cell recordings in the brains of monkeys, the neuropsychological examination of human patients containing a lesion in their visual cortex, and psychophysical experiments performed on healthy human adults. Of course, the scientific question raised by this body of empirical work is: what is the functional significance of this anatomical segregation? The major aim of *Ways of Seeing* is to provide a framework that can both provide a satisfactory answer to the scientific question and elucidate some of the ensuing philosophical issues: what is the meaning of the English verb "to see?" How does subjective visual experience yield objective knowledge of the world? How does conscious visual experience relate to thought? What is the contribution of phenomenally conscious visual experience to visually guided actions?

Primates in general, and humans in particular, are endowed with hands whose thumbs are separated from other fingers, a feature which enables them to dexterously manipulate objects. Thus, many human actions directed towards objects are hand actions: for example, humans can grasp the handle of a hammer with full-power grip or the handle of a teacup with precision grip. Until recently, there were two major competing versions of the two-visual-systems model of the primate visual system. First, Ungerleider and Mishkin (1982), who initially discovered the anatomical segregation in the visual cortex of non-human primates, thought of the ventral pathway in terms of the *What*-stream (specialized in object-identification) and of the dorsal pathway in terms of the *Where*-stream (specializing in the spatial localization of objects). A few years later, mostly on the basis of the examination of the effects of brain lesions on the visual abilities of human patients, Milner and Goodale (1995) agreed that the ventral stream underlies object-identification (or vision-for-perception); however, they argued that the dorsal stream is at the service of the visuomotor transformation (i.e., the visual guidance of hand actions directed towards objects).

In *Ways of Seeing*, we draw a basic functional distinction between two kinds of visual processing of objects: semantic and pragmatic processing. The former's purpose is the visual identification and recognition of objects; the latter's purpose is the visual guidance of actions towards objects. In healthy human beings, the two kinds of processing cooperate harmoniously. But this harmony can be disrupted either by brain lesions or by careful experimental design. We argue that, according to whether the task is perceptual or visuomotor, some visual stimuli can be processed

either semantically or pragmatically. We further argue that both the semantic and the pragmatic processing of objects can occur at various levels. Visual-form apperceptive agnosic patients, who cannot process the shape of visually presented objects, are impaired at a deeper level of semantic processing than are associative agnosic patients, who can represent the shape of a visually perceived object, but cannot match their representation of the shape of an object onto their concept (or onto the name) of the object stored in memory. Optic ataxic patients, who cannot accurately reach and grasp objects, are more deeply impaired in the pragmatic processing of the target of a visually guided action than are apraxic patients, whose visuomotor transformation is intact, but who cannot recognize the function of a tool, use it, or understand others' actions with tools.

In a nutshell, we agree with Milner and Goodale that Ungerleider and Mishkin's model is wrongly predicated on the assumption that the single function of the primate visual system is to build a unified conscious picture of the visible world. Against Ungerleider and Mishkin's model, Milner and Goodale were right to make room for the visuomotor transformation within the human visual cortex. However, we criticize Milner and Goodale's model in three respects: first, on the basis of the neuropsychological examination of neglect patients (with a lesion in their right inferior parietal lobe), we argue that the inferior parietal lobe (which is part of the dorsal pathway) is involved in the perception of spatial relations among perceived objects and that, therefore, Milner and Goodale were wrong in hypothesizing that activity in the ventral stream alone is both necessary and sufficient for the conscious visual perception of objects. Second, we argue that, by restricting visually guided action to the visuomotor transformation (i.e., to reaching, pointing towards, and grasping a target), Milner and Goodale vastly underestimate the scope and complexity of the pragmatic processing of visual inputs, which they reduce to its lowest level (the visuomotor transformation), at the expense of higher levels (such as the use and recognition of complex tools and the perception and understanding of actions involving the use of tools). Finally, we point out that human vision is not restricted to the perception of objects that they can also manipulate. Humans can perceive many things that they cannot manipulate, e.g., shadows, holes, gases, clouds, events and actions. In particular, they can see human agents act. But, we argue, seeing these things falls beyond the scope of the two-visual-systems hypothesis.

In *Ways of Seeing*, we endorse a broadly representationalist framework for elucidating the nature and content of both visual perception and visually guided actions. The semantic processing of a visual object gives rise to a visual percept. At the lowest level of semantic processing, what Dretske (1969) would call "non-epistemic" visual percepts, are visual representations with non-conceptual content. At a higher level of semantic

processing, epistemic visual percepts make visual information about the world available for conceptual thoughts (i.e., what philosophers call "the belief box"). The lowest level of pragmatic processing gives rise to a visuomotor representation of a target of action. We argue that some psychophysical experiments on size-contrast illusory displays provide evidence for the existence of a dissociation, in healthy human adults, between perceptual responses and visuomotor responses. We further argue that these experiments show that the visuomotor system can be fooled and made to misrepresent two-dimensional features of the display as three-dimensional properties. If visuomotor representations (that arise from low-level pragmatic processing) can misrepresent a visual display, it follows that they are genuine mental representations—or so we argue. Roughly speaking, we argue that a visuomotor representation of an object is the representation of what Gibson (1979) would call an affordance for action. The purpose of a visuomotor representation of a target of action is to provide visual information for the benefit of an agent's motor intention. Whereas beliefs have a mind-to-world direction of fit, intentions and desires have a world-to-mind direction of fit. Percepts that arise from higher-level semantic processing share the mind-to-world direction of fit of beliefs. We argue that visuomotor representations, like Millikan's (1996) "pushmi-pullyu" representations, are hybrid mental representations: they represent facts in a format suitable, not for forming beliefs (acquiring knowledge) about them, but for informing an agent's motor intention about how to act on an object.

One major task, then, of *Ways of Seeing* is to provide a principled distinction between the non-conceptual content of non-epistemic visual percepts and the non-conceptual content of visuomotor representations. Representing the colour of an object may be a crucial condition for identifying it, but not for grasping it. We argue that the main difference between the content of a visuomotor representation of an object and the content of a visual percept (at either a lower or higher level of semantic processing) of the same object lies in the representation of the object's spatial position. The job of a visual percept is object-identification and recognition: it must, therefore, convey visual information about objects to the human long-term memory and reasoning systems. To fill its purpose, it must encode information about the enduring properties of objects that are relevant for object-recognition over time. For an object to be identified and recognized at different times, in different lighting conditions or from different spatial perspectives, its enduring properties must be represented in a format that is object-dependent and viewer-independent. The job of a visuomotor representation is to track the visual features of a target that are of immediate relevance for an agent's grasping (or pointing towards) it. An agent cannot reach and grasp (or point to) a target unless her visual system computes its spatial location in an egocentric frame of reference.

Thus, for a visuomotor representation of a target to fill its function, it must represent the target's spatial location relative to the agent's egocentric frame of reference (i.e., centered on the agent's body).

Suppose now that you are faced with a basket of fruits containing pears, apricots, peaches, bananas, oranges, a melon, and a collection of red, yellow, and green apples. Suppose that you want to eat a red apple. If so, then you must first be able to reach it. You will not be able to reach it unless you represent its location in egocentric coordinates. If you reach it, then you will not be able to grasp it accurately unless you represent its shape, size, and orientation. This is what a visuomotor representation of the apple in the basket is good for: representing the absolute shape, size, and orientation of a target, whose location is represented in egocentric coordinates. But you could not form a visuomotor representation of your target in the basket unless you had visually selected it from the set of distractors. In order to do so, you must represent the shape, size, texture, and colour of a particular red apple and compare it to the shapes, sizes, textures, and colours of potentially competing targets. This is what a visual percept is good for. Now, you will not be able to form such a visual percept of the relative shape, size, texture, and colour of the relevant red apple unless you represent its location in some allocentric frame of reference centered on some item in the visual array. The fine-grained non-conceptual content of the visual percept of a selected red apple outstrips the conceptual content of a belief about the very same red apple that it is of a kind that makes good cider. The non-conceptual content of the percept is involved in the selection of the target of an action of prehension, but it does not guide the fine-tuning of the hand movement. On the basis of a visual epistemic percept of the content of the basket, what you might do is count the number of apples and come to form beliefs about the relative sets of different fruits in the basket—perhaps that there are more pears than apples or that there are more apricots than any other fruits there.

In-depth examination of deeply apperceptive visual-form-agnosic patient DF shows that, although she is unable to recognize the shape of visually presented objects, she nonetheless is able to reach for and grasp objects with close to normal precision grip. DF can form visuomotor representations, but she fails to form visual percepts of objects. She fails to be phenomenally aware of the size, shape, and orientation of the objects that she can grasp. Since she can grasp it, DF can represent the size, shape, and orientation of an object for the purpose of reaching for and grasping it, but not for recognizing it. She can do so only if and when she represents its spatial position in an egocentric frame of reference, even if not in an allocentric frame of reference. In *Ways of Seeing*, we assume that a necessary condition for perceptual (or phenomenal) awareness of an object is the ability to identify and recognize its visual features. We further assume that, unless a creature has the resources to make contrastive identifica-

tions among distinct instances of one and the same visual feature, she will not be able to recognize or identify it (the constraint of contrastive identification). Now, in a visuomotor representation of a target of action, the spatial position of the target is represented in an egocentric frame of reference centered on the agent's body. In a visual percept, the spatial position of the object is represented in an allocentric frame of reference. We argue that unless an object's spatial position is represented in some allocentric frame of reference (enabling the representation of spatial relations among several objects) one cannot become phenomenally aware of such visual attributes of an object as its size, shape, and orientation. Only if the spatial position of an object is represented in an allocentric frame of reference centered on some item in the visual array can the relative size, shape, and orientation of an object be available for comparison with the relative sizes, shapes, and orientations of neighbouring objects.

In a nutshell, the general picture espoused in *Ways of Seeing* is the following. Because thoughts, which have conceptual content, are both systematic and productive, they can rise above the limitations of perception and represent entities that are not in space. Because they can represent the spatial relations among objects in some allocentric frame of reference, visual percepts satisfy the constraint of contrastive identification and they can make one visually aware of objects in space by representing their relative visual features. Because they represent the spatial position of objects in egocentric coordinates, visuomotor representations enable an agent to act on a target. But by representing its absolute size, shape, and orientation, visuomotor representations fail to satisfy the constraint of contrastive identification, and thus, as such, they fail to make the agent phenomenally aware of the target's visual features.

References

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

Gibson, J. J.
1979 *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Millikan, R. G.
1996 "Pushmi-Pullyu Representations." In *Philosophical Perspectives*, Vol. 9: 185-200. Edited by J. Tomberlin. Atascadero, CA: Ridgeview.

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

Ungerleider, L., and M. Mishkin
1982 "Two Cortical Visual Systems." In *Analysis of Visual Behavior*. Edited by D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield. Cambridge, MA: MIT Press, pp. 549-86.