

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

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

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. 'Purposeful vision', as this approach is sometimes described, has emphasized the role of vision in the direct control of actions rather than its con-

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 linwaysl-346lled,\6((Figusual)

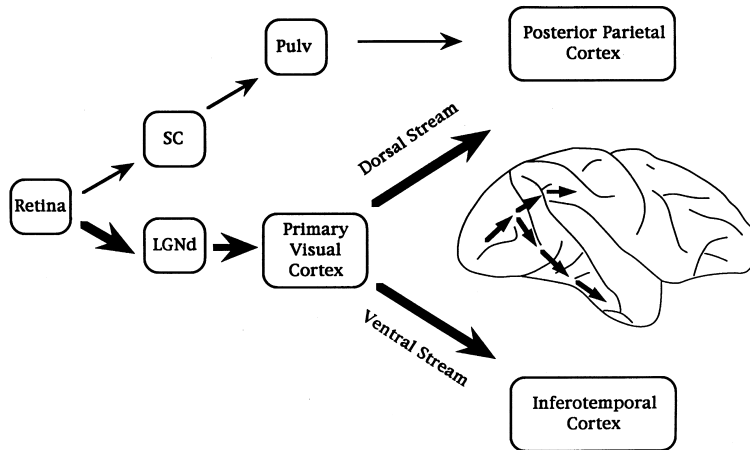


Fig. 1. Major routes whereby retinal input reaches the dorsal and ventral streams. The diagram of the macaque brain (right hemisphere) on the right of the figure shows the approximate routes of the cortico-cortical projections from the primary visual cortex to the posterior parietal and the inferotemporal 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

oughn50722(their)-52supioceisual5-41eaeimorat of objeirmain

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-

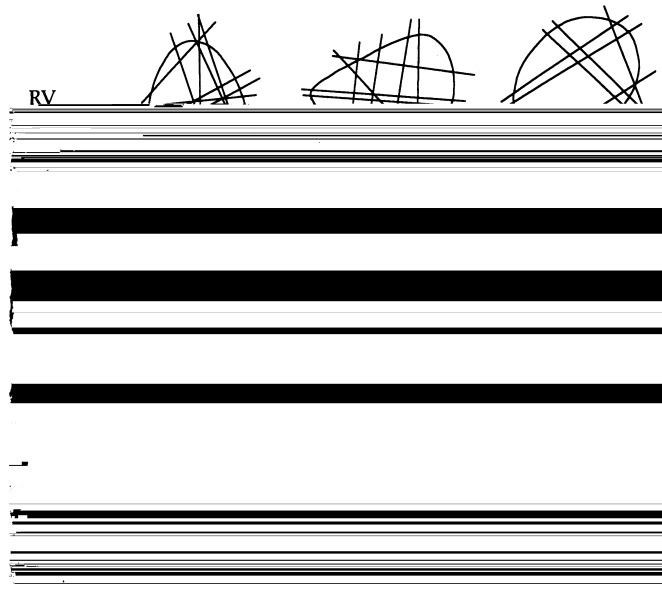


Fig. 2. The ‘grasp lines’ (joining points where the index finger and the thumb first made contact with the shape) selected by the optic ataxic patient (RV), the visual form agnostic patient (DF), and the control subject when picking up three of the twelve shapes. The four different orientations in which 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.

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 be 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

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

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 Bühlhoff, 1992; Humphrey and Khan, 1992; Rock and DiVita, 1987; Tarr, 1995; see also Biederman and Gerhardstein, 1995; Tarr and Bühlhoff, 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°

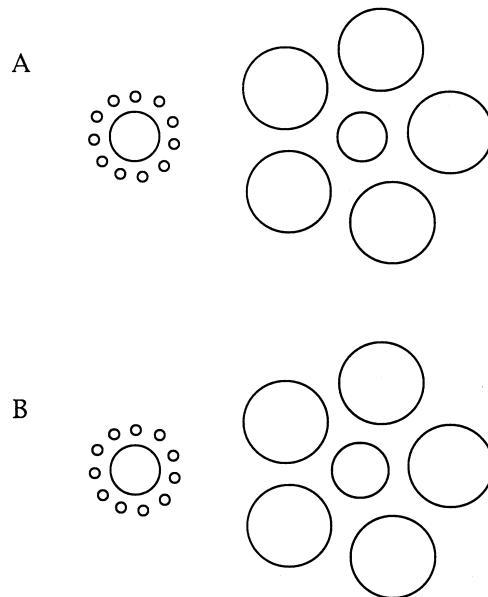


Fig. 4. The 'Ebbinghaus' illusion. The standard version of the illusion, the target circles in the centre of the two arrays appear to be different in size even though they are physically identical, as shown in (A). For most people, the circle in the annulus of smaller circles appears to be larger than 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 discs 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 discs appeared perceptually identical but were physically different in size were randomly alternated with trials in which the discs 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

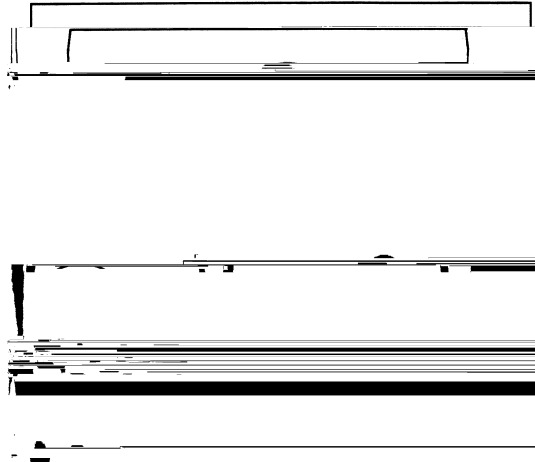


Fig. 5. A line drawing of our three-dimensional version of the Ebbinghaus illusion. Note the infra-red light emitting diodes (IREDS) attached to the finger, thumb and wrist of the subject.

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 is 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-

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

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

(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.

- 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.

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.