Robert Eamon Briscoe

Another Look at the Two Visual Systems Hypothesis

The Argument from Illusion Studies

1. Introduction

The purpose of this paper is to defend what I call the *action-oriented coding theory* (**ACT**) of spatially contentful visual experience. Integral to ACT is the view that conscious visual experience and visually guided action make use of a common subject-relative or 'egocentric' frame of reference. Proponents of the influential *two visual systems hypothesis* (**TVSH**), however, have maintained on empirical grounds that this view is false (Milner & Goodale, 1995/2006; Clark, 1999; 2001; Campbell, 2002; Jacob & Jeannerod, 2003; Goodale & Milner, 2004). One main source of evidence for TVSH comes from behavioral studies of the comparative effects of size-contrast illusions on visual awareness and visuomotor action. This paper shows that not only is the evidence from illusion studies inconclusive, there is a better, ACT-friendly interpretation of the evidence that avoids serious theoretical difficulties faced by TVSH.

2. The Action-Oriented Coding Theory (ACT)

Many philosophers have supposed that conscious visual experience and visually guided action make use of a common egocentric frame of reference (Taylor, 1979/1985; Evans, 1982; Campbell, 1994; Bermúdez, 1998; 2007;Grush, 1998; 2000; forthcoming; Gallagher, 2005). Christopher Peacocke's proposal that the representational content of a visual experience is given by a spatial type that he calls a *scenario* is one familiar elaboration of this view (1992, chap. 3). Individuating a

Correspondence:
R.E. Briscoe, Dept of Philosophy, 202 Ellis Hall, Ohio University, Athens, OH 45701, USA. *Email: rbriscoe@gmail.com*

Journal of Consciousness Studies, **15**, No. 8, 2008, pp. 35–62

scenario involves specifying which scenes — which ways of 'filling out' space around the perceiver at the time of the experience — are consistent with the content's correctness. Each such scene is constituted by an assignment of surfaces and properties (orientations, textures, colours, etc.) to points in a spatial coordinate system whose axes originate from the center of the perceiver's chest. According to Peacocke, scenario contents are involved both in locating the position of an object relative to the perceiver in visual experience and in generating intentional movements and actions targeted on or otherwise directed in relation to the object. I see the saltcellar as located *there* — where *there* is shorthand for a location more precisely specifiable using the egocentric axes *right*/*left*, *front*/*behind*, and *above*/*below* — and, so, it is *there* that I intend to reach when it is my purpose to grasp the saltcellar with my hand. The testimony of the senses is delivered in an egocentric language that the body understands.

Making a couple of points at the outset will hopefully help avert misunderstanding. First, talk of an 'egocentric' frame of reference need not be taken to imply that visual experience organizes the spatial layout of visible objects and surfaces around a *single bodily origin*, e.g., a point in the perceiver's torso (as in the framework Peacocke develops), or the perceiver's center of gravity, or the apex of the solid angle of the perceiver's visual field. When I see an object's egocentric location, I do not simply see its location relative to myself. Indeed, there is no privileged point in (or on) my body that counts as *me* for purposes of characterizing my perceived spatial relation to the object. Rather, my visual experience of the object may convey information about the object's location relative to *any* part of my body of which I am proprioceptively aware.[1] I may perceive, e.g., that a book on the shelf is closer to my right hand than to my left hand, above my waist, but below my chin, etc. Such perspectival, body-relative spatial information – which may be more or less *precise*, depending inter alia on the relevant effector (eye, head, hand, etc.), the object's distance in depth (Cutting & Vishton, 1995), and the visual structure of the object's background (Dassonville & Bala, 2004), and which may be

---

[1] As Anthony Marcel writes, 'the "ego" of an "egocentric" reference frame is whatever is the body part, and its indexical location, that stands in an intentional relation to a target, and this can be multiple points simultaneously' (2003, p. 84). I should note that our abilities to plan motor actions in respect of perceived objects does not require that spatial information in an eye-centered frame of reference be converted by so-called 'coordinate transformation' into *all* other effector-specific frames of reference (either serially or in parallel) prior to acting. Rather it is plausible that eye-centered spatial information about attended targets may be converted only when and as needed for current motor purposes. For a defence of such a 'conversion-on-demand' model, see Henriques *et al.* (2002).

more or less *salient*, depending inter alia on the specific task situation, the perceiver's expertise, and correlative demands on her attention — is plausibly part of the content of a visual experience of an object and is reflected in its phenomenology.[2]

Second, the notion that conscious visual experience conveys information to the perceiver about egocentric spatial properties is compatible with the claim that visual experience *also* conveys information to the perceiver about object- or scene-relative spatial properties.[3] When I view two candlesticks on the dining table, e.g., I see not only their direction and relative distance in depth from me, but also their 'allocentric' spatial relations to one another and to other visible objects and surfaces in the room. Notably, certain objects have intrinsic axes of their own in relation to which I am able to perceive the locations of things around them. When I view a cat, I see not only its location relative to my bodily axes, but also the locations of other objects to *its* intrinsic axes. Whether a toy mouse is perceived as to the right, or in front of, or behind the cat depends on the cat's orientation relative to the toy, not my own.

That visuomotor action targeted on an object depends on the perceiver's ability to locate the object in an egocentric frame of reference seems uncontroversial. As Anthony Marcel writes,

> In so far as motor specifications are in spatial terms, they must entail coordinates, even if only implicitly. It is logical necessary that such specifications are in egocentric coordinates, if we allow points on the body to be considered as points of origin for a spatial frame of reference. The starting-points of the body parts that are to implement the action must be specified in a common reference frame with the targets, so that the trajectory or movement accelerations and decelerations are produced. The targets of such movements must also be specified with respect to the starting-points of the relevant body parts or effectors… The only spatial description *common for all body parts and for external locations is an egocentric one* (2003, p. 84).

Further, the ability to locate distal objects (and surfaces) using an egocentric frame of reference seems integral to one's awareness of space as an arena for movement and world-involving engagement and, so, to one's sense of embodied agency. Indeed, to perceive an object's location in egocentric space is not passively to locate the object in a 3D, Cartesian coordinate system originating from one's own body, but

---

[2] I thus agree with Shaun Gallagher that 'in normal experience there is no phenomenal division between motor space, proprioceptive space, and perceptual space. Rather, conscious experience is normally of an intermodally seamless spatial system' (2005, p. 59).

[3] Thanks to John Schwenkler for a lively exchange concerning this point.

rather to locate the object in a motoric space of possible bodily actions (orienting, tracking, pointing, grasping, etc.). It should be clear why egocentric spatial representation is intrinsically action-oriented or agent-involving on this view: one's abilities to perceive objects in surrounding space are not constitutively separable from one's abilities to generate intentional movements and actions targeted on or otherwise directed in relation to those objects (Campbell, 1994).[4] Accordingly, in what follows, I shall refer to the view that conscious visual experience and visually guided action make use of a common egocentric frame of reference as the *action-oriented coding theory* (ACT). ACT, I should emphasize, is not committed to the claim that conscious visual experience is *necessary* for intelligent, environmentally responsive visuomotor action. In fact, ACT does not make any pronouncements concerning the extent to which visually guided action is possible *without* conscious visual experience.[5] ACT only claims that the same kind of egocentric spatial content is conveyed to the subject by conscious visual experience as is implicated in the subject's intentions for object-directed movement and action.

3. The Two Visual Systems Hypothesis

The assumption, integral to ACT, that both visual awareness and visuomotor action make use of egocentric frame of reference has a

---

[4] I should emphasize that this is a very different view of the role of bodily action in perception than that recently defended by Alva Noë and Susan Hurley. In order to hedge their 'enactive' approach against empirical objections premised on the two visual systems hypothesis Hurley & Noë (2007) deny that visual spatial awareness requires any understanding of how actually to interact with the objects we see. Visual awareness of an object's spatial properties, they argue, only requires understanding of how the object's 'look' or 'appearance' would vary as a function of possible bodily movements. For critical assessment of the enactive approach, see Briscoe (2008).

[5] Hence ACT is not open to objections faced by what John Campbell (2002) calls the 'Grounding Thesis'. According to the Grounding Thesis, the spatial parameters for one's visually based action on an object are fully determined by and, in this sense, 'grounded' in aspects of one's conscious visual experience of the object. When one reaches for an object, Campbell writes, 'the visual information that is being used in setting the parameters for action *must* be part of the content of [one's] experience of the object' (2002, p. 50, my emphasis). This view does not have much *prima facie* plausibility. One reason is that there are many examples of visually transduced information subserving finely tuned action in the absence of conscious seeing. In navigating a busy city sidewalk while conversing with a friend, or returning a fast tennis serve, or driving a car while deeply absorbed in thought, one's bodily responses and adjustments often seem to be prompted and guided by the nonconscious use of visual information. Another more obvious reason is phylogenetic. As Gareth Evans writes 'it seems abundantly clear that the evolution could throw up an organism in which… advantageous links [between sensory input and behavioral output] were established long before it had provided us with a conscious subject of experience' (1985, p. 387). But, if this is the case, then conscious visual experience clearly cannot be necessary for *all* environmentally responsive visuomotor action.

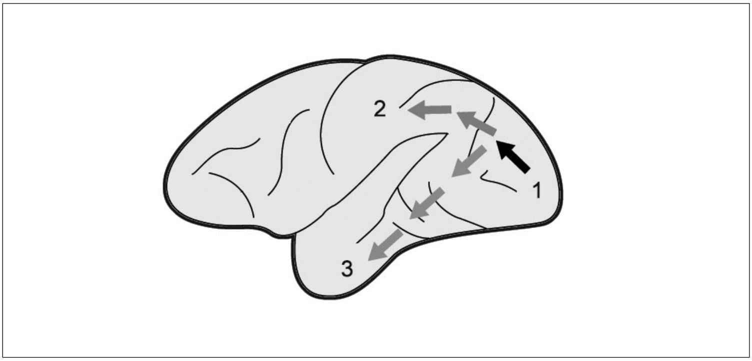

Figure. 1. A sideways view of the macaque monkey brain.

Dorsal processing stream from primary visual cortex (1) to posterior parietal cortex (2). Ventral processing stream from primary visual cortex (1) to inferotemporal cortex (3).

great deal of intuitive plausibility. Proponents of the influential *two visual systems hypothesis* (TVSH) have nonetheless maintained on empirical grounds that the assumption is false (Milner & Goodale, 1995/2006; Clark, 1999; 2001; Campbell, 2002; Jacob & Jeannerod, 2003; Goodale & Milner, 2004). According to TVSH, the primate brain comprises two, functionally dissociable visual systems: a phylogenetically ancient system subserving visually based action and a phylogenetically recent system subserving conscious visual awareness (figure 1). The former system is identified with the putative *dorsal* processing stream from primary visual cortex (V1) to posterior parietal cortex, while the latter system is identified with the putative *ventral* processing stream from primary visual cortex to inferotemporal cortex. The hypothesized 'action' and 'perception' systems can be succinctly distinguished as follows:

> *Action*: The action system is concerned with the nonconscious control and guidance of visually based actions. It contains an array of dedicated visuomotor modules that transform visual inputs into spatially directed motor outputs. Dorsal processing supporting the action system codes fine-grained metrical information about the absolute size, distance, and geometry of objects in an *egocentric* frame of reference. Upstream, i.e., 'bottom-up,' sources of 3-D spatial information to dorsal processing are quite limited. These include stereopsis (binocular disparity), vergence, retinal image size, and motion parallax.[6]

---

[6] As we shall see (§§4–5), however, additional sources of 3-D spatial information are available to the dorsal stream through cross connections with the ventral stream.

> *Perception*: The perception system subserves conscious, high-level recognition of objects and their task-relative significance or function. It is also implicated in the selection of targets for the visuomotor system, e.g., a hatchet, and in the selection of object-appropriate types of action in which to engage, e.g., taking hold of the hatchet by its handle. Crucially, ventral stream processing supporting the perception system codes only coarse-grained metrical information about the relative size, distance, and geometry of objects in an *allocentric* or *scene-based* frame of reference. Upstream sources of 3-D spatial information to ventral processing are quite extensive. In addition to binocular cues such as stereopsis and vergence, these include monocular, 'pictorial' cues such as occlusion, relative size, shading, and reflections, as well as gestalt principles of perceptual organization. Downstream, i.e., 'top-down,' sources of 3-D spatial information include stored knowledge about specific types of objects and scenes.

In representing spatial properties, the two hypothesized visual systems are thus taken to contrast significantly in respect of the metrics, the frames of reference, and the sources of spatial information that they respectively exploit.

Before proceeding, it is necessary to point out that matters are significantly complicated by two observations. First, there is growing evidence that certain dorsal processing areas subserve both action *and* visual awareness of objects in space (Rizzolatti & Matelli, 2003; Gallese, 2007). There is a strong correlation, as pointed out in §6 below, between damage to the superior part of the posterior parietal lobe and visuospatial attentional deficits. Subjects with such damage, e.g., may exhibit visual 'extinction' in which stimuli on the contralesional side of visual space are sometimes completely ignored, i.e., absent from visual experience. This suggests that visuospatial awareness cannot be neatly mapped onto ventral processing areas.

Second, not only are there multiple anatomical connections between the two putative processing streams, there are strong reasons to think that substantial interaction between them is *functionally necessary* for a wide variety of familiar actions. The movements one makes in picking up a cup of coffee, e.g., are determined not only by the cup's visible spatial properties — its shape, location, etc. — but also by its weight, how full the cup is, and by the temperature of the coffee (for related examples, see Jeannerod, 1997; Jacob & Jeannerod, 2003; and Glover, 2004). Plausibly, the mechanics of spatially directed actions involving high-level, stored object knowledge would be determined by *both* dorsal and ventral processing areas in normal subjects.[7] In tandem, these two observations suggest that the story of the relationship between the two streams in everyday action may be that of mythical Alpheus and Arethusa writ large.

Milner and Goodale explain the relationship between the two hypothesized systems by analogy with the relationship between a human operator and a semiautonomous robot guided by *tele-assistance* (1995/2006, pp. 231–4; 2004, 98–101). In tele-assistance, a remote human operator identifies a goal object, flags the target for the robot, and specifies an action on the target for the robot to perform. Once the relevant information has been communicated to the robot, the robot uses its own sensing devices and processors to determine which movements would enable it to achieve the remotely specified goal. John Campbell uses a similar analogy in order to explain the relationship between conscious seeing and visually based action:

> There is an obvious analogy with the behaviour of a heat-seeking missile. Once the thing is launched, it sets the parameters for action on its target in its own way; but to have it reach the target you want, you have to have it pointed in the right direction before it begins, so that it has actually locked on to the intended target (2002, p. 56).

Notably, both analogies assume that the target-selecting system (the ventral stream) has no difficulty in communicating to the target-engaging system (the dorsal stream) with which object it is to interact. This assumption, however, is quite substantial in view of the consideration that, according to TVSH, the two systems are locating objects using fundamentally different spatial frames of reference. (I shall return to this point in §5 below.)

4. The Argument from Illusion Studies

One of the main sources of evidence for TVSH comes from behavioral studies of the comparative effects of size-contrast illusions on visual

---

[7] Pisella *et al.* (2006) point out that damage to ventral processing areas may in fact have far more profound consequences for everyday action than damage to dorsal processing areas. Subjects with optic ataxia consequent upon damage to dorsal stream areas have difficulty in reaching toward objects in peripheral vision, but are easily able to compensate by foveating visual targets and slowing target-directed movements. Indeed, for this reason many optic ataxics do not fully notice their visuomotor deficits (Rossetti *et al.*, 2003; Rossetti *et al.*, 2005b). By contrast, damage to ventral processing areas in inferotemporal cortex may hinder or totally obstruct the performance of many familiar actions that depend on stored, object-specific knowledge, e.g., interacting with an object in a manner appropriate to its identity, purpose, or functional properties.

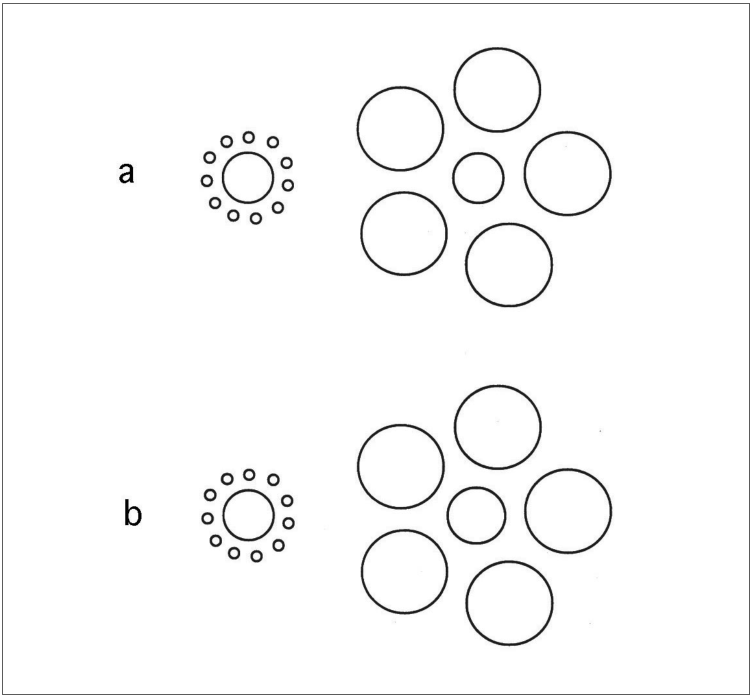

Figure. 2. Titchener Circles illusion

(a) The circles in the center of the two arrays perceptually appear to be different in size although they are physically identical.

(b) The circles in the center of the two arrays perceptually appear to be identical in size although they are physically different.

perception and visuomotor action.[8] Milner & Goodale, 1995/2006, chap. 6 and Goodale & Milner, 2004, chap. 6 in particular appeal to an experiment conducted by Aglioti *et al.* 1995 involving the Titchener (Ebbinghaus) Circles illusion.

In **figure 2a**, the two central circles perceptually appear to be different in size although they are physically identical, while in **figure 2b** the two central circles perceptually appear to be identical in size although they are physically different. (In **2b**, the central circle on the

---

[8] Additional sources of evidence for TVSH include empirical demonstrations that brain damage in profound visual form agnosia and optic ataxia can have different and separate effects on visual awareness and visuomotor action as well theoretical speculation about the computational demands respectively made by conscious seeing and visually guided engagement with objects. Space does not permit adequate discussion here, but for a review see Milner & Goodale (1995/2006) and Goodale & Milner (2004).

right has been enlarged in order to appear as the same size as the central circle on the left.) The illusion presumably arises due to the way the visual system is influenced by the contextual, pictorial depth cues in the display, in particular, the relative size of the circles. (Relative-size is a 'contextual' or 'context-dependent' pictorial depth cue because it involves comparisons between different objects in the scene. In addition to relative size, other contextual, pictorial depth cues include relative density, occlusion, height in the visual field, and aerial perspective. See Cutting & Vishton, 1995 and Palmer, 1999, chap. 5 for discussion.)

In **2a**, the circles in the annulus (ring) around the central circle on the left are much smaller than the circles in the annulus around the central circle on the right and, so, are perspectivally interpreted — together with the central circle on the left — as more distant in depth. But because the two central circles in **2a** are physically the same size (and, so, subtend the same visual angle), the central circle on the left is interpreted as being larger than the central circle on the right.

In the experiment, Aglioti and his colleagues constructed a 3-D version of the illusion, using thin solid disks (figure 3). Subjects were asked to pick up the central disk on the left if the two central disks appeared identical in size and to pick up the central disk on the right if they appeared different in size. The experimenters varied the relative size of the two target disks randomly so that in some trials physically different disks appeared perceptually identical in size, while in other trials physically identical disks appeared perceptually different in size. In selecting a disk in either trial condition, Milner and Goodale observe, 'subjects indicated their susceptibility to the visual illusion' (1995/2006, p. 169). Nonetheless, the effect of the illusion was found to be significantly more pronounced with respect to perception (as measured by the distance between thumb and forefinger in manual estimate of disk size) than with respect to action (as measured by grip aperture in prehension). Similar findings have been reported for a variety of other visual illusions including the Müller-Lyer illusion (Daprati & Gentilucci, 1997), the Ponzo illusion (Jackson & Shaw, 2000; Gonzalez *et al.*, 2007b), the Dot-in-Frame illusion (Bridgeman *et al.* 1997), and, recently, the Hollow-Face illusion (Króliczak *et al.*, 2006).

Milner and Goodale argue that the experimental findings provide support for the view that conscious seeing utilizes an *object-relative* metric in an allocentric or scene-based frame of reference, while visuomotor systems utilize an *absolute* metric in an egocentric frame

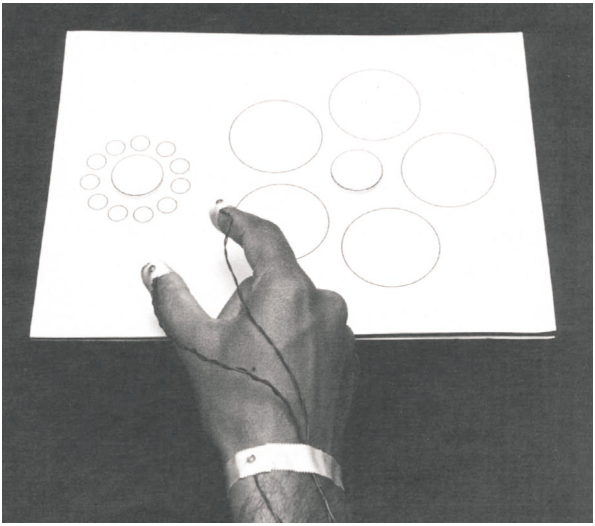

Figure. 3. The 3-D version of the Titchener Circles Illusion used by Aglioti *et al.* (1995).

The trajectory of the subject's grasping movement was recorded with two cameras that tracked infrared light-emitting diodes attached to the index finger, thumb, and wrist. (Reprinted from Aglioti *et al.* [1995] with permission from Elsevier.)

of reference. This would explain why a pictorial, size-contrast illusion may sometimes fool the eye, but not the hand.[9]

Clearly, this interpretation of the experimental findings is incompatible with ACT. ACT can accommodate evidence that, in addition to egocentric spatial information, conscious seeing *also* includes object- or scene-relative spatial information (spatial information that is either not normally accessed by or less heavily weighted by visuomotor action), but it cannot accommodate evidence that conscious seeing simply does not represent the layout of visible objects and surfaces in an egocentric frame of reference.

Fortunately for ACT, this interpretation is open to challenge. Although a final verdict on the comparative effects of visual illusions on action and perception is not yet in the offing, pertinent empirical considerations, to be further discussed below, include the following:

---

[9] See Jacob & Jeannerod (2003) for a similar assessment.

4.1. Many of the studies that have been cited as evidence for TVSH in fact indicate a theoretically significant — though comparatively less pronounced — effect of visual illusions on prehensile action. The original study by Aglioti *et al.* (1995), for instance, found that the Titchener circles illusion had a 2.5 mm effect on perception and a 1.6 mm effect on action (as measured, respectively, by the opening between index finger and thumb in a manual estimate of disk size and grip aperture). Similar findings concerning the effects of illusions on grasp position lead Ellis *et al.* (1999) to conclude that, in general, 'the motor system has access to *both* the illusory perceptual information (presumably obtained from the ventral stream) and the veridical information (presumably obtained from the dorsal stream)' (1999, p. 113). For an overview of the relative effects of various visual illusions on action and perception, see Glover (2004).

4.2. In certain contexts, object-directed actions are robustly influenced by visual illusions. First, under *monocular viewing conditions*, grasping is fully affected by the Titchener Circles illusion (Goodale & Milner 2004, p. 92). The presumption here is that, in the absence of binocular depth information provided by stereopsis and vergence, the dorsal stream automatically 'taps' pictorial depth information available in the ventral stream (Marotta *et al.*, 1997; Marotta *et al.*, 1998; Marotta & Goodale, 1998). Second, when a *brief delay* is imposed between the disappearance of a visual target and the initiation of action in tasks involving 'pantomimed' grasping or pointing, visuomotor mechanisms become fully susceptible to illusion (Goodale *et al.*, 1994; Bridgeman *et al.*, 1997; Hu & Goodale, 2000; Westwood & Goodale, 2003). The presumption here is that the longer time interval permits the dorsal stream to access spatial information temporarily stored in the ventral stream. Indeed, for this reason, visuomotor performance in optic ataxics and subjects with other forms of dorsal stream damage markedly improves with such delay (Milner *et al.*, 2003; Goodale *et al.*, 2004). Third, action is also highly susceptible to perceptual influence when movements are *awkward* and/or *unpracticed* (Gonzalez *et al.*, 2006; 2007b) and, notably, when movements are *slow* (Carey, 2001; Rossetti *et al.*, 2005; Króliczak *et al.*, 2006). Króliczak *et al.* (2006) found that even the high-level Hollow-Face illusion, in which a realistic, concave mask appears to be convex when illuminated from

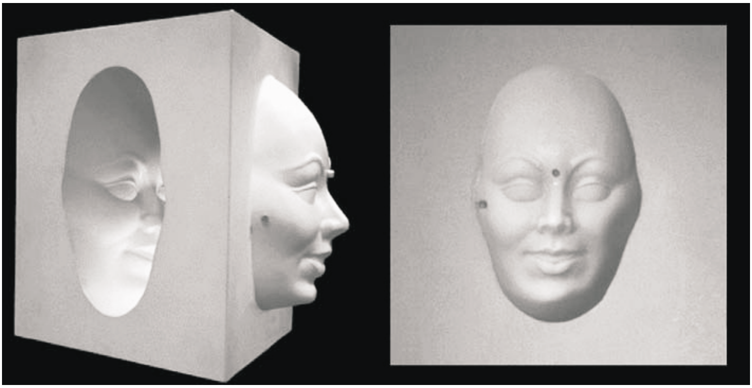

Figure. 4. The Hollow-Face Illusion

On the left is the apparatus used to display the normal (convex) and hollow (concave) faces. On the right is a front view of the hollow face. The face is illuminated from below, creating a pattern of shadows similar to the pattern created by illumination from above for normal, convex faces. Small magnets that serve as visual targets for manual flicking movements are show on the mask's cheek and forehead. (Reprinted from Króliczak *et al.* [2006] with permission from Elsevier.)

below (figure 4),[10] has a strong effect on slow flicking movements directed at magnets affixed on the facing surface of the mask. Finally, Gonzalez *et al*. (2006; 2007b) report that the effects of visual size illusions on grip aperture very much depends on *which hand* is used. Grasping with the left hand was found to be fully influenced by the Titchener Circles and Ponzo illusions in both right-handed and left-handed subjects. This finding suggests that the dorsal stream in the *right hemisphere*, i.e., the hemisphere contralateral to and controlling the left hand, may utilize the same sources of visuospatial information as are present in the ventral stream.

---

[10] This seems to be a top–down effect in which prior knowledge of face geometry and illumination conditions overrides bottom-up information provided by binocular disparity and other cues. The result is illusory convexity and reversed 3-D depth. For discussion, see Gregory (1997). I should note that the results reported by Króliczak *et al*. (2006) are controversial. In particular, Hartung *et al*. (2005), using a somewhat different experimental procedure, report that visual awareness and visually guided reaching are equally susceptible to the hollow face illusion. Importantly, Hartung *et al.* found that, in both cases, illusory hollow faces were estimated by subjects to be somewhat flatter than normal, convex faces. This suggests that the visual system combines bottom-up, depth-specific information with top-down object knowledge and does not completely discount one or the other.

4.3. The studies reviewed in 4.1 and 4.2 provide evidence for the context- and task-sensitive influence of visual illusions on reaching and grasping movements. There is a significant amount of evidence however that visual illusions also have a strong influence on the programming of *saccadic eye movements*. Saccades, e.g., consistently overshoot their targets when made between the endpoints of the subjectively longer, inward-pointing segment of the Müller-Lyer illusion and consistently undershoot their targets when made between the endpoints of the subjectively shorter, outward-pointing segment (Binsted & Elliott, 1999; Binsted *et al.*, 2001). DiGirolamo *et al.* (2001) and McCarley & DiGirolamo (2002) have suggested that the degree of influence of the illusion on oculomotor control is based in part on the type of saccade performed. Voluntary, endogenously driven saccades are influenced by the illusion to the same degree as conscious perception. Reflexive, exogenously driven saccades, by contrast, are also influenced by the illusion, but less pronouncedly so.[11] That said, the finding that even automatic, reflexive saccades are somewhat sensitive to pictorial visual illusions provides strong evidence for early interaction or 'crosstalk' between the two putative processing streams in oculomotor control — arguably a central component of all complex visuomotor performances (returning a fast tennis serve, driving a car, running down a trail, and so on).[12]

4.4. Many of the contextual, pictorial depth cues that sometimes give rise to visual illusions under contrived, ecologically aberrant viewing conditions actually enhance control and guidance of visuomotor action under ecologically normal conditions, i.e., the sorts of terrestrial viewing conditions in which the human visual system evolved. Thus numerous studies have found that object-directed movements are much more accurate when made in a visually structured environment, e.g., against a textured background, than when made in a visually unstructured environment (Proteau & Masson, 1997; Coello & Magne, 2000; Coello & Rossetti, 2004). Indeed, were the general tendency of

---

[11] But for claims to the contrary see Knox & Bruno (2007).

[12] Indeed, independent evidence for this conclusion is provided by a large body of experimental work on overt visual attention, i.e., *gaze*. A multitude of studies have found that high-level, semantic knowledge has a robust influence on the deployment of overt visual attention both when viewing a scene (Hoffman & Subramanian, 1995; Rock & Mack, 1998; Findlay & Gilchrist, 2003) and when engaging in specific visuomotor tasks (Ballard *et al.*, 1995; Hayhoe, 2000; Hayhoe & Ballard, 2005).

contextual depth cues processed in the ventral stream to override or distort accurate sources 3-D spatial information independently available to visuomotor action, the evolutionary propagation of mechanisms devoted to their uptake in vision would make little biological sense. For a brief review of studies of the role played by contextual depth cues in visuomotor action, see Dassonville & Bala (2004).

4.5. Studies 4.1–4.4 above are consistent with the view that perception is somewhat less refractory than action to visual illusions in certain experimental contexts. Proponents of a 'single representation' model of action and perception, however, have suggested that the results reported by Aglioti *et al.* (1995) are the product of an unnoticed asymmetry between the perceptual task and the motor task in the study's experimental design (Pavani *et al.*, 1999; Franz *et al.*, 2000; Franz, 2001; 2003; Franz *et al.*, 2003). In the study by Aglioti and colleagues, *two* disks each surrounded by an annulus of circles were displayed (as in figure 3 above). Consequently, when engaging in the perceptual task, subjects had to calculate the sizes of two disks at the same time. However, when engaging in the motor task, subjects had to calculate the size of only *one* disk at a time. The studies reported in the aforementioned series of papers found that, when this asymmetry is eliminated, i.e., in 'single-context' experiments involving only one disk surrounded by an annulus of circles, and when standard perceptual measures are utilized, visual illusions have *identical effects* on action and perception. Similar objections to the methodology used in Aglioti *et al.* (1995) are raised by Vishton & Fabre (2003) and Vishton (2004). These studies also find that in single-context experiments, the effects of visual illusion on visual perception are significantly attenuated, sometimes to the same level as observed with reaching and grasping behaviours.[13]

4.6. Similar scepticism has recently been motivated with respect to reported differential affects on action and perception in studies of the 'dot-in-frame' illusion (also known as the Induced Roelofs Effect). When a small visual target is surrounded by a large rectangular frame positioned so that the frame's center is offset from

---

[13] Proponents of TVSH have suggested that the effect of the Titchener Circles illusion on action can be explained by the hypothesis that the dorsal stream interprets the annulus of circles in experimental displays as *obstacles* that must be avoided in grasping the central disk (Haffenden & Goodale, 2000; Haffenden *et al.*, 2001; Jacob & Jeannerod, 2003, chap. 4). For criticisms of this hypothesis, see Franz *et al.* (2003).

> the observer's bodily midline, the apparent egocentric location of the target is *shifted* in the direction opposite to the frame's offset. Bridgeman *et al.* (1997) argued that rapid pointing and jabbing are refractory to the illusion, and interpreted the putative finding as evidence that separate spatial representations are respectively used in visual perception and visually guided behaviour. Dassonville *et al*. (2004) and Dassonville & Bala (2004) have recently argued, however, that action *is* influenced by the illusion inasmuch as the surrounding frame — presented to subjects seated in complete darkness — distorts the observer's proprioceptive awareness of her own bodily midline. Under ecologically normal viewing conditions, the center of a subject's visual field serves as an accurate indicator of egocentric 'straight-ahead'. When viewed against a black, visually unstructured background, the large rectangular frame 'attracts' the subject's awareness of her midline in the direction of the frame's offset. In consequence, actions that are directed toward targets within the distorted egocentric reference frame are fully accurate, since the errors of target localization cancel the errors of motor guidance.

Let us now take stock. The studies reviewed in 4.5–4.6 above deny that visual illusions actually affect and action and perception differently. Hence they deny that illusion studies provide any evidence for TVSH. In order to build the strongest possible case for ACT, I shall merely register this here, and instead focus in what follows on the studies earlier reviewed in 4.1–4.4. Assuming their empirical validity for purposes of argument, what is their broad significance for ACT?

On the one hand, the studies reviewed in 4.1–4.4 above clearly seem to indicate that the dorsal stream, especially in the right hemisphere, has ready *access* to sources of spatial content in the ventral stream. Whether and the extent to which the dorsal stream makes use of contextual depth cues and other sources of 3-D spatial information in the ventral stream appears to vary with its task-specific needs and resources (see Clark, 2001, pp. 507–8). Indeed, points made in 4.4 suggest that accessing or 'tapping' such information in the ventral stream, when feasible, generally serves to enhance control and guidance of visuomotor action. But, if this is the case, then the dorsal stream is likely to make use of spatial information in the ventral stream whenever it can afford to do so. The findings reviewed in 4.1–4.4, in short, militate against a robust, i.e., context- and task-invariant, dissociation of the two putative processing streams at the level of spatial content. They suggest a much more complicated picture, one in which the degree of interaction between the two streams depends inter alia on which side of the body (and, so, which hemisphere, right or left) is involved in the action, the availability to dorsal processing of its own bottom-up sources of binocular visual information, and, crucially, time constraints on performance.[14]

On the other hand, many of the studies reviewed in 4.1–4.3 do seem to indicate that, under certain conditions, e.g., when engaging in rapid or automatic reaching with the right hand under binocular viewing conditions or when making reflexive eye movements, visual illusions may have a measurably more *pronounced* effect on perception than on action. Why is this the case? One answer, of course, is provided by TVSH, by the hypothesis that visually guided action uses an egocentric frame of reference incorporating absolute metrical information, while conscious visual awareness uses an allocentric frame of reference incorporating object- or scene-relative metrical information. This hypothesis would explain why perception is less refractory than action to visual illusions involving contextual, pictorial depth cues (see Goodale & Milner, 2004, pp. 73–6). I think that a strong case for ACT will have been made, if I can show that there is another hypothesis, i.e., another plausible interpretation of the studies reviewed in 4.1–4.3, that is not only consistent with ACT, but also raises fewer serious theoretical difficulties than TVSH. This is my objective in the following section.

5. The Integration Hypothesis

According to the alternative interpretation, there is an ACT-friendly explanation for why perception, i.e., conscious visual awareness, may be less refractory to visual illusion than action. The explanation is not, as proponents of TVSH suggest, that perception simply is not in the business of coding egocentric spatial properties, but rather that, in coding egocentric spatial properties, e.g., the distances and orientations of visible surfaces in depth, perception sometimes integrates a wider variety of (fallible) sources of spatial information than does action. In consequence, perception sometimes runs a comparatively

---

[14] I am here focusing on interactions between the two streams at the level of visuospatial content. I am not including higher-level interactions mediated by stored object knowledge in the ventral stream. The point I wish to make is that there is evidence for substantial interaction between the two streams in normal subjects even if we bracket the role played by stored object knowledge in enabling high-level, semantically rich visuomotor engagements with the world.

greater risk of falling subject to visuospatial illusions. I shall call this the 'integration hypothesis.'

The integration hypothesis is supported by abundant psychophysical evidence that perception of 3-D spatial layout involves a linear, weighted averaging of independently variable sources of depth-specific information, including binocular disparity, motion - parallax, occlusion, perspective, texture gradients, shading, and reflections as well as stored, top-down knowledge of object geometry.[15] Less theoretically contentious than the claim that egocentric spatial properties are not represented in perception is simply the hypothesis that their representation in perception involves such a weighted averaging of fallible depth cues and that, sometimes, especially in contrived, ecologically aberrant viewing conditions, certain contextual depth cues may erroneously override or 'veto' other more reliable sources of spatial information. Since the dorsal stream does not attach much relative importance to contextual depth cues in situations in which action is fast or automatic (Dijkerman *et al.*, 1996; Humphrey *et al.*, 1996; Marotta *et al.*, 1997; Mon-Williams *et al.*, 2001) it is less likely to be mislead by them in those situations when these cues are inaccurate. However, when the dorsal stream's preferred sources of spatial information are unavailable or when it has time on its hands (pun intended),[16] the dorsal stream will make use of outputs from ventral processing and, consequently, visuomotor action will run a correspondingly greater risk of falling subject to illusion.

To sum up: Were visual illusions in fact shown sometimes to have a more pronounced effect on perception than on action, this finding would not evidence the absence of egocentric spatial coding in

---

[15] For an introductory overview, see Cutting & Vishton (1995) and Bruce *et al.* (2003, chap. 7). For more detailed treatment, see Deneve & Pouget (2004), Ernst & Bülthoff (2004), and Yuille & Kersten (2006). For discussion of the role of prior knowledge of face geometry in both visual awareness and visually guided action, see Hartung *et al.* (2005).

[16] A study by Greenwald *et al.* (2005) provides strong experimental evidence that depth cue integration for visuomotor control depends in part on the *time course* of cue processing. Greenwald *et al.* (2005) found that in a manual, object placement task, binocular cues provided by disparity and vergence influenced online control for short duration movements much more than monocular cues provided by texture gradients and object contours (a result replicated by Knill, 2005). They also found that the relative influence of binocular cues was more pronounced for the shorter online control phase of movement than for the longer planning phase. These findings were consistent with evidence provided by a 'temporal decorrelation' analysis of trial data that binocular cues are processed faster — and, so, become available for use sooner — than monocular cues in visuomotor control (Greenwald *et al.*, 2004, pp. 1982–3). The relevant point is that the role that a source of depth specific information plays in online control is a function not just of its estimated reliability and consequent weighting (as well as task-specific processing demands), but also of the temporal dynamics of the cue integration process.

conscious visual awareness, as proponents of TVSH maintain. Rather it would simply evidence the greater sensitivity, in certain cases, of egocentric spatial coding in conscious visual awareness to potentially erroneous sources of contextual, depth-specific information (available in the ventral stream) than egocentric spatial coding in visuomotor action.

(Siegel, 2006, points out that the less 'committal' are the contents of visual experience, the less potential for misperception there is. If properties of a particular kind *F* are not represented in visual experience, then the subject cannot incorrectly perceive an object as having or being a certain *F*. Conversely, the more committal are the contents of visual experience, i.e., the more varied the kinds of properties that are taken to be represented in visual experience, the more potential for misperception there is. I am concerned to make a somewhat different point: the more varied are the fallible sources of information used by the visual system to detect egocentric spatial properties, the more varied are the ways in which the visual system may be occasionally misinformed about them. This is the case even if integrating more sources of information *typically* serves to increase the accuracy with which the visual system detects egocentric spatial properties.)

I have shown that there is an ACT-friendly interpretation of the studies reviewed in 4.1–4.3 above, i.e., the integration hypothesis. I shall now proceed to show that the interpretation provided by the integration hypothesis avoids three serious theoretical difficulties that confront TVSH.

First, the idea that perception and action utilize fundamentally different spatial coding systems, integral to TVSH, gives rise to a serious problem about how the two putative systems manage to *communicate* with one another. As Goodale and Milner write, 'the two systems are using entirely different frames of reference — speaking a different language in fact — and yet somehow the ventral stream has to tell the dorsal stream which object to act upon' (2004, p. 101). The problem leads them to speculate that the ventral stream engages in what might be called 'backward flagging.' According to this view, higher-order areas in the ventral stream working together with other cognitive systems can use back-projections to primary visual cortex (V1), the common retinotopic source of information to both streams, in order to 'flag' or 'highlight' targets for the dorsal stream to engage (1995/2006, pp. 231–4; 2004, p. 102). Once a target has been highlighted on the retinal map in primary visual cortex, the dorsal stream can then compute its position relative to relevant parts of the body and initiate action.

It is well known that higher-order areas can prime lower-order areas in visual processing. Shifts of visual attention, as demonstrated by neurophysiological research, e.g., can actively modulate cell response patterns in the putative visual processing hierarchy as early as primary visual cortex (Ito & Gilbert, 1999). There are two serious theoretical difficulties with the backward flagging view, however. The first difficulty is that many of the studies reviewed in 4.1–4.4 above point to a significant amount of high-level crosstalk or leakage between the two streams in normal subjects. Indeed, it seems clear that much more than mere retinotopic object location may be leaked to the dorsal stream from the ventral stream. In the study reported in Króliczaka *et al.* (2006), e.g., high-level knowledge of faces and normal illumination conditions stored in the ventral stream appears to override low-level depth information provided by binocular disparity in the dorsal stream when action is slow. Moreover, there is evidence, as we saw, that the dorsal stream in the right hemisphere — controlling action on the left side of the body — may utilize substantially the same sources of 3-D visuospatial information as are present in the ventral stream (Gonzalez *et al.*, 2006; 2007b). These considerations suggest that not only are there direct, high-level communication links between the two streams, but also that some representational contents in the ventral stream are already in a format that the dorsal stream is able to understand.

The second theoretical difficulty is that experimental data on delayed or 'pantomimed' pointing and grasping, much of it garnered by Milner and Goodale themselves, indicate that the dorsal stream is able to tap briefly stored visual representations in the ventral stream shortly after the visual target has *disappeared*. Since, in relevant cases, the object is no longer seen, there is no area on the retinotopic map in primary visual cortex corresponding to the object for the ventral stream to flag. In order to initiate action in respect of a target after its disappearance, it again seems that the dorsal stream must be able to make direct use of spatial information (stored in visual memory) in the ventral stream.

One important merit of the ACT-friendly interpretation of the evidence provided by the integration hypothesis is that it bypasses the communication problem and the need to postulate something like backward flagging. Since, according to this interpretation, *both* perception, i.e., conscious visual awareness, and action make use of an egocentric frame of reference, perception has no problem when it comes to telling action upon which object to act. The testimony of the senses is delivered in a language that the body understands.

Perhaps a more significant merit of the ACT-friendly interpretation is that it comports with the widely accepted view that 'mid-level'[17] representation of 3-D surface layout in a viewer-centered, i.e., egocentric, frame of reference plays a crucial role in a variety of putative *ventral* processing tasks. In particular, a wide array of experimental and phenomenological evidence suggests that high-level object recognition, the ventral stream's putative *raison d'etre*, is significantly dependent for input on a more general purpose competence to perceive scenes in terms of surfaces egocentrically arrayed in depth (Marr, 1982; Nakayama *et al.*, 1995; Fleming & Anderson, 2004). Thus, Nakayama *et al.* (1995) argue that 'we cannot think of object recognition as proceeding from image properties … there needs to be an explicit parsing of the image into surfaces [in a viewer-centred frame of reference]. Without such parsing of surfaces, object recognition cannot occur' (p. 15). If this is correct, then a third serious theoretical difficulty faced by TVSH is that the ventral stream in order to perform its reputed functional role must, contrary to TVSH, generate or have access to representations of visible surface layout in an egocentric frame of reference.

In concluding this section, I would like briefly to return to the findings mentioned earlier about hemispheric differences with respect to the sources of depth-specific information used in visuomotor control (see section 4.2 above). Gonzalez *et al.* (2007b) write:

> There is some suggestion that the visuomotor networks controlling grasping make use of information derived from vergence and retinal image size, rather than pictorial cues, to fine-tune grip aperture… Whatever the specific cues might be, left-hemisphere mechanisms appear to play a crucial role in the control of these skilled movements. But this seems to be true only for the visual control of skilled movements with the right hand. Even though practice might make movements with the left hand faster and more efficient, the visual control of these movements cannot escape the relational metrics that characterize visual perception. Only movements with the right hand, perhaps because they have direct sensorimotor links with the left hemisphere, make use of real world metrics (ms. p. 7, in press).

---

[17] Mid-level vision is so called because it is poised in the putative visual-processing hierarchy between bottom-up, 'low-level' image analysis and top-down, 'high-level' object recognition. Mid-level vision represents only the orientation and relative distances of non-occluded object surfaces (and surface properties such as color and texture) in a viewer-centered, i.e., egocentric, frame of reference. It does not concern itself with object identities or their complete, volumetric shapes. Hence, there is an affinity between mid-level vision and what David Marr (1982) called the '2½-D sketch' of a scene. For arguments that the representational contents of visual awareness are to be identified with those of mid-level vision, see Jackendoff (1987, chaps. 9 and 14) and Prinz (2000).

One source of evidence for this view comes from a recent study of spontaneous hand preference in a ‘natural’ grasping task involving pieces of Lego© arrayed on a table (Gonzalez *et al.*, 2007a). Left-handers were found to be far more likely than right-handers to use their non-dominant hand to pick up pieces of Lego© anywhere on the table, even on the side *opposite* to that hand. In fact, left-handers used their left hand for precise, pincer grasping only 47.8% of the time, while right-handers used their right hand 80.2% of the time.

I think that the evidence garnered by Gonzalez and her colleagues as well as other findings (Hopkins *et al.*, 2002; 2005; Lewis *et al.*, 2006) provide fairly compelling evidence for the existence of a left-hemisphere specialization in the visual control of skilled prehension. The suggestion, however, that subjects do not make use of an egocentric frame of reference and real world metrics when using their left hand strikes me as theoretically profligate. A much more parsimonious interpretation — one notably in keeping with the integration hypothesis — would simply be that visuomotor processing controlling the left hand in the right hemisphere integrates *more* sources of depth-specific information, some of which are presumably coming via ventral stream, than does visuomotor processing controlling the right hand.[18] In other words, subjects *do* make use of an egocentric frame of reference when using their left hand, but they draw on a wider variety of depth cues than they do when using their right hand. This interpretation would explain why subjects are more susceptible to visual illusions when using their left hand than when using their right hand: the more varied are the independently variable (and fallible) sources of information used by the visuomotor system to encode the spatial properties of visible objects, the more varied are the ways in which the visuomotor system may be occasionally misinformed about them.[19]

The main point, then, is that the ACT-friendly integration hypothesis can explain differences between left- and right-handed visuomotor performance without recourse to the suggestion that subjects do not make use of an egocentric frame of reference and real-world metrics when using their left hand. Proponents of TVSH are lead to this

---

[18] At least when action is fast and fluent. As mentioned earlier, when action is either slow or unpracticed, visuomotor processing controlling the right-hand in the left hemisphere also appears to tap information in the ventral stream.

[19] It is important to emphasize once more that in ecologically normal viewing conditions (as opposed to informationally impoverished experimental viewing conditions) pictorial depth cues typically tend to corroborate — not to distort or override — other sources of 3-D spatial information. The evolutionary propagation of their uptake in vision would otherwise make little biological sense.

suggestion by evidence that the same sources of spatial information are implicated in visuomotor performance involving the left-hand as are implicated in conscious visual awareness and, crucially, by their assumption that conscious visual awareness only makes use of allocentric frames of reference and relational metrics. However, if, as the integration hypothesis claims, conscious visual awareness localizes objects in an egocentric frame of reference, but does so using more potentially erroneous sources of depth-information than are available to the dorsal stream in the left hemisphere (at least when action is fast and fluent), then there is no need to countenance the possibility that subjects make use of totally different spatial frames of reference when using their left and right hands respectively.

6. Conclusion: The Dorsal Stream and Conscious Visual Experience

The claim that personal-level visual awareness and visually based action both utilize an egocentric spatial content base depends for its plausibility on clearly distinguishing between representational *content* and representational *vehicle*. What is common to visual awareness and visually based action, according to ACT, is to be typed at the level of *content*. Hence, conscious seeing can deliver us with the same kind of egocentric spatial content as is utilized in visuomotor action, even if the representational *vehicles* that respectively support conscious seeing and visuomotor action are sometimes physically different.

It should be noted in this connection that, in defending ACT, it is possible to remain fairly noncommittal about the specific contributions of subpersonal ventral and dorsal stream processing, respectively, to spatially contentful visual awareness in healthy subjects. In particular, it is possible to resist Milner and Goodale's theoretical characterization of the ventral stream as a functionally independent 'vision for perception' system.[20] While it seems clear that dorsal stream processing by itself is functionally insufficient for normal visuospatial awareness — as evidenced in part by the profound *visual form agnosia* consequent upon trauma to the ventral stream in Milner and Goodale's famous subject DF (Milner & Goodale, 1995/2006) — there is significant neuropsychological evidence that much dorsal stream processing may nonetheless be functionally necessary. Thus Balint syndrome, caused by damage to the superior parietal lobe, is

---

[20] For recent neuropsychological resistance to TVSH, see Rizzolatti & Matelli (2003); Gallese (2005; 2007); Pisella *et al.* (2006); and Coello *et al.* (2007).

characterized not only by *optic ataxia*, i.e., deficits in reaching toward and grasping targets in peripheral, contralesional vision, but also by 'sticky' or paralyzed gaze (Moreaud, 2003). In consequence, a subject with Balint syndrome may have great difficulty in perceiving more than one object or part of space at a time. Damage to the superior parietal lobe may also sometimes result in 'visual extinction' in which stimuli on the contralesional side of visual space are completely absent from the subject's visual experience.[21] Last, there is preliminary evidence that deficits in reaching toward objects in non-foveal vision in optic ataxics may be due in part to impaired *perception* in the peripheral visual field (Michel & Hénaff, 2004; Rosetti *et al.*, 2005a,b). Since the relevant deficits are caused by damage to dorsal stream areas — in particular, areas in the superior parietal lobe and intraparietal sulcus — this would again suggest that such damage may result not only in (pure) visuomotor deficits, but also in perceptual deficits. The claim that conscious visual experience makes use of an egocentric frame of reference, accordingly, then, should not be understood as the claim that the ventral stream considered in isolation, i.e., apart from its many and complex interactions with the dorsal steam (and various other areas in the brain), causally supports or enables conscious visual experience of egocentric space.[22] Detailing those interactions will surely be the goal of much future neuropsychological research.

Acknowledgments

For helpful discussion of ideas presented in this paper I am grateful to Larry Hardesty, Ruth Garrett Millikan, and Alva Noë. Two anonymous referees provided detailed comments and suggestions that resulted in significant improvements.

References

Aglioti, S., Goodale, M. and DeSouza, J. (1995), 'Size contrast illusions deceive the eye but not the hand', *Current Biology*, **5**, pp. 679–85.

---

[21] For a useful assessment, see Milner (1997).

[22] Researchers in the Gibsonian tradition have long maintained that visual awareness of 3-D space has a crucially *pragmatic* significance for the perceiving animal. On their view, there is no purely passive visual awareness of an object's shape, size, location, etc. Rather to be visually aware of such spatial properties is, in the first instance, to understand (if only implicitly) the various possibilities that they afford for bodily movement and action. If this Gibsonian view — integral to ACT (see section 2 above) — is correct, and, if, as neuropsychological evidence suggests, object-directed visuomotor action is crucially dependent on processing in the dorsal stream, then it follows that the dorsal stream may actually have a quite central role to play in normal, visuospatial awareness. Adequately developing and defending this contention unfortunately, however, must remain a task for another occasion.

Ballard, D., Hayhoe, M. and Pelz, J. (1995), 'Memory representations in natural tasks', *Journal of Cognitive Neuroscience*, **7**, pp. 66–80.

Ballard, D., Hayhoe, M., Pook, P. and Rao, R. (1997), 'Deictic codes for the embodiment of cognition', *Behavioral and Brain Sciences*, **20**, pp. 723–67.

Battaglia-Mayer, A. and Caminiti, R. (2002), 'Optic ataxia as the result of the breakdown of the global tuning fields of parietal neurones', *Brain*, **125**, pp. 225–37.

Bermúdez, J. (1998), *The Paradox of Self-Consciousness* (Cambridge, MA: MIT Press).

Bermúdez, J. (2007), 'From two visual systems to two forms of content?', *Psyche*, **13**(2) (http://psyche.cs.monash.edu.au/symposia/seeing/Bermudez.pdf).

Bermúdez, J., Marcel, A. and Eilan, N. (ed. 1995), *The Body and the Self* (Cambridge, MA: MIT Press.)

Binsted, G. and Elliott, D. (1999), 'The Müller-Lyer illusion as a perturbation to the saccadic system', *Human Movement Science*, **18**, pp. 103–17.

Binsted, G., Elliott, D., Helsen, W. and Chua, R. (2001), 'Eye-hand coordination in goal-directed aiming', *Human Movement Science*, **20**, pp. 563–85.

Briscoe, R. (2008), 'Vision, action, and make-perceive', *Mind and Language*, **23**, pp. 457–96.

Bruce, V., Greene, P. and Georgeson, M. (ed. 2003), *Visual Perception: Physiology, Psychology, and Ecology*, 4th ed. (London: Psychology Press.)

Campbell, J. (1994), *Past, Space, and Self* (Cambridge, MA: MIT Press).

Campbell, J. (2002), *Reference and Consciousness* (Oxford: OUP).

Carey, D. (2001), 'Do action systems resist visual illusions?', *Trends in Cognitive Sciences*, **5**, pp. 109–13.

Clark, A. (1999), 'Visual awareness and visuomotor action', *Journal of Consciousness Studies*, **6**(11–12), pp. 1–18.

Clark, A. (2001), 'Visual experience and motor action: Are the bonds too tight?', *The Philosophical Review*, **110**, pp. 495–519.

Coello, Y. and Magne, P. (2000), 'Determination of target position in a structured environment: Selection of information for action', *European Journal of Cognitive Psychology*, **12**, pp. 489–519.

Coello, Y. and Rossetti, Y. (2000), 'Planning and controlling action in a structured environment: Visual illusion without dorsal Stream', *Behavioral and Brain Sciences*, **27**, pp. 29–31.

Coello, Y., Danckert, J., Blangero, A. and Rossetti, Y. (2007), 'Do visual illusions probe the visual brain? Illusions in action without a dorsal visual stream', *Neuropsychologia*, **45**, pp. 1849–58.

Cutting, J. and Vishton, P. (1995), 'Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth', in W. Epstein and S. Rogers, eds., *Perception of Space and Motion* (San Diego: Academic Press).

Daprati, E. and Gentilucci, M. (1997), 'Grasping an illusion', *Neuropsychologia*, **35**, pp. 1577–82.

Dassonville, P. and Bala, J. (2004), 'Perception, action, and Roelofs Effect: A mere illusion of dissociation', *PLoS Biology* **2**(11: e364), pp. 1936–45.

Dassonville, P., Bridgeman, B., Bala, J., Thiem, P. and Sampanes, A. (2004), 'The Induced Roelofs Effect: Two visual systems or the shift of a single reference frame?' *Vision Research*, **44**, pp. 603–11.

Deneve, S. and Pouget, A. (2004), 'Bayesian multisensory integration and cross-modal spatial links', *Journal of Physiology*, **98**, pp. 249–58.

Dijkerman, H., Milner, A.D. and Carey, D. (1996), 'The perception and prehension of objects oriented in the depth plane. I. Effects of visual form agnosia', *Experimental Brain Research*, **112**, pp. 442–51.

Ellis, R., Flanagan, J., and Lederman, S. (1999), 'The influence of visual illusions on grasp position', *Experimental Brain Research*, **125**, pp. 109–14.

Ernst, M. and Bülthoff, H. (2004), 'Merging the senses into a robust percept', *Trends in Cognitive Sciences*, **8**, pp. 162–9.

Evans, G. (1982), *The Varieties of Reference*, J. McDowell, ed. (Oxford: Oxford University Press).

Findlay, J. and Gilchrist, I. (2003), *Active Vision: The Psychology of Looking and Seeing* (Oxford: Oxford University Press).

Fleming, R. and Anderson, B. (2004), 'The perceptual organization of depth', in L. Chalupa and J. Werner, eds., *The Visual Neurosciences* (Cambridge, MA: MIT Press).

Franz, V. (2001), 'Action does not resist visual illusions', *Trends in Cognitive Sciences*, **5**, pp. 457–59.

Franz, V. (2003), 'Manual size estimation: A neuropsychological measure of perception?', *Experimental Brain Research*, **151**, pp. 471–77.

Franz, V., Gegenfurtner, K., Bülthoff, H. and Fahle, M. (2000), 'Grasping visual illusions: No evidence for a dissociation between perception and action', *Psychological Science*, **11**, pp. 20–25.

Franz, V., Bülthoff, H., and Fahle, M. (2003), 'Grasp effects of the Ebbinghaus illusion: Obstacle avoidance is not the explanation', *Experimental Brain Research*, **149**, pp. 470–77.

Gallagher, S. (2005), *How the Body Shapes the Mind* (Oxford: OUP).

Gallese, V. (2005), 'Embodied simulation: From neurons to phenomenal experience', *Phenomenology and the Cognitive Sciences*, **4**, pp. 23–48.

Gallese, V. (2007), 'The "conscious" dorsal stream: Embodied simulation and its role in space and action conscious awareness', *Psyche*, **13** (http://psyche.cs.monash.edu.au/).

Glover, S, (2004), 'Separate visual representations in the planning and control of action', *Behavioral and Brain Sciences*, **27**, pp. 3–78.

Gonzalez, C., Ganel, T. and Goodale, M. (2006), 'Hemispheric specialization for the visual control of action is independent of handedness', *Journal Neurophysiology*, **95**, pp. 3496–501.

Gonzalez, C., Whitwell, R., Morrissey, B., Ganel, T. and Goodale, M. (2007a), 'Left handedness does not extend to visually guided precision grasping', *Experimental Brain Research*, **182**, pp. 275–79.

Gonzalez, C., Ganel, T., Whitwell, R., Morrissey, B. and Goodale, M. (2007b), 'Practice makes perfect, but only with the right hand: Sensitivity to perceptual illusions with awkward grasps decreases with practice in the right but not the left hand', *Neuropsychologia*, in press.

Goodale, M. and Humphrey, K. (1998), 'The objects of action and perception', *Cognition*, **67**, pp. 181–207.

Goodale, M. and Milner, A.D. (2004), *Sight Unseen* (Oxford: Oxford University Press).

Goodale, M., Jakobson, L. and Keillor, J. (1994), 'Differences in the visual control of pantomimed and natural grasping movements', *Neuropsychologia*, **32**, pp. 1159–78.

Goodale, M., Westwood, D., and Milner, A.D. (2004), 'Two distinct modes of control for object-directed action', *Progress in Brain Research*, **144**, pp. 131–44.

Greenwald, H., Knill, D. and Saunders, J. (2005), 'Integrating visual cues for motor control: A matter of time', *Vision Research*, **45**, pp. 1975–89.

Gregory, R. (1997), ‘Knowledge in perception and illusion’, *Philosophical Transactions of the Royal Society of London* B, **352**, pp. 1121–28.

Grush, R. (1998), ‘Skill and spatial content’, *Electronic Journal of Analytic Philosophy***, 6** (http://ejap.louisiana.edu/EJAP/1998/grusharticle98.html).

Grush, R. (2000), ‘Self, world, and space: The meaning and mechanisms of ego- and allocentric spatial representation’, *Brain and Mind***, 1**, pp. 59–92.

Grush, R. (forthcoming), ‘Skill theory v2.0: Dispositions, emulation, and spatial perception’, to appear in *Synthese*.

Hartung, B., Schrater, P., Bülthoff, H., Kersten, D., and Franz, V. (2005), ‘Is prior knowledge of object geometry used in visually guided reaching?’, *Journal of Vision*, **5**, pp. 504–14.

Hayhoe, M. (2000), ‘Vision using routines’, *Visual Cognition*, **7**, pp. 43–64.

Hayhoe, M. and Ballard, D. (2005), ‘Eye movements in natural behavior’, *Trends in Cognitive Sciences*, **9**, pp. 188–94.

Henriques, D., Medendorp, W., Khan, A. and Crawford, J. (2002), ‘Visuomotor transformations for eye–hand coordination’, *Progress in Brain Research,* **140**, pp. 329–40.

Hoffman, J. and Subramanian, B. (1995), ‘Saccadic eye movements and visual selective attention’, *Perception and Psychophysics*, **57**, pp. 787–95.

Hopkins, W., Cantalupo, C., Freeman, H., Russell, J., Kachin, M. and Nelson, E. (2005), ‘Chimpanzees are right-handed when recording bouts of hand use’, *Laterality*, **10**, pp. 121–30.

Hopkins, W., Cantalupo, C., Wesley, M., Hostetter, A. and Pilcher, D. (2002), ‘Grip morphology and hand use in chimpanzees (pan troglodytes): Evidence of a left hemisphere specialization in motor skill’, *Journal of Experimental Psychology*, **131**, pp. 412–23.

Hu, Y. and Goodale, M. (2000), ‘Grasping after a delay shifts size-scaling from absolute to relative metrics’, *Journal of Cognitive Neuroscience*, **12**, pp. 856–68.

Humphrey, G., Symons, L., Herbert, A. and Goodale, M. (1996), ‘A neurological dissociation between shape from shading and shape from edges’, *Behavioural Brain Research*, **76**, pp. 117–25.

Hurley, S. and Noë, A. (2007), ‘Can hunter-gatherers hear color?’ in G. Brennan, R. Goodin, and M. Smith, eds., *Common Minds: Essays in Honor of Philip Pettit* (Oxford: Oxford University Press).

Ito, M. and Gilbert, C. (1999), ‘Attention modulates contextual influences in the primary visual cortex of alert monkeys’, *Neuron*, **22**, pp. 593–604.

Jackendoff, R. (1987), *Consciousness and the Computational Mind* (Cambridge, MA: MIT Press).

Jackson, S. and Shaw, A. (2000), ‘The Ponzo illusion affects grip-force but not grip-aperture scaling during prehension movements’, *Journal of Experimental Psychology HPP*, **26**, pp. 418–23.

Jacob, P. and Jeannerod, M. (2003), *Ways of Seeing* (Oxford: Oxford University Press).

Jeannerod, M. (1997), *The Cognitive Neuroscience of Action* (Oxford: Blackwell).

Knill, D. (2005), ‘Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception’, *Journal of Vision*, **5**, pp. 103–15.

Knox, P. and Bruno, N. (2007), ‘When does action resist visual illusion? The effect of Müller–Lyer stimuli on reflexive and voluntary saccades’, *Experimental Brain Research*, **181**, pp. 277–87.

Króliczak, G., Heard, P., Goodale, M., and Gregory, R. (2006), ‘Dissociation of perception and action unmasked by the hollow-face illusion’, *Brain Research*, **1080**, pp. 9–16.

Lewis, J., Phinney, R., Brefczynski-Lewis, J. and DeYoe, E. (2006), 'Lefties get it "right" when hearing tool sounds', *Journal of Cognitive Neuroscience* **18**, pp. 1314–30.

Marcel, A. (2003), 'The sense of agency: Awareness and ownership of action' in J. Roessler and N. Eilan, eds., *Agency and Self-Awareness: Issues in Philosophy and Psychology* (Oxford: Oxford University Press).

Marotta, J., Behrmann, M., and Goodale, M. (1997), 'The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia', *Experimental Brain Research*, **116**, pp. 113–121.

Marotta, J., DeSouza, J., Haffenden, A. and Goodale, M. (1998), 'Does a monocularly presented size-contrast illusion influence grip aperture?', *Neuropsychologia*, **36**, pp. 491–97.

Marotta, J. and Goodale, M. (1998), 'The role of learned pictorial cues in the programming and control of grasping', *Experimental Brain Research*, **121**, pp. 465–70.

Marr, D. (1982), *Vision* (New York: W.H. Freeman and Sons).

McCarley, J. and DiGirolamo, G. (2002), 'One visual system with two interacting visual streams', *Behavioral and Brain Sciences*, **24**, pp. 112–13.

Michel, F., and Hénaff, M. (2004), 'Seeing without the occipito-parietal cortex: Simultagnosia as a shrinkage of the attentional visual field', *Behavioural Neurology*, **15**(1–2), pp. 3–13.

Millikan, R. (2006), 'Postulating perceptual representations in a way that actually supports Gibson's central insights'. Lecture at the Boston Colloquium for Philosophy of Science, January 30, 2006.

Milner, A.D. (1997) 'Neglect, extinction, and the cortical streams of visual processing', in P. Thier and H. Karnath, eds., *Parietal Lobe Contributions to Orientation in 3D Space* (New York: Springer).

Milner, A.D. and Goodale, M. (1995/2006), *The Visual Brain in Action*, 2nd edition (Oxford: Oxford University Press).

Milner, A.D., Dijkerman, H., McIntosh, R., Rossetti, Y. and Pisella, L. (2003), 'Delayed reaching and grasping in patients with optic ataxia', *Progress in Brain Research*, **142**, pp. 225–42.

Mon-Williams, M., Tresilian, J., McIntosh, R. and Milner, A.D. (2001), 'Monocular and binocular distance cues: Insights from visual form agnosia I (of III)', *Experimental Brain Research*, **139**, pp. 127–36.

Nakayama, K., He, Z. and Shimojo, S. (1995), 'Visual surface representation: A critical link between lower-level and higher-level vision', in S. Kosslyn and D. Osherson, eds., *Visual Cognition* (Cambridge, MA: MIT Press).

Palmer, S. (1999), *Vision Science*: *Photons to Phenomenology* (Cambridge, MA: MIT Press).

Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., and Farnè, A. (1999), 'Are perception and action affected differently by the Titchener circles illusion?', *Experimental Brain Research*, **127**, pp. 95–101.

Peacocke, C. (1992), *A Study of Concepts* (Cambridge, MA: MIT Press).

Pisella, L., Binkofski, F., Lasek, K., Toni, I. and Rossetti, Y. (2006), 'No double-dissociation between optic ataxia and visual agnosia: Multiple sub-streams for multiple visuo-manual integrations', *Neuropsychologia* ,**44**, pp. 2734–48.

Prinz, J. (2000), 'A neurofunctional theory of visual consciousness', *Consciousness and Cognition*, **9,** pp. 243–59.

Proteau, L. and Masson, G. (1997), 'Visual perception modifies goal-directed movement control: Supporting evidence from a visual perturbation paradigm', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **50A**, pp. 726–41.

Rizzolatti, G. and Matelli, M. (2003), 'Two different streams form the dorsal visual system: Anatomy and functions', *Experimental Brain Research*, **153**, pp. 146–57.

Rossetti, Y., Pisella, L. and Vighetto, A. (2003), 'Optic ataxia revisited: visually guided action versus immediate visuomotor control', *Experimental Brain Research*, **153**, pp. 171–79.

Rossetti, Y., Goldenberg, G. and Rode, G. (2005a), 'Current issues in sensorimotor rehabilitation', in H. Freund, M. Jeannerod, M. Hallett, and Leguarda, eds., *Higher-Order Motor Disorders* (Oxford: Oxford University Press).

Rossetti, Y., Revol, P., McIntosh, R., Pisella, L., Rode, G., Danckert, J., Tilikete, C. Dijkerman, H. Boisson, D., Vighetto, A., Michel, F., and Milner, A.D. (2005b), 'Visually guided reaching: Bilateral posterior parietal lesions cause a switch from fast visuomotor to slow cognitive control', *Neuropsychologia*, **43**, pp. 162–77.

Siegel, S. (2006), 'Which properties are represented in perception?', in T. Gendler and J. Hawthorne, eds., *Perceptual Experience* (Oxford: OUP).

Taylor, C. (1979/1995), 'The validity of transcendental arguments', in *Philosophical Arguments* (Cambridge, MA: Harvard University Press).

Vishton, P. (2004), 'Human vision focuses on information relevant to a task, to the detriment of information that is not relevant', *Behavioral and Brain Sciences*, **27**, pp. 53–54.

Vishton, P. and Farbe, E. (2003), 'Effects of the Ebbinghaus illusion on different behaviors', *Spatial Vision*, **16**, pp. 377–92.

Westwood, D. and Goodale, M. (2003) 'Perceptual illusion and the real-time control of action', *Spatial Vision*, **16**, pp. 243–54.

Yuille, A. and Kersten, D. (2006). 'Vision as Bayesian inference: Analysis by synthesis?', *Trends in Cognitive Sciences*, **10**, pp. 301–8.

Paper received December 2007