COGNITION

Cognition 67 (1998) 181–207

The objects of action and perception

Melvyn A. Goodale*, G. Keith Humphrey

Department of Psychology, University of Western Ontario, London, ON N6A 5C2, Canada

Abstract

Two major functions of the visual system are discussed and contrasted. One function of vision is the creation of an internal model or percept of the external world. Most research in object perception has concentrated on this aspect of vision. Vision also guides the control of object-directed action. In the latter case, vision directs our actions with respect to the world by transforming visual inputs into appropriate motor outputs. We argue that separate, but interactive, visual systems have evolved for the perception of objects on the one hand and the control of actions directed at those objects on the other. This 'duplex' approach to high-level vision suggests that Marrian or 'reconstructive' approaches and Gibsonian or 'purposive-animate-behaviorist' approaches need not be seen as mutually exclusive, but rather as complementary in their emphases on different aspects of visual function. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Vision; Action; Perception

1. Introduction

It is a common assertion that the fundamental task of vision is to construct a representation of the three-dimensional layout of the world and the objects and events within it. But such an assertion begs at least two fundamental and interrelated questions. First, what is vision? Second, what is the nature of the representation that vision delivers? These questions, which are central to the entire research enterprise in understanding human vision, form the framework for the present paper. In attempting to answer these questions, we will contrast what we believe are two major functions of the visual system. One function of vision is the creation of an internal model or percept of the external world – a model that can be used in the

---

* Corresponding author. Tel.: +1 519 6612070; fax: +1 519 6613961; e-mail: mgoodale@julian.uwo.ca

PII S0010-0277(98)00017-1

recognition of objects and understanding their interrelations. Most research in object vision has concentrated on this function (witness the current volume). There is another function of vision, however, which is concerned not with object recognition, but with object-directed action. In this case, vision guides our actions with respect to the world by transforming visual inputs into appropriate motor outputs. We will suggest that separate, but interacting, visual systems have evolved for the perception of objects on the one hand and the control of actions directed at those objects on the other. This 'duplex' approach to high-level vision suggests that Marrian or 'reconstructive' approaches and Gibsonian or 'purposive-animate-behaviorist' approaches need not be mutually exclusive and may be actually complementary.

2. What is vision?

Vision gives us sight. In other words, vision gives us an experience of the world beyond our immediate body surface, a world full of objects and events that are imbued with meaning and significance. Research in human psychophysics and perception has concentrated almost entirely on the way in which the visual system delivers this visual experience (for related discussions of this issue see Georgeson, 1997; Watt, 1991, 1992). Although a good deal of this research has concentrated on 'low-level' visual computations, even here it has been generally assumed that the mechanisms supporting such computations are all part of the same general-purpose system dedicated to the construction of the visual percept. This fascination with what and how we 'see' has meant that many other functions of vision have either been ignored or been assumed to depend on the same mechanisms supporting sight. This preoccupation with vision as sight was nicely described 20 years ago by Weimer (1977):

> Since the time of Aristotle the mind has been regarded as intrinsically sensory in nature, as a passive black box or window that is (somehow) sensibly impressed with input from the environment. A root metaphor of mind has evolved from the common-sense, everyday experience of looking at the world. Vision, conceived as the passive reception of information that both exists and possesses an intrinsic psychological character independently of the organism, became the paradigm exemplar of mental processing (p. 268).

For most people then vision is synonymous with sight; there is nothing more to vision than visual experience. Even Marr, who was perhaps the most influential visual theorist in recent years, appears to endorse this 'plain man's' conception of vision (see p. 3 of Marr, 1982). Yet there is plenty of evidence that much of the work done by the visual system has nothing to do with sight or experiential perception. The pupillary light reflex, the synchronization of circadian rhythms with the local light-dark cycle, and the visual control of posture are but three examples of a range of visually modulated outputs where we have no direct experience of the controlling stimuli and where the underlying control mechanisms have little to do with our perception of the world. Yet most contemporary accounts of vision, while acknowledging the existence of these ‘extraperceptual’ visual phenomena, still assume that the main function of the visual system is the construction of some sort of internal model or percept of the external world (for a detailed discussion of this issue, see Goodale, 1983a, 1988, 1997). In such accounts, phenomena such as the pupillary light reflex are seen as simple servomechanisms which, while useful, are not part of the essential machinery for the construction of the visual percept. But, as we shall see later, the visual control of much more complex behaviours, such as reaching out and grasping an object, also appear to depend on mechanisms that are functionally and neurally separate from those mediating our perception of that object. Indeed, the origins of vision may be related more to its contribution to the control of action than to its role in conscious perception, a function which appears to be a relative newcomer on the evolutionary scene (Goodale, 1983a, 1988; Goodale et al., 1996).

2.1. Vision for acting on the world

Vision in many animals can be studied without appealing to the idea of vision as sight. The reason for this, of course, is that vision evolved in animals, not to enable them to ‘see’ the world, but to guide their movements through it. Indeed, the visual system of most animals, rather than being a general-purpose network dedicated to reconstructing the rather limited world in which they live, consists instead of a set of relatively independent input-output lines, or visuomotor ‘modules’, each of which is responsible for the visual control of a particular class of motor outputs.

While evidence for separate visuomotor modules can be found in a broad range of anatomical, electrophysiological, and behavioral studies, some of the most compelling demonstrations have been provided by experiments with so-called ‘rewired’ frogs. Because the amphibian brain is capable of far more regeneration following damage than the mammalian brain, it is possible to ‘re-wire’ some retinal projections, such as those going to the optic tectum in the midbrain, while leaving all the other retinal projections intact. Thus, the retinotectal projections can be induced to project to the optic tectum on the same side of the frog’s brain instead of to the optic tectum on the opposite side, as is the case in the normal animal. In one such experiment, these unfortunate creatures were shown to demonstrate ‘mirror-image’ feeding – directing their snapping movements to positions in space that were mirror-symmetrical to the location of prey objects (Ingle, 1973). They also showed mirror-image predator avoidance and jumped towards rather than away from the looming visual stimuli. These results suggest that the optic tectum plays a critical role in the visual control of these patterns of behavior in the frog. Remarkably, however, the same ‘rewired’ frogs showed quite normal visually-guided barrier avoidance as they locomoted from one place to another, even when the edge of the barrier was placed in the visual field where mirror-image feeding and predator avoidance could be elicited. As it turns out, the reason they showed normal visual control of barrier avoidance is quite straightforward; the retinal projections to the pretectum, a structure in the thalamus just in front of the optic tectum, were still intact and had not been redirected to the opposite side of the brain. A number of lesion studies have shown that this structure plays a critical role in the visual control of barrier avoidance (Ingle, 1980, 1982). Thus, it would appear that there are at least two independent visuomotor systems in the frog: a tectal system, which mediates visually elicited prey-catching and predator-avoidance, and a pretectal system which mediates visually guided locomotion around barriers. In fact, more recent work suggests that there may be upwards of five or more distinct visuomotor networks in the amphibian brain, each with its own set of retinal inputs and each controlling different arrays of motor outputs (Ewert, 1987; Ingle, 1991).

The results of such studies, which point to a good deal of modularity in the organization of the visuomotor circuitry in the frog, do not fit well with the common view of a visual system dedicated to the construction of a general-purpose representation of the external world. Although the outputs from the different visuomotor systems described above need to be coordinated, it makes no sense to argue that the different actions controlled by these networks are guided by a single visual representation of the world residing somewhere in the animal's brain. Of course, the idea of separate visuomotor channels is consistent with the views of some visual theorists who have argued that vision does more than mediate perception and subserves the visual control of many the different actions that organisms carry out in their daily lives. 'Purposive vision', as this approach is sometimes described, has emphasized the role of vision in the direct control of actions rather than its contribution to constructing percepts of the world in which those actions might unfold (e.g. Aloimonos, 1990).

While there is certainly plenty of evidence to suggest that visuomotor modularity of the kind found in the frog also exists in the mammalian brain (e.g. Ellard and Goodale, 1986, 1988; Goodale, 1983b, 1996; Goodale and Carey, 1990; Goodale and Milner, 1982), the very complexity of day-to-day living in many mammals, particularly in higher primates, demands much more flexible organization of the circuitry. In monkeys (and thus presumably in humans as well), there is evidence that many of the phylogenetically ancient visuomotor circuits that were present in more primitive vertebrates are now modulated by more recently evolved control systems in the cerebral cortex (for review, see Milner and Goodale, 1995). Thus, the highly adaptive visuomotor behavior of humans and other higher primates is made possible by the evolution of another layer of control in a series of hierarchically organized networks. This idea is reminiscent of the views of John Hughlings Jackson (e.g. Jackson, 1875), an eminent nineteenth-century British neurologist who was heavily influenced by concepts of evolution. Jackson tried to explain the effects of damage to human cerebral cortex by suggesting that such damage removed the more highly evolved aspects of brain function, so that what one saw in the performance of many patients was the expression of evolutionarily older mechanisms residing elsewhere in the brain. The emergence of more flexible visuomotor control has not been accomplished entirely by cortical modulation of older circuitry however. The basic subcortical circuitry has itself changed to some extent and new visuomotor control systems have also emerged in which visual control of an almost limitless range of motor outputs is possible. Nevertheless, as we shall see later, for the most part, these networks have remained functionally and neurally separate from those mediating our visual perception of the world.

2.2. Vision for perceiving the world

Although the need for more flexible visuomotor control was one of the demands on the evolving primate brain, another was related to the need to identify the objects, to understand their significance and causal relations, to plan a course of action, and to communicate with other members of the species. In short, the emergence of cognitive systems and complex social behavior created a whole new set of demands on vision and the organization of the visual system. Direct sensory control of action was not enough. As interactions with the world become more complicated and subtle, motor outputs became quite arbitrary with respect to sensory input. In fact, many animals particularly humans and other primates, behave as though their actions are driven by some sort of internal model of the world in which they live. The representational systems that use vision to generate such models or percepts of the world must carry out very different transformations on visual input than the transformations carried out by the visuomotor modules described earlier (the nature of these differences will be explored later). Moreover, these systems, which generate our perception of the world, are not linked directly to specific motor outputs but are linked instead to cognitive systems involving memory, semantics, spatial reasoning, planning, and communication. But even though such higher-order representational systems permit the formation of goals and the decision to engage in a specific act without reference to particular motor outputs, the actual execution of an action may nevertheless be mediated by dedicated visuomotor modules that are not dissimilar in principle from those found in frogs and toads. In summary, vision in humans and other primates (and perhaps in other animals as well) has two distinct but interactive functions: (1) the perception of objects and their relations, which provides a foundation for the organism’s cognitive life, and (2) the control of actions directed at (or with respect to) those objects, in which specific sets of motor outputs are programmed and guided ‘on-line’.

3. Action and perception systems in the primate brain: dorsal and ventral streams

The evolution of separate systems for visual perception and for the visual control of action is reflected in the organization of the visual pathways in the primate cerebral cortex. Over fifteen years ago, Ungerleider and Mishkin (1982) identified two distinct ‘streams of processing’ in the macaque monkey brain: a so-called ventral stream projecting from primary visual cortex to inferotemporal cortex and a so-called dorsal stream projecting from primary visual cortex to posterior parietal cortex (Fig. 1). Although one must always be cautious when drawing homologies between monkey and human neuroanatomy (Crick and Jones, 1993), it seems likely that the visual projections from the primary visual cortex to the temporal and parietal

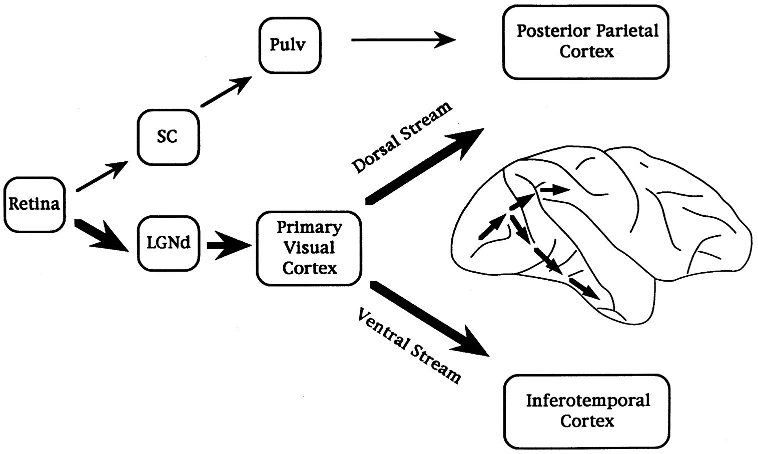

Fig. 1. Major routes whereby retinal input reaches the dorsal and ventral streams. The diagram of the macaque brain (right hemisphere) on the right of the figure shows the approximate routes of the cortico-cortical projections from the primary visual cortex to the posterior parietal and the inferotemporal cortex respectively. LGNd, lateral geniculate nucleus, pars dorsalis; Pulv, pulvinar; SC, superior colliculus.

lobes in the human brain may involve a separation into ventral and dorsal streams similar to that seen in the macaque brain. Ungerleider and Mishkin (1982) suggested, on the basis of a number of electrophysiological studies in the monkey, that the ventral stream plays a critical role in object vision, enabling the monkey to identify an object, while the dorsal stream is involved in spatial vision, enabling the monkey to localize the object in space. This interpretation, in which a distinction is made between identification and localization, is similar to an earlier functional dichotomy proposed by Schneider (1969), who argued that primary visual cortex plays an essential role in identifying visual stimuli while the more ancient midbrain structure, the superior colliculus (another name for the optic tectum in mammals), is responsible for localizing the stimulus. Ungerleider and Mishkin (1982) have taken this same distinction and moved it into the cerebral cortex. More recently, however, Goodale and Milner (1992) (and Milner and Goodale, 1995) have offered a re-interpretation of the apparent differences in the visual processing carried out by the two streams of processing emanating from primary visual cortex. Rather than emphasizing differences in the visual information handled by the two streams (object vision versus spatial vision or ‘what’ versus ‘where’), their account has instead focused on the difference in the requirements of the output systems that each stream of processing serves.

According to Goodale and Milner, the ventral stream plays the major role in constructing the perceptual representation of the world and the objects within it, while the dorsal stream mediates the visual control of actions directed at those objects (for a more detailed discussion, see Goodale and Milner, 1992; Milner and Goodale, 1995). In other words, processing within the ventral stream allows the monkey to recognize an object, such as a ripe piece of fruit dangling from a tree, while processing within the dorsal stream provides critical information about the location, size, and shape of that fruit so that the animal can accurately reach out and grasp it with its hand or mouth. Notice that in this account, information about object attributes, such as size, shape, orientation, and spatial location, are processed by both streams but the nature of that processing is very different. The functional distinction is not between 'what' and 'where', but between the way in which the visual information about a broad range of object parameters are transformed either for perceptual purposes or for the control of goal-directed actions. This is not to say that the distribution of retinogeniculate inputs does not differ between the two streams, but rather that the main difference lies in the nature of the transformations that each stream performs on those two sets of inputs.

3.1. Neuropsychological studies of the dorsal stream

In the intact brain, the two streams of processing work together in a seamless and unified fashion. Nevertheless, by studying individuals who have sustained brain damage that spares one of these systems but not the other, it is possible to get a glimpse of how the two streams differ in the way they each deal with incoming visual information. For example, patients who have sustained damage to the superior portion of the posterior parietal cortex, the major terminus of the dorsal stream, are unable to use visual information to reach out and grasp objects in the hemifield contralateral to the lesion. Clinically, this deficit is called optic ataxia (Bálint, 1909). Such patients have no difficulty using other sensory information, such as proprioception, to control their reaching; nor do they usually have difficulty recognizing or describing objects that are presented in that part of the visual field. Thus, their deficit is neither 'purely' visual nor 'purely' motor; it is a visuomotor deficit.

Observations in several laboratories have shown that patients with optic ataxia not only have difficulty reaching in the correct direction, but they also show deficits in their ability to adjust the orientation of their hand when reaching toward an object, even though they have no difficulty in verbally describing the orientation of the object (e.g. Perenin and Vighetto, 1988). Such patients can also have trouble adjusting their grasp to reflect the size of an object they are asked to pick up – although again their perceptual estimates of object size remain quite accurate (Jakobson et al., 1991; Goodale et al., 1993). To pick up an object successfully, however, it is not enough to orient the hand and scale the grip appropriately; the fingers and thumb must be placed at appropriate opposition points on the object's surface. To do this, the visuomotor system has to compute the outline shape or boundaries of the object. In a recent experiment (Goodale et al., 1994b), a patient (RV) with bilateral lesions of the occipitoparietal region, was asked to pick up a series of small, flat, non-symmetrical smoothly contoured objects using a precision grip, which required her to place her index finger and thumb in appropriate positions on either side of each object. If the fingers were incorrectly positioned, the computation of the correct opposition points ('grasp points') can be achieved only if the overall shape or form of the object is taken into account. Despite the fact that the patient could readily distinguish these objects from one another, she often failed to place her fingers on the appropriate grasp points when she attempted to pick up the objects (Fig. 2).

Such studies suggest that it is not only the spatial location of the object that is apparently inaccessible for controlling movement in patients with dorsal-stream lesions, but the intrinsic characteristics of the object as well. It would be incorrect to characterize the deficits in these patients simply in terms of a disturbance of spatial vision. In fact, in one clear sense their 'spatial vision' is quite intact, since they can often describe the relative location of objects in the visual field contralateral to their lesion, even though they cannot pick them up (Jeannerod, 1988). This pattern of deficits is quite consistent with Goodale and Milner's proposal that the dorsal stream plays a critical role in the visuomotor transformations required for skilled actions, such as visually guided prehension – in which the control of an accurate grasp requires information about an object's location as well as its orientation, size, and shape. It should be emphasized, however, that not all patients with damage to the posterior parietal region have difficulty shaping their hand to correspond to the structural features and orientation of the target object. Some have difficulty with hand postures, some with controlling the direction of their grasp, and some with foveating the target. Indeed, depending upon the size and locus of the lesion, a patient can demonstrate any combination of these visuomotor deficits (for review, see Milner and Goodale, 1995). Different sub-regions of the posterior parietal cor-

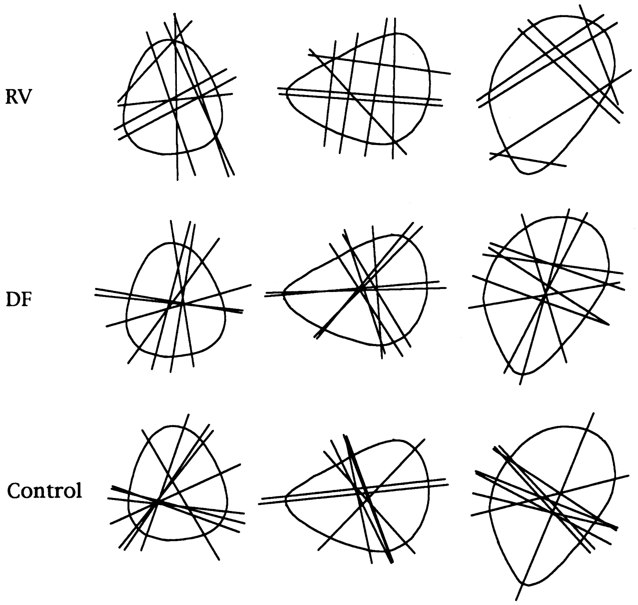

Fig. 2. The 'grasp lines' (joining points where the index finger and the thumb first made contact with the shape) selected by the optic ataxic patient (RV), the visual form agnosic patient (DF), and the control subject when picking up three of the twelve shapes. The four different orientations in which each shape was presented have been rotated so that they are aligned. No distinction is made between the points of contact for the thumb and finger in these plots.

tex, it appears, support transformations related to different motor outputs. Such modularity within the dorsal stream means that a particular skilled action would invoke certain combinations of these visuomotor networks and other actions would invoke quite different combinations.

3.2. Neuropsychological studies of the ventral stream

As we have just seen, patients with optic ataxia often have difficulty reaching towards and/or picking up objects that they have no difficulty identifying. Are there patients who show the opposite pattern of visual deficits and spared visual abilities? In other words, are there patients who can grasp objects quite accurately despite their failure to recognize what it is they are attempting to pick up? One such patient is DF, a young woman who developed a profound visual form agnosia following near-asphyxiation by carbon monoxide. Not only is she unable to recognize the faces of her relatives and friends or the visual shape of common objects, but she is also unable to discriminate between such simple geometric forms as a triangle and a circle. DF has no difficulty identifying people from their voices and she has no problem identifying objects placed in her hands. Her perceptual problems are exclusively visual. Moreover, her deficit, seems largely restricted to the form of objects. She can use color and other surface features to identify objects and she can even use shape from shading to some extent (Humphrey et al., 1994, 1996; Servos et al., 1993). What she seems unable to perceive are the contours of objects – no matter how the contours are defined (Milner et al., 1991). Thus, she cannot identify, shapes whose contours are defined by differences in luminance or color, or by differences in the direction of motion or the plane of depth. Not surprisingly, DF is also unable to recognize shapes that are defined by the similarity or proximity of individual elements of the visual array. A selective deficit in form perception with spared color and other surface information is characteristic of the severe visual agnosia that sometimes follows an anoxic episode. Although MRI shows a pattern of diffuse brain damage in DF that is consistent with anoxia, most of the damage was evident in the ventrolateral region of the occipital lobe sparing primary visual cortex.

The profound deficit in DF’s form perception cannot be explained by disturbances in ‘low-level’ sensory processing. In perimetry testing, she was able to detect luminance-defined targets at least as far out as 30 degrees (Milner et al., 1991). Her spatial contrast sensitivity also appeared to be normal above about 10 cycles/degree and was only moderately impaired at lower spatial frequencies (of course, even though she could detect the presence of the gratings used to measure her contrast sensitivity, she could not report their orientation; see also Humphrey et al., 1991). But the most compelling reason to doubt that her perceptual deficit is due to some sort of low level disturbance in processing is the fact that in another domain, visuomotor control, she remains exquisitely sensitive to the form of objects! Thus, despite her inability to recognize the shape, size, and orientation of objects, she shows strikingly accurate guidance of hand and finger movements directed at those very same objects. Thus, when she was presented with a pair of rectangular blocks of the same or different dimensions, she was unable to distinguish between them. Even when she was asked to indicate the width of a single block by means of her index finger and thumb, her matches bore no relationship to the dimensions of the object and showed considerable trial to trial variability. In contrast, when she was asked simply to reach out and pick up the block, the aperture between her index finger and thumb changed systematically with the width of the object as the movement unfolded, just as in normal subjects (Goodale et al., 1991). In other words, DF scaled her grip to the dimensions of the object she was about to pick up, even though she appeared to be unable to perceive those object dimensions.

A similar dissociation was seen in DF’s responses to the orientation of stimuli. Thus, when presented with a large slot which could placed in one of a number of different orientations, she showed great difficulty in indicating the orientation of the slot either verbally or even manually by rotating a hand-held card (see Fig. 3, left). Nevertheless, when she was asked simply to reach out and insert the card, she performed as well as normal subjects, rotating her hand in the appropriate direction as soon as she began the movement (see Fig. 3, right). Finally, even though DF could not discriminate between target objects that differed in outline shape, she could nevertheless pick up such objects successfully, placing her index finger and thumb on stable grasp points (see Fig. 2).

Findings such as these are difficult to reconcile with the idea of Ungerleider and Mishkin (1982) that object vision is the preserve of the ventral stream – for here we have a patient in whom a profound loss of object perception exists alongside a preserved ability to use object features such as size, outline shape, and orientation to guide skilled actions. Such a dissociation, of course, is consistent with the idea

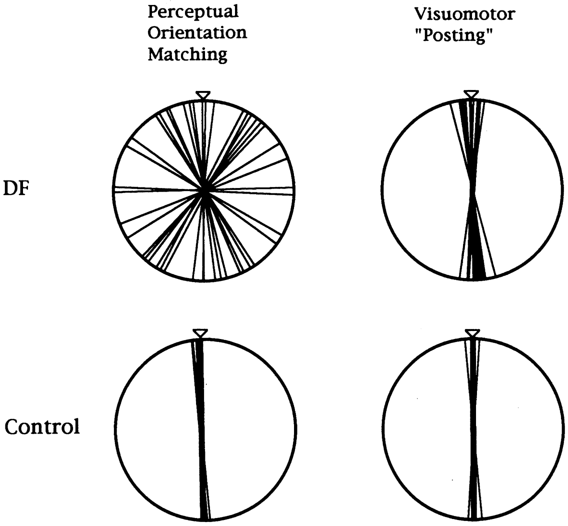

Fig. 3. Polar plots of the orientation of the hand-held card when DF and a control subject were each asked to rotate the card to match the orientation of the slot (left column) or to ‘post’ the card into the slot (right column). The orientation of the card on the visuomotor task was measured at the instant before the card was placed in the slot. In both plots, the actual orientations of the slot have been normalized to vertical.

proposed by Goodale and Milner (1992) that there are separate neural pathways for transforming incoming visual information for action and perception. Presumably it is the latter and not the former that is compromised in DF. In other words, the brain damage that she suffered as a consequence of anoxia appears to have interrupted the normal flow of shape and contour information into her perceptual system without affecting the processing of shape and contour information by the visuomotor modules comprising her action system. If, as Goodale and Milner have suggested, the perception of objects and events is mediated by the ventral stream of visual projections to inferotemporal cortex, then DF should show evidence for damage relatively early in this pathway. Certainly, the pattern of damage revealed by MRI is consistent with this interpretation; the major focus of cortical damage is in the ventrolateral region of the occipital cortex, an area that is thought to be part of the human homologue of the ventral stream. Primary visual cortex, which provides input for both the dorsal and ventral streams, appears to be largely intact. Thus, although input from primary visual cortex to the ventral stream may have been compromised in DF, input from this structure to the dorsal stream appears to be essentially intact. In addition, the dorsal stream, unlike the ventral stream, also receives input from the superior colliculus via the pulvinar, a nucleus in the thalamus (see Fig. 1). Input to the dorsal stream from both the superior colliculus (via the pulvinar) and the lateral geniculate nucleus (via primary visual cortex) could continue to mediate well-formed visuomotor responses in DF.

Nevertheless, it must not be forgotten that DF’s problems arose, not from a discrete lesion, but from anoxia. Therefore, the brain damage in DF, while localized to some extent, is much more diffuse than it would be in a patient with a stroke or tumour. For this reason, any attempt to map the striking dissociation between perceptual and visuomotor abilities in DF onto the ventral and dorsal streams of visual processing must be regarded as tentative. The proposal is strengthened, however, by observations in the patients described earlier whose pattern of deficits is complementary to DF’s and whose brain damage can be confidently localized to the dorsal stream.

4. Electrophysiological and behavioural studies in the monkey

The functional division of labour between the two streams proposed by Goodale and Milner is also supported by a large number of studies in the macaque monkey. Thus, monkeys which show profound deficits in object recognition following inferotemporal lesions are nevertheless as capable as normal animals at picking up small food objects (Klüver and Bucy, 1939), at catching flying insects (Pribram, 1967), and at orienting their fingers in a precision grip to grasp morsels of food embedded in small slots placed at different orientations (Buchbinder et al., 1980). In short, these animals behave much the same way as DF: they are unable to discriminate between objects on the basis of visual features that they can clearly use to direct their grasping movements. In addition, there is a long history of electrophysiological work showing that cells in this area are tuned to specific objects or object features.

Moreover, the responses of these cells are not affected by the animal's motor behavior, but are instead sensitive to the reinforcement history and significance of the visual stimuli that drive them (for review, see Goodale, 1993; Milner and Goodale, 1995). Indeed, sensitivity to particular objects can be created in ensembles of cells in inferotemporal cortex simply by training the animals to discriminate between different objects (Logothetis et al., 1995). Finally, there is evidence for a specialization within separate regions of the ventral stream for the coding of certain categories of objects, such as faces and hands, which are of particular social significance to the monkey (for review, see Logothetis and Sheinberg, 1996; Perrett et al., 1995).

In contrast to cells in the ventral stream, most visually-sensitive cells in the dorsal stream are modulated by the concurrent motor behavior of the animal (e.g. Hyvärinen and Poranen, 1974; Mountcastle et al., 1975). In reviewing the electrophysiological studies that have been carried out on the posterior parietal cortex, Andersen (1987) concluded that most neurons in these areas 'exhibit both sensory-related and movement-related activity'. The activity of some visually-driven cells in this region have been shown to be linked to saccadic eye movements; the activity of others to whether or not the animal is fixating a stimulus; and the activity of still other cells to whether or not the animal is engaged in visual pursuit or is making goal-directed reaching movements (e.g. Snyder et al., 1997). Some cells in the posterior parietal area that fire when monkeys reach out to pick up objects are selective not for the spatially directed movement of the arm, but for the movements of the wrist, hand, and fingers that are made prior to and during the act of grasping the target (Hyvärinen and Poranen, 1974; Mountcastle et al., 1975). In a particularly interesting recent development, Sakata and his colleagues have shown that many of these so-called 'manipulation' cells are visually selective and are tuned for objects of a particular shape (Sakata et al., 1992; Taira et al., 1990; for review see Sakata and Taira, 1994; Sakata et al., 1997). These manipulation neurons thus appear to be tied to the properties of the goal object as well as to the distal movements that are required for grasping that object. Finally, it should be noted that lesions in the posterior parietal area in the monkey produce deficits in the visual control of reaching and grasping similar in many respects to those seen in humans following damage to the homologous region (e.g. Haaxma and Kuypers, 1975; Ettlinger, 1977). This review of the monkey literature is clearly far from complete. Interested readers are directed to Milner and Goodale (1993; 1995).

5. Neuro-imaging studies in humans

Ten years ago little was known about the organization of the cerebral visual pathways beyond V1 in humans. With the advent of functional neuroimaging, however, a wealth of data has suddenly become available. The careful work of Tootell et al. (1996) has revealed an organization of visual areas in the human brain that is remarkably similar to that seen in the macaque. Although clear differences in the topography of these areas emerges as one moves from monkey to human, the functional separation into a ventral occipitotemporal and a dorsal occipitoparietal pathway appears to be preserved. Thus, areas in the occipitotemporal region appear to be specialized for the processing of colour, texture, and form differences of objects (e.g. Puce et al., 1996; Price et al., 1996; Malach et al., 1995; Kanwisher et al., 1996). In contrast, regions in the posterior parietal cortex have been found that are activated when subjects engage in visually guided movements such as saccades, reaching movements, and grasping (Matsumura et al., 1996).

As in the monkey, there is evidence for specialization within the occipitotemporal and occipitoparietal visual pathways. Thus, activation studies have identified regions in the occipitotemporal pathway for the processing of faces that are distinct from those involved in the processing of other objects (Kanwisher et al., 1997; Gauthier et al., 1997). Similarly, there is evidence that different areas in and around the intraparietal sulcus are activated when subjects make saccadic eye movements as opposed to manual pointing movements towards visual targets (e g. Kawashima et al., 1996).

Thus, the neuroimaging data are consistent with the idea of two visual streams. In addition, the results of several studies indicate that areas in the posterior parietal cortex are involved in the visual control of action while areas in the occipitotemporal region appear to play a role in object recognition.

6. Differences in the visual transformations mediating action and perception

The division of labour within the organization of the cerebral visual pathways in primates reflects the two important trends in the evolution of vision in higher vertebrates that were identified earlier. First, the emergence of a dorsal ‘action’ stream reflects the need for more flexible programming and on-line control of visually guided motor outputs. It is interesting to note that this stream is intimately connected not only with the primate forebrain but also with those brainstem structures such as the superior colliculus and various pontine nuclei that play a critical role in the programming and control of movement in all vertebrates (Milner and Goodale, 1995). Thus, one way that the dorsal stream may mediate the visual control of skilled actions is by modulating the activity of these more phylogenetically ancient visuomotor networks.

Second, the emergence of a ventral ‘perception’ stream which can parse the visual array into discrete objects and events means that animals like ourselves can use perceptual representations of those objects and their relations for long-range planning, communication and other cognitive activities. Although a separate system for this kind of reconstructive visual activity is evident in the cerebral cortex of many mammals (Goodale and Carey, 1990), it is particularly well-developed in humans and other higher primates. Indeed, the ventral stream projections to the inferotemporal cortex, which is intimately connected with structures in the medial temporal lobe and prefrontal cortex involved in long-term memory and other cognitive activities, is exquisitely poised to serve as interface between vision and cognition. In short, while the dorsal stream allows us visual control of our movements through the world, it is the ventral stream that gives us sight.

The distinction between vision for perception and vision for action is similar in many respects to a distinction that Neisser (1989, 1994) has drawn between what he calls ‘recognition systems’ on the one hand and ‘direct perception systems’ on the other. According to Neisser, recognition systems mediate the identification and classification of objects through the accumulation of evidence in relation to stored representations. Recognition is always defined as a relation of present input to the past, i.e. in relation to stored information about objects. Direct perception for Neisser is something quite different. Following Gibson (1979), he proposes that the direct perception system provides information about where we are, where objects are, and what physical actions those objects can afford – information that is provided as the animal moves through the world and interacts with it. He proposes that Gibson’s concept of affordance be limited to the notion of a ‘physical affordance’ since such affordances are entirely specified by the physics of light and do not depend on stored semantic knowledge about the objects – which is the business of the recognition systems. The possibility of picking up an object such as a rock and throwing it does not depend on identifying the object but rather the ‘fit’, or physical relationship, between our effector organs and the object. In short, the task of direct perception is the programming and on-line control of action.

The type of information stored and used in recognition is quite different from that used in the control of action. One example that Neisser (1989) uses to illustrate the difference between recognition and direct perception is the way each system deals with the orientation of objects. As several papers in this volume attest to, a large amount of recent research on object orientation has been concerned with the effects of object orientation on recognition. Our recognition of an object often suffers greatly if its orientation does not match the orientation that we have experienced in the past (e.g. Edelman and Bülthoff, 1992; Humphrey and Khan, 1992; Rock and DiVita, 1987; Tarr, 1995; see also Biederman and Gerhardstein, 1995; Tarr and Bülthoff, 1995; for review see Jolicoeur and Humphrey, 1998). In sharp contrast, our ability to direct a well-formed grasp at an object is not dependent on prior familiarity with a particular orientation; in fact, we do not need to recognize the object to grasp it efficiently.

While Neisser’s distinction between recognition and direct perception converges on our own ideas to some extent, there are some critical differences. For us, the action system (similar to Neisser’s direct perception system) is entirely concerned with providing visual information for the programming and control of motor outputs. This system contains an array of dedicated visuomotor modules which, when activated in various combinations, transform visual inputs into directed motor acts. Neisser, however, suggests that our perception of the spatial location of objects and their relations is dependent on the direct perception system; the recognition system for Neisser seems to be concerned only with identifying and classifying objects. In our scheme, the visuomotor modules that make up the action system do not participate in the construction of perceptual representations of the layout or disposition of objects for cognitive purposes. Instead, it is the perception system which does this.

Like Neisser, we see the perception system (similar to his recognition system) as being intimately linked with cognitive processes such as long-term memory; but unlike Neisser, we see the perception system as providing information, not only about the identity of objects, but also about their spatial and temporal relations. In our scheme, the perception system delivers our experience of the world and the objects within it.

Clearly what distinguishes the perception system from the visuomotor modules making up the action system is the way in which the visual world is represented in the brain. Of course, the notion of representation is one of the central ideas in perception and cognition, although the type(s) of representations used in visual perception and the very notion of representation itself have been the source of much debate. Nevertheless, the goal of visual perception is often taken to be the creation of a representation that is in some sense an internal model of the three-dimensional world. In this sense, a representation is a reconstruction of the world (for further critical discussion of this approach see Ballard and Brown, 1992; Churchland et al., 1994; Tarr and Black, 1994 and accompanying commentaries). This approach to vision is exemplified by Marr (1982) who concentrated on the representation of information about objects for the purposes of recognition. According to this approach, the major task of recognition is to reconstruct a detailed and accurate model or replica of the three-dimensional world on the basis of the two-dimensional data present at the retinas.

Presumably, the proposed representation is not only important for recognition, but plays a crucial role in other cognitive activities related to spatial reasoning and the semantics of objects and scenes. It is the construction of this kind of representation that we see as the major function of the perception system – a kind of ‘general purpose’ representation that can serve as the substrate upon which a large range of cognitive operations can be mounted (in fact, the cognitive operations are themselves intimately involved in the construction of the representation upon which they operate). Of course, the nature of representations used for recognition and other cognitive acts is far from settled. A large proportion of recent research in object vision has been directed at uncovering the nature of this presentation as other papers in this volume attest. It is also clear that although Marr’s approach to object recognition has been very influential, recognition need not entail reconstruction in the way he envisaged.

Our perception of the world certainly appears remarkably rich and detailed. Nevertheless much of this perceptual representation is ‘virtual’ and is derived from memory rather than visual input, (e.g. McConkie and Currie, 1996; O’Regan, 1992; Rensink et al., 1997). Much of the metric information about objects and their relations is inaccurate and even unavailable (for review, see Intraub, 1997). And in any case, the metrical information is not computed with reference to the observer as much as it is to other objects in a visual array (Goodale and Haffenden, 1998). Indeed, if perceptual representations were to attempt to deliver the real metrics of all objects in the visual array, the computational load would be astronomical. The solution that perception appears to have adopted is to use world-based coordinates – in which the real metric of that world need not be computed. Only the relative position, orientation, size and motion of objects is of concern to perception. For example, we can watch the same scene unfold on television or on a movie screen without be confused by the enormous change in the coordinate frame.

As soon as we direct a motor act towards an object, an entirely different set of constraints applies. We can no longer rely on the perception system's 'general purpose' representation. We could not, for example, direct actions towards what we see on television, however compelling and 'real' the depicted scene might be. To be accurate, the actions must be finely tuned to the metrics of the real world. Moreover, different actions will engage different effectors. As a consequence, the computations for the visual control of actions must not only take into account the real metrics of the world, they must be specific to the particular motor output required. Directing a saccadic eye movement, for example, will demand different transformations of visual input to motor output from those required to direct a manual grasping movement. The former will involve coordinate systems centred on the retina and/or head, while the latter will involve shoulder and/or wrist centered coordinates. While it is theoretically possible that a highly sophisticated 'general-purpose' representation could accommodate such transformations, such a possibility seems unlikely and unnecessary. Indeed, as we saw earlier, the empirical evidence from a broad range of studies indicates that visuomotor control in humans and other primates is organized in much the same way as it is in simpler vertebrates, such as the frog (for review see Milner and Goodale, 1995). Moreover, these different visuomotor modules work in real time with only limited 'memory'. In other words, once a movement is made the visuomotor coordinates used to program and guide that movement are lost. Even if the movement is not performed, the coordinates cannot be stored much beyond a second or two (e.g. Gnadt et al., 1991; Goodale et al., 1994a). Perception of course has a much longer time course and stores information perhaps in some cases for a lifetime. In summary, the visuomotor modules within the action system transform sensory information directly into motor output rather than using reconstructions of visual scenes. Moreover, as will be described below, such transformations are not available to consciousness in the way that outputs of perceptual processes usually are.

7. Dissociations between action and perception in normal subjects

Although the visual fields of the two eyes together span about 200°, most of our perceptual experience is confined to the few degrees subtended by the foveal and parafoveal region. In short, we see what we are looking at. Yet as we move through the world, stepping over curbs, negotiating doorways, and grasping door handles, we often utilize visual information from the far periphery of vision. This differential use of the fovea and peripheral visual fields by perception and action systems may explain why in the monkey there is differential representation of these regions in the ventral and dorsal streams. The receptive fields of cells in the inferotemporal cortex almost always include the fovea and very little of the far peripheral visual fields whereas cells in the posterior parietal cortex have a very large representation of the peripheral visual fields (Baizer et al., 1991). Indeed, in some areas of the dorsal stream, such as the parieto-occipital area, the portion of cortex devoted to the fovea is no larger than would be expected on the basis of the extent of the visual field it subtends; i.e. there is no ‘cortical magnification’ of central vision (Gattass et al., 1985).

If a similar retinotopic organization of cortical areas exists in the human brain, then one might expect that the visual control of motor behavior might be quite sensitive to differences in visual stimuli presented in the far peripheral visual field whereas perceptual judgements of the same stimuli might be relatively insensitive. In a recent experiment, Goodale and Murphy (1997) presented subjects with five different rectangular objects of the same overall size but different dimensions. These objects were presented randomly at different retinal eccentricities that varied from 5 to 70° and subjects were required to categorize each object into one of five previously learned categories or, in another block of trials, to reach out and grasp the object across its longitudinal axis. As one might expect, the variability of the subjects’ perceptual categorizations increased substantially as the objects were presented at more and more eccentric locations. In sharp contrast, the relationship between the aperture of their grasp (before contact) and the width of the object was as well-tuned at 70° as it was at 5°. There was also another striking difference between subjects’ perceptual judgements of the width of the objects and the calibration of their grasp. Although the subjects reported that objects did not look as wide in the far periphery as the same objects in more parafoveal regions, the aperture of their grasp was actually larger for objects in the peripheral visual field (even though the grasp continued to be well-calibrated with respect to the object’s dimensions). These dissociations between verbal reports and visuomotor control again emphasize the specialization of different parts of the visual system for perception and action.

Another way to demonstrate the distinction between perception and action systems in vision is to look at the way each system deals with objects embedded in pictorial illusions. Pictorial illusions are, of course, favourite ways of illustrating interpretive and context-sensitive aspects of visual perception. Consider the Ebbinghaus (or Titchener Circles) Illusion for a moment. In this familiar illusion, two target circles of equal size, each surrounded by a circular array of either smaller or larger circles, are presented side by side (see Fig. 4a). Subjects typically report that the target circle surrounded by the array of smaller circles appears larger than the one surrounded by the array of larger circles, presumably because of the difference in the contrast in size between the target circles and the surrounding circles. It is also possible to make the two target circles appear identical in size by increasing the actual size of the target circle surrounded by the array of larger circles (see Fig. 4b).

While our perceptual judgements of what we see are clearly affected by the manipulations of the stimulus array, there is good reason to believe that the calibration of size-dependent motor outputs, such as grip aperture during grasping, would not be. After all, when we reach out to pick up an object, particularly one we have not seen before, the visuomotor networks controlling grasping must compute the size (and distance) of the object accurately if we are to pick it up efficiently. It is not

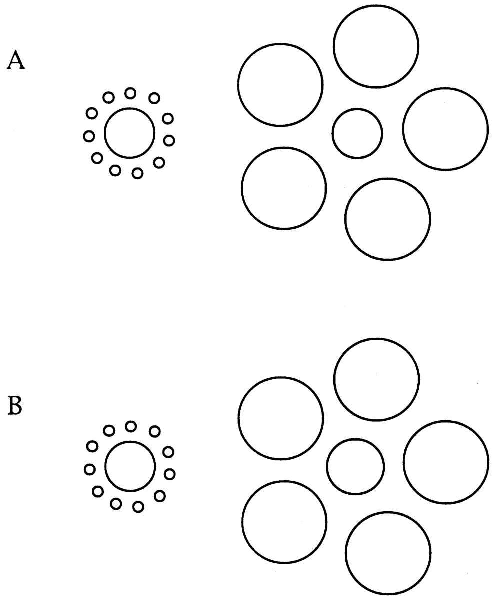

Fig. 4. The ‘Ebbinghaus’ illusion. The standard version of the illusion, the target circles in the centre of the two arrays appear to be different in size even though they are physically identical, as shown in (A). For most people, the circle in the annulus of smaller circles appears to be larger than the circle in the annulus of larger circles. (B) Shows a version of the illusion in which the target circle in the array of large circles has been made physically larger than the other target circle. The two target circles should now appear to be perceptually equivalent in size.

enough to know that the target object is larger or smaller than surrounding objects; the visuomotor module controlling hand aperture must compute its real size. For this reason, one might expect grip scaling to be refractory to size-contrast illusions.

To test this possibility, Aglioti et al. (1995) developed a three-dimensional version of the Ebbinghaus Illusion in which two thin ‘poker-chip’ discs were used as the target circles. The disks were arranged as pairs on a standard Ebbinghaus annular circle display (see Fig. 5) drawn on a white background and positioned directly in front of the subject. Trials in which the two disks appeared perceptually identical but were physically different in size were randomly alternated with trials in which the disks appeared perceptually different but were physically identical. The left-right position of the arrays of large and small circles was of course randomly varied throughout. Subjects (all of whom had normal vision) were given the following instructions: if the discs appear equal in size, pick up the one on the right; if they appear different, pick up the one on the left. Subjects used their right hand and grip aperture was tracked using standard opto-electronic recording.

Although there was considerable individual variation, all the subjects remained sensitive to the size-contrast illusion throughout testing. In other words, their choice of disk was affected by the contrast in size between the disks and the surrounding circles. As a consequence, they treated disks that were actually physically different

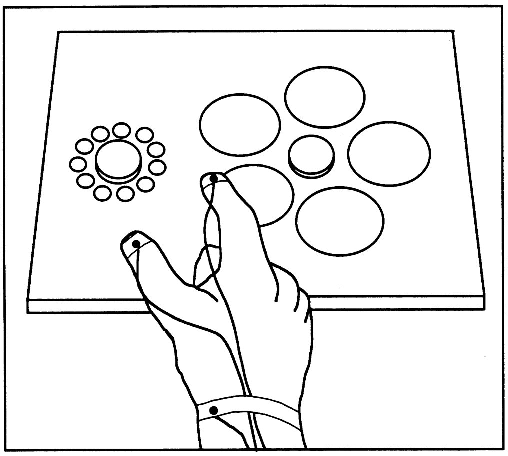

Fig. 5. A line drawing of our three-dimensional version of the Ebbinghaus illusion. Note the infra-red light emitting diodes (IREDs) attached to the finger, thumb and wrist of the subject.

in size as perceptually equivalent and they treated disks that were physically identical as perceptually different. Remarkably, however, the scaling of their grasp was affected very little by these beliefs. Instead, the maximum grip aperture, which was achieved approximately 70% of the way through the reach towards the disk, was almost entirely determined by the true size of that disk. Thus, on trials in which the two disks were perceived as being the same size, subjects opened their hand wider for the larger disk than they did for the smaller one. An example of such a case in illustrated in Fig. 6a. In fact, as shown in Fig. 6b, the difference in grip aperture for large and small disks was the same for trials in which the subject believed the two disks were equivalent in size (even though they were different) as it was for trials in which the subject believed the two disks were different in size (even though they were identical). In short, the calibration of grip size seemed to be largely impervious to the effects of the size-contrast illusion. This difference in the susceptibility of perceptual judgements and the visual control of prehension was replicated in a recent study in which subjects had no opportunity to compare their hand opening with the goal object during the execution of the movement (Haffenden and Goodale, 1998).

The dissociation between perceptual judgements and the calibration of grasping is not limited to the Ebbinghaus Illusion. The vertical-horizontal illusion is one in which a vertical line that bisects a horizontal line appears longer than the horizontal line even though both lines are in fact the same length. Vishton and Cutting (1995) have recently demonstrated that even though subjects show the usual bias in their judgements of line length, they did not show a bias when they attempted to reach out and ‘grasp’ the lines. The relative insensitivity of reaching and grasping to pictorial illusions has also been demonstrated for the Müller-lyer illusion (Gentilucci et al., 1996) and the Ponzo illusion (Ian Whishaw, personal communication).

But why should perception be so susceptible to these illusions while the calibra-

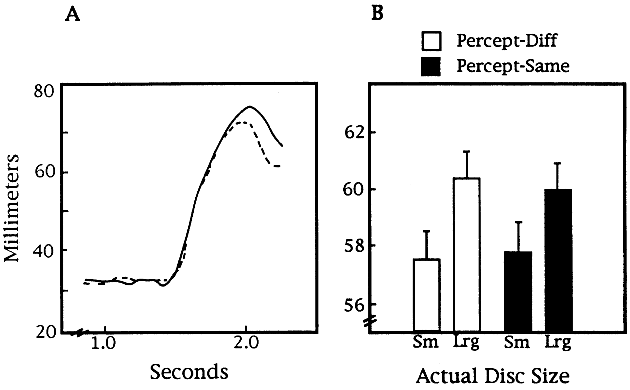

Fig. 6. Graphs illustrating grip aperture in different testing conditions. (A) Representative grip aperture profiles for one subject picking up a large disc (solid line) and a small disc (broken line) on separate trials in which he judged the two discs to be identical in size (even though, of course, the two discs were physically quite different). In both cases, the disc was located on the left hand side of the display. (B)The mean maximum grip aperture for the 14 subjects in different testing conditions. The two solid bars on the right indicate the maximum aperture on trials in which the two discs were judged to be perceptually the same even though they were physically different in size. The two open bars on the left indicate the mean maximum aperture on trials in which the two discs were judged to be perceptually different even though they were physically the same size (either two large discs or two small discs).

tion of grasp is not. Take the Ebbinghaus illusion for example. It is possible that the illusion arises from a straightforward relative-size scaling mechanism, whereby an object that is smaller than its immediate neighbors is assumed to be smaller than a similar object that is larger than its immediate neighbors (for review, see Coren and Girgus, 1978). It is also possible that a computation relating image size and distance is responsible for the illusion. If the array of smaller circles is assumed to be more distant from the observer than the array of larger circles, then the target circle within the array of smaller circles will also be perceived as more distant, and therefore larger, than the target circle of equivalent retinal image size within the array of larger circles. In other words, the illusion might simply be a consequence of the perception system’s attempt to make size constancy judgments on the basis of an analysis of the entire visual array (Gregory, 1963).

Mechanisms such as these, in which the relations between objects in the visual array play a crucial role in scene interpretation, are clearly central to the operation of the perception system. As we gaze across the landscape, some of the objects within our field of view will be perceived, in an obligatory fashion, as larger or closer than others. Perception is by its very nature relative. In contrast, the execution of a goal-directed act like manual prehension depends on metrical computations that are centered on the target itself. Moreover, the visual mechanisms within the action system that mediate the control of the grasping movements must compute the real distance of the object (presumably on the basis of reliable cues such as stereopsis and retinal motion). As a consequence, computation of the retinal image size of the object coupled with an accurate estimate of distance will deliver the true size of the object for calibrating the grip – and such computations may be quite insensitive to the kinds of pictorial cues that drive our perception of familiar illusions. Thus, the very act by means of which subjects indicate their susceptibility to the illusion (i.e. picking up one of the two target circles) is itself unaffected by the visual information driving that illusion. This paradox demonstrates that what we think we ‘see’ is not always what guides our actions. It also provides evidence for the parallel operation of the two kinds of visual processing that we described earlier, each apparently designed to serve quite different purposes, and each characterized by quite different properties.

8. The action/perception distinction in computational vision

We would suggest that the distinction between vision for perception and vision for action is relevant to some aspects of a current debate in the computational vision literature. The debate could be characterized as one between ‘behaviorist or purposive’ approaches and ‘reconstructive’ approaches to vision. Here we will make some general remarks that capture only some aspects of the various positions in the debate as there are many theoretical divergences within both ‘behaviorist’ and ‘reconstructionist’ proposals that we are overlooking. We are obviously oversimplifying in the belief that such a caricature captures some significant agreement and divergences in general orientation that can be mapped on to the distinctions in perception and action systems that we have proposed. The reader should refer to Tarr and Black (1994) and the accompanying commentaries on that paper for more detail. Other relevant papers that could be consulted are Ballard and Brown (1992), Churchland et al. (1994), Jolion (1994), Sloman (1989) and the collection of papers edited by Aloimonos (1992).

Researchers who espouse behaviorist approaches to vision are often quite sympathetic to the general framework of Gibson (1979) in his emphasis on the active, exploratory nature of perception. As a consequence they concentrate on visual behavior such as obstacle avoidance, reaching and grasping, gaze control, and other aspects of behavior guided by visual input. It is essentially a ‘motor’ view of perception (Churchland et al., 1994; Watt, 1993). We would suggest that many of the visuomotor transformations that occupy the attention of these researchers are part of the dorsal action system and its associated subcortical and cortical networks. Thus, the preoccupation with visually guided actions that characterizes behaviorist approaches to vision has meant that most of the visual mechanisms that are being studied are those found in the dorsal stream.

In contrast, the reconstructive approach (e.g. Marr, 1982) concentrates on the creation of a replica of the world ‘out there’ on the basis of the sensory input data present at the retinas. In a sense, the approach of Marr is a ‘passive’ view of perception in which the representation is central and the external behavior of the organism is largely ignored (Ballard and Brown, 1992). For Marr, vision is ‘an information processing task’ and visual science is conceived as an ‘inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions’

(p. 3). This approach to vision need not be seen as opposing the behaviorist approach. Indeed, we would suggest that reconstruction of the external world is exactly the kind of activity which we believe is carried out by the ventral stream. Of course, as noted above, there is considerable debate about the way in which visual mechanisms and stored representations interact in visual perception. Whatever the particular mechanisms might be that underlie recognition and other perceptual/cognitive operations, it is the ventral stream, we believe, that carries them out. Moreover, the 'awareness' that typically accompanies much of visual perception may also depend, in part, on the ventral stream pathways (for some speculative accounts of routes to visual awareness, see Crick and Koch, 1995; Goodale and Milner, 1992; Milner and Goodale, 1995).

Thus, it seems to us that at least some aspects of the arguments about the relative merits of 'behaviorist' and 'reconstructionist' approaches to vision are misplaced. The two approaches are concerned with different visual systems that have different agendas (see also Neisser, 1989, 1994). Of course we realize that the drawing of parallels between 'reconstructionist' and 'behaviorist' approaches and the ventral and dorsal streams of visual processing will in no way settle the many issues that concern researchers in computational vision. Nevertheless, we hope our suggestion will supply some useful distinctions for thinking about these issues.

9. Getting it together: interactions between action and perception

Throughout this paper, we have been advancing the idea that the ventral perception system and the dorsal action system are two independent and decidedly different visual systems within the primate brain. We realize that in doing this we have overstated our position to some extent. This was a deliberate attempt to counter the tendency in object vision research to focus on issues such as recognition and other cognitive operations, without taking into account the actions that are performed on objects and the particular visuomotor transformations necessary for such actions. It is obvious that systems for action and those for perception must interact and cooperate in the control of behavior. Nevertheless, as we have tried to show, the computations carried out by the two systems complement one another.

One way to think about the interaction between the two streams, an interaction that would take advantage of the differences in their computational constraints, is in terms of a 'teleassistance' model. In teleassistance, a human operator uses a symbolic code to communicate with a robot that actually performs the required motor act on the marked goal object (Pook and Ballard, 1996). In terms of this teleassistance metaphor, the perceptual-cognitive system in the ventral stream, with its rich and detailed representations of the virtual scene, would be the human operator. Processes in the ventral stream identify a particular goal and flag the relevant object in the scene, perhaps by means of an attention-like process. Once a particular goal object has been flagged, dedicated visuomotor networks in the dorsal stream (in conjunction with related circuits in premotor cortex, basal ganglia, and brainstem) can then be activated to perform the desired motor act. Mechanisms in the dorsal stream, while not delivering anything like the visual detail provided by perception, do provide accurate information about the goal object in effector-specific frames of reference – and provide this information quickly. This means that a flagged object in the scene will be processed in parallel by both ventral and dorsal stream mechanisms – each transforming the visual information in the array for different purposes.

Thus, once a goal object has been selected for goal-directed action, the two systems process incoming visual information simultaneously – even though the nature of the visual information that is transformed might be rather different. Such simultaneous activation will, of course, provide us with visual experience (via the ventral stream) during the performance of a skilled action (mediated by the dorsal stream). For example, when we reach out to pick up an interesting book, we not only form our grasp according to the dimensions and location of that book but at the same time we might also perceive that it is one we have not seen before. Moreover, certain objects such as tools demand that we grasp the object in a particular way so that we can use it properly. In such a case both streams would have to interact fairly intimately in mediating the final motor output. Certainly there is evidence that, on the neural level, the two systems are interconnected allowing for communication and cooperation between them (reviewed in Goodale and Milner, 1992; Milner and Goodale, 1995). Thus, although there is clearly a division of labour between the perception and action systems, this division reflects the complementary role the two systems play in the production of adaptive behaviour.

Acknowledgements

The preparation of this manuscript was helped in part by grants from the Medical Research Council of Canada to M.A.G. and the Natural Sciences and Engineering Research Council to G.K.H.

References

Aglioti, S., DeSouza, J.F.X., Goodale, M.A., 1995. Size-contrast illusions deceive the eye but not the hand. Current Biology 5, 679–685.

Aloimonos, Y., 1990. Purposive and qualitative active vision. In: Proceedings of the International Conference on Pattern Recognition, Atlantic City, NJ, pp. 346–360.

Aloimonos, Y., 1992. Purposive, qualitative, active vision. CVGIP:Image Understanding 5, 3–129.

Andersen, R.A., 1987. Inferior parietal lobule function in spatial perception and visuomotor integration. In: V.B. Mountcastle, F. Plum, S.R. Geiger (Eds.), Handbook of Physiology, Section 1: The Nervous System, Vol. V, Higher Functions of the Brain, Part 2. American Physiological Association, Bethesda, MD, pp. 483–518.

Baizer, J.S., Ungerleider, L.G., Desimone, R., 1991. Organization of visual input to the inferior temporal and posterior parietal cortex in macaques. Journal of Neuroscience 11, 168–190.

Bálint, R., 1909. Seelenlämung des ‘Schauens’, optische Ataxie, räumliche Störung der Aufmerksamkeit. Monatschrift für Psychiatrie und Neurologie 25, 51–81.

Ballard, D.H., Brown, C.M., 1992. Principles of animate vision. CVGIP: Image Understanding 56, 3–21.

Biederman, I., Gerhardstein, P.C., 1995. Viewpoint-dependent mechanisms in visual object recognition:

reply to Tarr and Bülthoff, 1995. . Journal of Experimental Psychology: Human Perception and Performance 21, 1506–1514.

Buchbinder, S., Dixon, B., Hyang, Y.-W., May, J.G., Glickstein, M., 1980. The effects of cortical lesions on visual guidance of the hand. Society for Neuroscience Abstracts 6, 675.

Churchland, P.S., Ramachandran, V.S., Sejnowski, T.J., 1994. A critique of pure vision. In: C. Koch, J.L. Davis (Eds.), Large-scale neuronal theories of the brain. MIT Press, Cambridge, MA, pp. 23–60.

Coren, S., Girgus, J.S., 1978. Seeing is Deceiving: The Psychology of Visual Illusions. Lawrence Erlbaum Associates, Hillsdale, NJ.

Crick, F., Jones, E., 1993. Backwardness of human neuroanatomy. Nature 361, 109–110.

Crick, F., Koch, C., 1995. Are we aware of neural activity in primary visual cortex? Nature 275, 121–123.

Edelman, S., Bulthoff, H.H., 1992. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Research 32, 2385–2400.

Ellard, C.G., Goodale, M.A., 1986. The role of the predorsal bundle in head and body movements elicited by electrical stimulation of the superior colliculus in the Mongolian gerbil. Experimental Brain Research 64, 421–433.

Ellard, C.G., Goodale, M.A., 1988. A functional analysis of the collicular output pathways: A dissociation of deficits following lesions of the dorsal tegmental decussation and the ipsilateral collicular efferent bundle in the Mongolian gerbil. Experimental Brain Research 71, 307–319.

Ettlinger, G., 1977. Parietal cortex in visual orientation. In: F.C. Rose (Ed.), Physiological Aspects of Clinical Neurology. Blackwell, Oxford, pp. 93–100.

Ewert, J.-P., 1987. Neuroethology of releasing mechanisms: prey-catching in toads. Behavioral and Brain Sciences 10, 337–405.

Gattass, R., Sousa, A.P.B., Covey, E., 1985. Cortical visual areas of the macaque: possible substrates for pattern recognition mechanisms. In: C. Chagas, R. Gattass, C.G. Gross (Eds.), Pattern Recognition Mechanisms. Vatican City: Pontifical Academy of Sciences, pp. 1–20.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., Gore, J.C., 1997. Levels of categorization in visual recognition studies using functional magnetic resonance imaging. Current Biology 7, 645–651.

Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M.C., Toni, I., 1996. Visual illusion and action. Neuropsychologia 34, 369–376.

Georgeson, M., 1997. Guest editorial: vision and action: you ain't see nothin' yet..… Perception 26, 1–6.

Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston, MA, 332 pp.

Gnadt, J.W., Bracewell, R.M., Andersen, R.A., 1991. Sensorimotor transformation during eye movements to remembered visual targets. Vision Research 31, 693–715.

Goodale, M.A., 1983a. Vision as a sensorimotor system. In: T.E. Robinson (Ed.), Behavioral Approaches to Brain Research. Oxford University Press, New York, pp. 41–61.

Goodale, M.A., 1983b. Neural mechanisms of visual orientation in rodents: targets versus places. In: A. Hein, M. Jeannerod (Eds.), Spatially Oriented Behavior. Springer-Verlag, Berlin, pp. 35–61.

Goodale, M.A., 1988. Modularity in visuomotor control: from input to output. In: Z. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective. Ablex, Norwood, NJ, pp. 262–285.

Goodale, M.A., 1993. Visual pathways supporting perception and action in the primate cerebral cortex. Current Opinion in Neurobiology 3, 578–585.

Goodale, M.A., 1996. Visuomotor modules in the vertebrate brain. Canadian Journal of Physiology and Pharmacology 74, 390–400.

Goodale, M.A., 1997. Visual routes to perception and action in the cerebral cortex. In: M. Jeannerod (Ed.), Handbook of Neuropsychology, Vol. 11, Elsevier, Amsterdam, pp. 91–109.

Goodale, M.A., Carey, D.P., 1990. The role of cerebral cortex in visuomotor control. In: B. Kolb, R.C. Tees (Eds.), The Cerebral Cortex of the Rat, Ablex, Norwood, NJ, pp. 309–340.

Goodale, M.A., Haffenden, A., (1998). Frames of reference for perception and action in the human visual system. Neuroscience and Biobehavioral Reviews, in press.

Goodale, M.A., Jakobson, L.S., Keillor, J.M., 1994a. Differences in the visual control of pantomined and natural grasping movements. Neuropsychologia, 31, 1994, 1159–1178.

Goodale, M.A., Jakobson, L.S., Servos, P., 1996. The visual pathways mediating perception and prehension. In: R. Flanagan, P. Haggard, A. Wing (Eds.), Sensorimotor Control of the Hand. Academic Press, New York, pp. 15–31.

Goodale, M.A., Meenan, J.P., Bülthoff, H.H., Nicolle, D.A., Murphy, K.S., Racicot, C.I., 1994. Separate neural pathways for the visual analysis of object shape in perception and prehension. Current Biology 4, 604–610.

Goodale, M.A., Milner, A.D., 1982. Fractionating orienting behavior in rodents. In: D.J. Ingle, M.A. Goodale, R.J.W. Mansfield (Eds.), Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp. 267–299.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends in Neurosciences 15, 20–25.

Goodale, M.A., Milner, A.D., Jakobson, L.S., Carey, D.P., 1991. A neurological dissociation between perceiving objects and grasping them. Nature 349, 154–156.

Goodale, M.A., Murphy, K., 1997. Action and perception in the visual periphery. In: P. Their, H.-O. Karnath (Eds.), Parietal Lobe Contributions to Orientation in 3D Space. Springer-Verlag, Heidelberg, pp. 447–461.

Goodale, M.A., Murphy, K., Meenan, J.-P., Racicot, C., Nicolle, D.A., 1993. Spared object perception but poor object- calibrated grasping in a patient with optic ataxia. Society for Neuroscience Abstracts 19, 775.

Gregory, R.L., 1963. Distortions of visual space as inappropriate constancy scaling. Nature 199, 678–680.

Haaxma, R., Kuypers, H.G.J.M., 1975. Intrahemispheric cortical connexions and visual. guidance of hand and finger movements in the rhesus monkey. Brain 98, 239–260.

Haffendale, A., Goodale, M.A., 1998. The effect of pictorial illusion on prehension and perception. Journal of Cognitive Neuroscience 10, 122–136.

Humphrey, G.K., Goodale, M.A., Gurnsey, R., 1991. Orientation discrimination in a visual form; agnosic: evidence from the McCollough effect. Psychological Science 2, 331–335.

Humphrey, G.K., Goodale, M.A., Jakobson, L.S., Servos, P., 1994. The role of surface information object recognition: studies of a visual form agnosic and normal subjects. Perception 23, 1457–1481.

Humphrey, G.K., Khan, S.C., 1992. Recognizing novel views of three-dimensional objects. Canadian Journal of Psychology 46, 170–190.

Humphrey, G.K., Symons, L.A., Herbert, A.M., Goodale, M.A., 1996. A neurological dissociation between shape from shading and shape from edges. Behavioural Brain Research 76, 117–125.

Hyvärinen, J., Poranen, A., 1974. Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. Brain 97, 673–692.

Ingle, D.J., 1973. Two visual systems in the frog. Science 181, 1053–1055.

Ingle, D.J., 1980. Some effects of pretectum lesions in the frog's detection of stationary objects. Behavioural Brain Research 1, 139–163.

Ingle, D.J., 1982. Organization of visuomotor behaviors in vertebrates. In: D.J. Ingle, M.A. Goodale, R.J.W. Mansfield (Eds.), Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp. 67–109.

Ingle, D.J., 1991. Functions of subcortical visual systems in vertebrates and the evolution of higher visual mechanisms. In: R.L. Gregory, J. Cronly-Dillon (Eds.), Vision and Visual Dysfunction, Vol. 2, Evolution of the Eye and Visual System London. Macmillan, Basingstoke, pp. 152–164.

Intraub, H., 1997. The representation of visual scenes. Trends in Cognitive Sciences 1, 217–222.

Jackson, J.H., 1875. Clinical and Physiological Researches on the Nervous System. Churchill, London.

Jakobson, L.S., Archibald, Y.M., Carey, D.P., Goodale, M.A., 1991. A kinematic analysis of reaching an grasping movements in a patient recovering from optic ataxia. Neuropsychologia 29, 803–809.

Jeannerod, M., 1988. The Neural and Behavioural Organization of Goal-directed Movements. Oxford University Press, Oxford.

Jolicoeur, P., Humphrey, G.K., 1998. Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In: V. Walsh, J. Kulikowski (Eds.), Visual Constancies: Why Things Look As They Do. Cambridge University Press, Cambridge, pp. 69–123.

Jolion, J.-M.., 1994. Computer vision methodologies. CVGIP: Image Understanding 59, 53–71.

Kanwisher, N., Chun, M.M., McDermott, J., Ledden, P.J., 1996. Functional imaging of human on visual recognition. Cognitive Brain Research 5, 55–67.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience 17, 4302–4311.

Kawashima, R., Naitoh, E., Matsumura, M., Itoh, H., Ono, S., Satoh, K., Gotoh, R., Koyama, M., Inoue, K., Yoshioka, S., Fukuda, H., 1996. Topographic Representation in Human Intraparietal Sulcus of Reaching and Saccade. NeuroReport 7, 1253–1256.

Klüver, H., Bucy, P.C., 1939. Preliminary analysis of functions of the temporal lobes of monkeys. Archives of Neurological Psychiatry 42, 979–1000.

Logothetis, N.K., Pauls, J., Poggio, T., 1995. Shape representation in the inferior temporal cortex of monkeys. Current Biology 5, 552–563.

Logothetis, N.K., Sheinberg, D.L.., 1996. Visual object recognition. Annual Review of Neuroscience 19, 577–621.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Leddedn, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B.H., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences USA 92, 8135–8139.

Matsumura, M., Kawashima, R., Naito, E., Satoh, K., Takahashi, T., Yanagisawa, T., Fukuda, H., 1996. Changes in rCBF during grasping in humans examined by PET. NeuroReport 7, 749–752.

Marr, D., 1982. Vision. Freeman, San Francisco, CA.

McConkie, G.W., Currie, C.B., 1996. Visual stability across saccades while viewing complex pictures. Journal of Experimental Psychology Human Perception and Performance 22, 563–581.

Milner, A.D., Goodale., M.A., 1993. Visual pathways to perception and action. In: T.P. Hicks, S. Molotchnikoff, T. Ono (Eds.), The Visually Responsive Neuron: From Basic Neurophysiology to Behavior, Progress in Brain Research, Vol. 95. Elsevier, Amsterdam, pp. 317–338.

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

Milner, A.D., Perrett, D.I., Johnston, R.S., Benson, P.J., Jordan, T.R., Heeley, D.W., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E., Davidson, D.L.W., 1991. Perception and action in visual form agnosia. Brain 114, 405–428.

Mountcastle, V.B., Lynch, J.C., Georgopoulos, A., Sakata, H., Acuna, C., 1975. Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. Journal of Neurophysiology 38, 871–908.

Neisser, U., 1989. Direct perception and recognition as distinct perceptual systems. Address presented to the Eleventh Annual meeting of the Cognitive Science Society, Ann Arbor, MI.

Niesser, U., 1994. Multiple systems: a new approach to cognitive theory. European Journal of Cognitive Psychology 6, 225–241.

O'Regan, J.K., 1992. Solving the 'real' mysteries of visual perception: The world as an outside memory. Canadian Journal of Psychology 46, 461–488.

Perenin, M.-T., Vighetto, A., 1988. Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. Brain 111, 643–674.

Perrett, D., Benson, P.J., Hietanen, J.K., Oram, M.W., Dittrich, W.H., 1995. When is a face not a face? In: R. Gregory, J. Harris, P. Heard, D. Rose (Eds.), The Artful Eye. Oxford University Press, Oxford, pp. 95–124.

Pook, P.K., Ballard, D.H.., 1996. Deictic human/robot interaction. Robotics and Autonomous Systems 18, 259–269.

Price, C.J., Moore, C.J., Humphreys, G.W., Frackowiak, R.S.J., Friston, K.J., 1996. The neural regions subserving object recognition and naming. Proceedings of the Royal Society of London (B) 263, 1501–1507.

Pribram, K.H., 1967. Memory and the organization of attention. In: D.B. Lindsley, A.A. Lumsdaine (Eds.), Brain Function, Vol. IV, UCLA Forum in Medical Sciences 6, University of California Press, Berkley, CA, pp. 79–112.

Puce, A., Allison, T., Asagari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human faces, letterstrings, and textures: a functional magnetic resonance imaging study. Journal of Neuroscience 16, 5205–5215.

Rensink, R.A., O'Regan, J.K., Clark, J.J., 1997. To see or not to see: the need for attention to perceive changes in scenes. Psychological Science 8, 368–373.

Rock, I., DiVita, J., 1987. A case of viewer-centered object perception. Cognitive Psychology 19, 280–293.

Sakata, H., Taira, M., Mine, S., Murata, A., 1992. Hand-movement-related neurons of the posterior parietal cortex of the monkey: their role in visual guidance of hand movements. In: R. Caminiti, P.B. Johnson, Y. Burnod (Eds.), Control of Arm Movement in Space: Neurophysiological and Computational Approaches. Springer-Verlag, Berlin, pp. 185–198.

Sakata, H., Taira, M., 1994. Parietal control of hand action. Current Opinion in Neurobiology 4, 847–856.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y., 1997. The TINS Lecture: The Parietal Association Cortex in Depth Perception and Visual Control of Hand Action. Trends in Neurosciences 20, 350–357.

Schneider, G.E., 1969. Two visual systems: brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. Science 163, 895–902.

Servos, P., Goodale, M.A., Humphrey, G.K., 1993. The drawing of objects by a visual form agnosic: contribution of surface properties and memorial representations. Neuropsychologia 31, 251–259.

Sloman, A., 1989. On designing a visual system: towards a Gibsonian computational model of vision. Journal of Experimental and Theoretical Artificial Intelligence 1, 289–337.

Snyder, L.H., Batista, A.P., Andersen, R.A., 1997. Coding of intention in the posterior parietal cortex. Nature 386, 167–170.

Taira, M., Mine, S., Georgopoulos, A.P., Murata, A., Sakata, H., 1990. Parietal cortex neurons of the monkey related to the visual guidance of hand movement. Experimental Brain Research 83, 29–36.

Tarr, M.J., 1995. Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bulletin and Review 2, 55–82.

Tarr, M., Black, M., 1994. A computational and evolutionary perspective on the role of representation in vision. CVGIP: Image Understanding 60, 65–73.

Tarr, M.J., Bülthoff, H.H., 1995. Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein, 1993. Journal of Experimental Psychology: Human Perception and Performance 21, 1494–1505.

Tootell, R.B.H., Dale, A.M., Sereno, M.I., Malach, R., 1996. New images from human visual cortex. Trends in Neurosciences 19, 481–489.

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. In: D.J. Ingle, M.A. Goodale, R.J.W. Mansfield (Eds.). Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp. 549–586.

Vishton, P.M., Cutting, J.E., 1995. Veridical size perception for action: reaching vs. estimation. Investigative Ophthalmology and Visual Science 36 (Suppl.), 358.

Watt, R.J., 1991. Understanding Vision. Academic Press, London, pp. 301.

Watt, R.J., 1992. Visual analysis and the representation of spatial relations. In: G.W. Humphreys (Ed.), Understanding Vision. Basil Blackwell, Oxford, 19–38 pp.

Watt, R.J., 1993. Issues in shape perception. Image and Vision Computing 11, 389–394.

Weimer, W.B., 1977. A conceptual framework for cognitive psychology: motor theories of mind. In: R. Shaw, J. Bransford (Eds), Perceiving Acting and Knowing: Toward an Ecological Psychology. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 267–311.