Mid-Range Action-Driving Visual Information and the Milner and Goodale 'Two Visual Systems' Hypothesis

Abstract

Milner and Goodale (1995, 2006, 2008) have advanced a justly influential theory of the structure of the human visual system. In broad outline, Milner and Goodale hold that the ventral neural pathway is associated with recognition and experiential awareness, and with a kind of indirect control of action. And they hold that, by contrast, the dorsal neural stream is associated with the non-conscious, direct control of visually informed action. Most of the relevant empirical research has focused on the visual control of close-in, "personal space," reaching and/or grasping. While their view has not escaped controversy and debate, we think that Milner and Goodale have a compelling story to tell about the visual guidance of such close in, "personal space" action. However, the question of whether their scheme applies to behavior with visually specified targets that are outside of the space accessible by simply reaching and grasping is largely unstudied. And it should be studied, since this is the arena of much human behavior. We provide good reason to think that the Milner and Goodale scheme may well not apply to important classes of mid-range, 'action space' behavior. But the matter is in the end to be decided in the lab. We describe a research program to do this.

David J. Bennett[1]

Brown University

Patrick S. Foo[2]

University of North Carolina, Asheville

Milner and Goodale (1995, 2006, 2008)[3] have advanced a justly influential theory of the structure of the human visual system. Their view and program has shaped a good deal of research in the vision sciences. It has also drawn the attention of philosophers (cf. Clark, 2001). In broad outline, Milner and Goodale hold that the ventral neural pathway is associated with recognition and experiential awareness and with a kind of indirect control of action. They also hold that, by contrast, the dorsal stream is associated with the non-conscious, direct (in a sense) control of visually informed action. Most of the relevant research in visual neuroscience, however, has focused on the visual control of close-in, "personal space," reaching and grasping. While their views have been

[1] Brown University, Box 1918, Providence, RI 02912 (David_Bennett@Brown.edu)

[2] University of North Carolina, Asheville, Asheville, NC 28804 (pfoo@unca.edu)

[3] When we make reference simply to views of "Milner and Goodale" we are referring to themes found in many places in their work, including in the over-views cited here, as well as Goodale and Milner (2004).

vigorously debated, we think that Milner and Goodale have a compelling story to tell about this kind of "personal space" action. The question of whether their scheme applies to behavior with visually specified aims or targets in mid-range, "action space," however, is largely unstudied. This issue should be studied, since the mid-range is the arena of much human behavior. We provide good reason to think that the Milner and Goodale scheme may well not apply to important classes of mid-range action. This includes locomotion to targets at certain perceived distances from a viewpoint, and the passage through openings viewed initially from some distance. But the matter is, in the end, to be decided in the lab. We describe a research program to do this.

In section 1 below, we clarify the Milner and Goodale "two visual systems" scheme of visual processing. We also clarify the distinction between close-in, "personal space" and mid-range, "action space." In section 2, we give reasons to think that the Milner and Goodale scheme does not hold of behavior driven by the visual detection of the egocentric distance/depth of targets of mid-range, action space behavior, and we describe experiments that would decide the matter. We also clarify the logic of tests of the Milner and Goodale two visual systems scheme. In section 3, we argue that there is reason to believe that the Milner and Goodale scheme does not apply to action guided by the mid-range visual detection of body-scaled size. We again describe experiments that would decide the matter. In section 4, we take up Milner's and Goodale's presentation of anecdotal evidence that their patient D.F.[4] is able to navigate well in the mid-range (Milner & Goodale, 2004). We also address a brief empirical report on D.F.'s stepping performance (Patla & Goodale, 1996). The results of this study and the Milner and Goodale anecdotal observations may have led people to think that conclusions reached about close-in, personal space behavior generalize to mid-range, action space behavior. But we raise doubts about any such conclusion. Section 5 sums up and points forward.

Appendix I describes "virtual reality" (VR) experimental settings. This serves as background for our experiment proposals. When we allude to VR settings in proposing experiments we always intend what is described in this appendix as "free walking" "immersive" VR. Appendix II briefly argues that our hypotheses are compatible with what is currently known from visual neuroscience about the architecture of the human visual system.

In the present paper we are only able to present hypotheses, and propose tests of them: experiments that we hope others will run. We lack the (very significant, VR) resources to run the tests ourselves. Still, we think there is value in pointing out that certain important questions concerning the Milner and Goodale proposals both remain open and are susceptible to test. We think it likely that one reason this is not appreciated (in the areas we point to) is lack of full clarity about aspects of the logic of testing the Milner and Goodale hypotheses. One aim of the current paper is to bring some further

[4] This is the famous patient of Milner and Goodale, who suffered significant brain damage as a result of accidental carbon monoxide poisoning, concentrated especially in ventral stream areas (for recent review of what is known about D.F.'s brain damage see the epilogue in Milner & Goodale, 2006). D.F. can perform many visually guided actions adequately but is severely impaired in tests of experiential awareness. (See also Appendix II below, "The view from visual neuroscience.")

clarity to the logic of experiment design and of experiment assessment in this area of research.

The Basic Milner and Goodale Two Visual Systems Scheme

Milner and Goodale (1995, 2006, 2008; see also Goodale & Milner, 2004) propose a scheme of visual processing with the following features:

Ventral	Dorsal
1. Ventral neural locus	dorsal neural locus
2. Relative/allocentric (and more)	absolute/egocentric (and more)
3. Conscious	non-conscious
4. Indirectly action guiding	directly action guiding

Table 1

Here we use the headings "ventral" and "dorsal" to summarize two broad kinds of visually-based processing, with certain hypothesized characteristics listed below these headings. Of course, the first of these characteristics concerns the (proposed) ventral and dorsal neural loci of processing. It is then claimed that the psychological-coding properties associated with rows 2 – 4 have (left to right) ventral and dorsal stream or pathway neural loci. To sharpen the theory we must clarify the distinctions 2 – 4.

It may seem initially odd that what is probably the clearest, simplest, sharpest distinction in the lot concerns consciousness (row 3). The root distinction for our purposes is that between neural-informational processes that shape conscious experiential awareness, and those that don't. We do not make any substantive commitments about why or how such shaping of experiential awareness takes place. It is also important that there are strong, well founded expectations that visual sensitivity to the stimuli we describe will clearly and dramatically shape conscious, experiential awareness. These will very likely not be borderline or controversial cases, where we need developed theories of consciousness to referee whether there is indeed conscious, experiential awareness of the kind described.[5] This we take to be a desideratum in experiments testing the Milner and Goodale scheme. Our chief hypothesis or conjecture is that there are important cases of visually guided mid-range action where certain visual information is both saliently reflected in experiential awareness and also directly guides behavior. Pace the Milner and Goodale Table 1 scheme.

[5] Hereafter, whenever we speak only of "experience" we mean to imply that the subject is, ipso facto, in a conscious state.

The nature of the row 4 distinction between the direct and indirect control of action is not quite so immediately evident. An attempt at complete sharpening would no doubt take up another full paper. But Milner and Goodale (1995, 2006, 2008) invoke an intuitive distinction that will suffice to get us started, after a little massaging. We really need two distinctions. To begin, Milner and Goodale distinguish a kind of "indirect" (Milner & Goodale, 2008, p. 775) guidance of action (that they associate with the ventral stream[6]). Milner and Goodale hold that such indirect control involves the setting of aims or goals of action, and the selection of broad action types. So, there might, say, be a decision to "eat that apple" (goal), by first "reaching" for it (action type). In contrast (on the Milner and Goodale view), there is also "direct," online, sustained shaping of fine grained movement by visual information. This is needed in order to actually achieve the selected aim. An example would be the visual control of precise reach trajectory and grasp dynamics in reaching out to pick up an apple. Milner and Goodale hold that this sort of direct control of action is accomplished without the guiding information being reflected in conscious awareness. They propose that such non-conscious online visual guidance constitutes the "direct foundation of action" (Milner & Goodale, 2008, p. 775). On their view, this kind of direct control of action is associated with dorsal stream processing.[7]

We also want to explore the possibility, however, that some actions can be directly shaped by visual informational that fixes experiential awareness as well. We need the conceptual tools to frame this hypothesis. Consequently, we will in general invoke a weaker notion of "direct" guidance of action, with the "non-consciously" dropped from the foregoing initial characterization. In this second sense we can ask whether the information that directly drives certain kinds of behavior is also reflected in experiential awareness.

Row 2 really alludes to a set of distinctions. First, Milner and Goodale (2006) invoke a familiar distinction between "allocentric" spatial coding (roughly, world-centered) and "egocentric" coding (roughly, subject or body-centered). According to Milner and Goodale, the former kind of spatial coding is associated with ventral processing and the latter with dorsal processing. (In what follows, the first spatial-coding term introduced is similarly associated with ventral processing by Milner and Goodale, and the second term with dorsal processing.) Milner and Goodale, however, also seem to associate this (allocentric vs. egocentric) coding distinction with a distinction between what they call "relative" and "absolute" coding of spatial attributes (Milner & Goodale, 2006, p. 239). These two distinctions are apparently associated as well by Milner and

[6] Actually, no doubt many brain areas are engaged here, including the frontal lobes in planning. Milner and Goodale do not deny this.

[7] Milner and Goodale (2008) suggest that the non-conscious action guiding process that they associate with the dorsal stream should be broken down further. Specifically, they hold that "the dorsal stream plays a central role in the programming of actions (i.e. the pre-specification of movement parameters), as well as their on-line control." We are not sure we entirely understand this. If the "motor programming" idea is endorsed to begin with, then presumably part of what it is for action to be under "on-line control" is that visual information sets and updates parameters of the muscle-signaling "motor programs." In any case, it seems adequate for our purposes to work with the distinction between direct and indirect control of action as broadly framed in the text. No doubt finer distinctions can be made and should be for some purposes (cf. Loomis & Beall, 1998, pp. 272-277)

Goodale with a distinction between coarse, "ball park" coding of spatial attributes and coding with "precision" (Milner & Goodale, 2006, pp. 90-91, 239). Furthermore, in one place there is an apparent implicit contrast drawn between (presumably) "non-coordinate" coding and what Milner and Goodale allude to as "coordinate" coding (Milner & Goodale, 2006, p. 91). As far as we understand these distinctions (and we are especially uncertain about the last), they seem different in the sense that the various distinctions need not line up. A charitable reading of Milner and Goodale would say they are making the following claims: i) As an empirical matter each side of the four row 2 distinctions are found together, and ii) the left and right sides of these row 2 pairs are also found to empirically align with the remaining properties distinguished in the two (ventral and dorsal) columns. We will not, however, discuss these (row 2) distinctions further in the current paper.

Finally, before proceeding, a word on the terms "personal space" and "action space" that we have used in framing our basic conjecture (roughly, that the Milner and Goodale scheme may well not apply to "action space" behavior, even though it may fit "personal space" behavior). These terms are drawn from Cutting and Vishton (1995). Cutting and Vishton (1995) remark that "the zone immediately surrounding the observer's head generally within arm's reach and slightly beyond, is quite personal." They "somewhat arbitrarily" limit such "personal space" to a distance of 2 meters (m). Of "action space" Cutting and Vishton (1995) remark, "In the circular region just beyond personal space is a space of an individual's public action…we move quickly within this space, we can talk (even teach) within it without too much difficulty, and if needs be we could toss something to a compatriot or throw a projectile at an object or animal." According to Cutting and Vishton (1995), this space stretches out to about 30 m (past which "vista space" is said to begin). These notions could no doubt be refined in theoretically useful ways (cf. DeVignemont, 2009), but they will do for current purposes. A chief observation of the current paper is that experiments have yet to be run to determine whether the Milner and Goodale scheme applies to actions with visually present targets situated in "action space."

Detection of Egocentric Distance and Mid-Range Behavior

Here we explore whether it is likely that the Milner and Goodale scheme applies to behavior guided by the visual detection of action space egocentric depth/distance. We begin with nuts and bolts, motivating and describing experiments that would decide the matter. More theoretical-methodological issues then arise naturally from a consideration of these proposed experiments.

A recent study, Bruggeman, Yonas, and Konzak (2007), suggests that it may well be conscious experience of egocentric distance/depth (or the information that fixes such experiences) that guides mid-range action, at least under some conditions. To appreciate the implications of the Bruggeman et al. (2007) results we first need to describe aspects of this study.

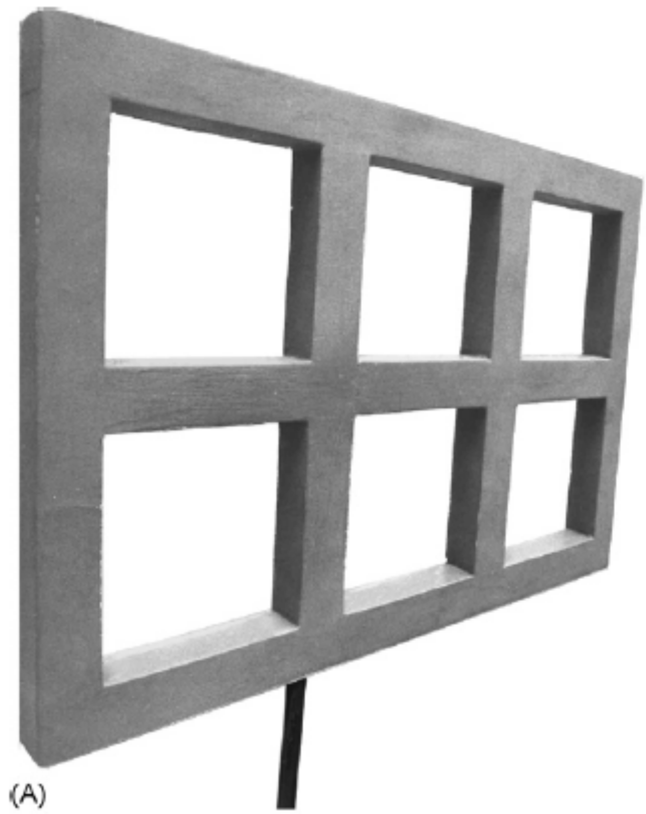


Figure 1. The Trapeziodal Ames window (see Bruggeman, Yonas, & Konzak, 2007). This appears to be a rectangular window rotated in depth, when in fact it is an unrotated trapezoid.

Bruggeman et al. (2007) used a physically constructed Ames trapezoidal window (ATW), shown above in figure 1. Participants viewed this ATW at less than half a meter. This is an object that appears to be a rectangle rotated in depth, but is in fact a trapezoidal form perpendicular to the line of sight. Bruggeman et al. had participants move their hands rapidly to the side of either the left or right edge of the ATW (without touching the object) and point to the position in depth of these outside edges. They compared performance (using a between subject design) under both monocular and binocular viewing. Using a kind of comparison task they also assessed the level of the perceptual illusion. That is, they assessed how rotated in depth the stimulus consciously appeared as a result of the (misleading) perspective information.

As gauged by the perception-comparison task, the illusion was powerful under both monocular and binocular viewing. The pointing behavior under monocular viewing closely tracked the conscious (illusory) perception of the depth of the outer edges. But under binocular viewing the pointing was instead closely linked to the actual positions of the left and right edges of the window. So under binocular viewing there was a dissociation between the conscious perception and the information driving the pointing. This was presumably because the pointing was guided by the effective (at the close viewing) stereo specification of the actual positions of the outside edges, which, of course, were, in fact, at the same depth/distance from the participants.

Bruggeman et al. (2007) interpret their results as in keeping with the Milner and Goodale two visual systems picture, and in an important respect they are. Stereo information has long been held to be especially important in guiding close-in, personal space reaching and grasping (cf. Cutting & Vishton, 1995; Servos & Goodale, 1994). In this experimental setting, it is apparently the stereo information, when available, that (accurately) drives the reaching and pointing and not the misleading perspective information (which still shapes conscious experience). This is indeed in keeping with the Milner and Goodale scheme.

But the Bruggeman et al. (2007) monocular results suggest that consciously perceived egocentric depth or spatial structure may well importantly drive mid-range, action-space behavior with similar stimuli. After all, in the monocular condition Bruggeman et al. found that behavior was driven by the consciously perceived egocentric depths/locations, the illusory depths/locations, of the object edges. And it is widely held that outside what we have called “personal space” stereo information for depth and spatial-structure is of increasingly diminished potency and usefulness (cf. McKee, Levi & Browne, 1990; Cutting & Vishton, 1995). In what the authors describe as a somewhat preliminary study, Allison, Gillam and Veceillo (2009) did find that adding stereo in already information rich settings can increase the accuracy of distance assessments out to (at least) 18 m. Nevertheless, the review and analysis of Cutting and Vishton (1995), in basic keeping with orthodoxy, quite strongly suggests that stereo information probably plays a rapidly diminishing and at best subsidiary role outside of personal space (at least in many spatial vision tasks). Bruggeman et al. found that it was the misleading perspective information that drove behavior when stereo information was not available. This suggests that the Milner and Goodale scheme may not apply to mid-range, action-space behavior.[8] This is at least a plausible hypothesis that should be carefully explored.

One possibility would be to run the deciding experiments in a free-walking, immersive VR environment. The general idea would be to show a simulated, rather large field ATW and then remove it. The task of participants would be to walk a distance corresponding to the distance that the left/right sides of the ATW were from them. The simulated ATW might be shown sitting on a pole joining its lower mid-point to a textured ground surface. Viewing should be in stereo (in at least one condition) and at roughly 8-10 m (simulated). We would expect that, first, the simulated ATW would still consciously appear rotated in depth as it did, with close viewing in the Bruggeman et al.

[8] Bruggeman et al. (2007) do find a reaction time advantage under stereo viewing. But while statistically significant, it is very small, on the order of 12 msec. Such a small difference does not suggest that the monocularly driven pointing behavior was somehow halting and initiated by some high-level, very different, inferential (or whatever) process.

study under both mono and binocular viewing. At this viewing distance, however, we would expect that the consciously perceived rotation in depth (specified by the perspective information) would now at least importantly drive the (action space) locomotion/walking and not the potentially (or 'in principle') disambiguating stereo/parallax information. If so, then Milner and Goodale would not be correct for an important class of mid-range visually guided action.

A possible drawback of using VR in this way is that it is well-known that estimated egocentric distance, as gauged by walking, is strongly under-estimated in VR environments (cf. Kunz, Wouters, Smith, Thompson, & Cream-Regehr, 2009). Nevertheless, for such a VR study, our prediction would be that the pattern in the results would reflect consciously perceived rotation/depth/distance. Thus, for the edge that consciously looked closer, walking distance would be shorter. If this is indeed the pattern observed, then this would still be informative. This said, it is no doubt advisable to also do the just described experiment with real objects in a non-VR environment. In this setting, participants might turn away from the (physical) ATW and attempt to walk a distance corresponding to the distance to the previously perceived left/right edges of the ATW.

In the preceding proposed experiments the behavior of participants would proceed after the target stimulus was withdrawn or was out of view. And it might be complained that with such "open loop" behavior we wouldn't know for sure if the information present initially (but then withdrawn) was really involved in the visual guidance of action in everyday settings.

We don't think that the reasoning here is decisive (see footnote 11 below). But we do agree that closed loop experiments should be run as well, both with real objects and using VR. Here the stimulus would remain in view and participants would walk to positions just outside the outer edges of the ATW. Under these conditions, we predict a more complex pattern in the action, but one with important, informative structure. Specifically, we conjecture that there will be differences in initial acceleration and in initial speed corresponding to the perceived rotation in depth of the ATW. Accordingly, when headed at the edge that is perceived to be further away, participants will, we conjecture, at least initially move/accelerate faster/greater than they would when headed at the edge that is perceived to be closer.[9] It is also possible, however, that stereo information, especially, will take over and control the approach to a stopping point as the ATW is neared.[10]

[9] It is worth noting that in the empirically based and tested Warren and Fajen model of the behavioral dynamics of human walking in a cluttered environment distance to obstacles/objects is an important parameter controlling walking acceleration (Warren & Fajen, 2004).

[10] And possibly also the structure information generated by the optic flow once subjects begin to move. With rotating Ames windows, however, the misleading perspective information dominates at least conscious experience. But, again, the experiments need to be done.

In sum, we think there is potential insight to be gained from both the open and closed loop experiments, with both real objects and using VR.[11]

We close this section with some step-back reflections on the logic of our proposed study and on the (at least implicit) logic of other related tests of the Milner and Goodale proposals about visual system structure and function.

In our proposals we have followed (largely implicit) custom in requiring that the following three conditions hold for there to be evidence against the Milner and Goodale scheme. First (1), there must be some information present that specifies the actual shape in space of the object, if a real (non simulated) object is used. Or with VR, there must be information that potentially disambiguates the stimulus, specifying (in the current case) an unrotated trapezoid. In the current case, this information could be stereo depth order information (for example) and/or perhaps optic flow structure information in the closed loop experiments. Second (2), there must be a clear perceptual illusion, at least of a sort. So the conscious percept must not track the relevant attributes of the actual object or (in VR) the relevant attributes specified by the disambiguating information. Finally (3), behavior must follow the illusory conscious appearances and not the shape in space of the actual object (if an actual object is used). Or, if in VR settings, then the behavior must follow the conscious percept of (here) a rotated rectangle and not the disambiguating information specifying an unrotated trapezoid.

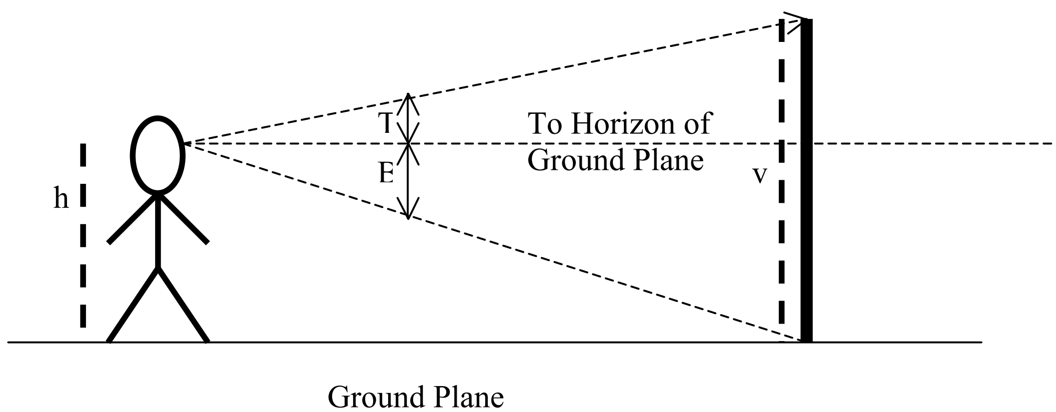
We do agree that if these conditions are met, then that would be evidence that the Milner and Goodale scheme does not fit the evidence for the case at hand. This is what we expect will happen with the "action space" depth/distance/spatial-structure stimuli and tasks we describe above.

But the interesting general, long term project may in the end simply be to determine just how (if affected at all) conscious experience varies with various kinds of sensed, action-driving visual information. In some settings there may be a tight coupling between experiential awareness and visual information that (say, specifically) directly drives mid-range action. There may also be no such coupling in other settings, possibly even with information controlling at least some actions aimed at targets in the mid-range (as we will discuss in section 4 below). We need to know these rules in order to more fully understand the explanation of visually guided behavior in general, and, more specifically, in order to understand the functional-behavioral role of conscious experience (or at least the role of the information that fixes or shapes conscious experience).

Detection of Body-Scaled Size and Mid-Range Behavior

[11] We do think that the open loop experiments should be in the mix. So, for example, while the reasoning is not simple, it has been argued in detail that it is the initially consciously perceived egocentric distance that drives blind walking responses (Loomis, Da Silva, Philbeck, & Fukusima, 1996).

Here is an illustration of horizon ratio geometry:



h = eye level (taken to be = 1)
v = height of object
T,B = visual angles
$v/h = (\tan(T) + \tan(B))/\tan(B)$. If v is small relative to its distance from the perceiver, then:
v/h = (approx.) $(T + B)/ B$ → v = (approx.) $h * (T + B)/ B$.
This is size in eye level units.

Figure 2. Horizon Ratio Geometry

Consistent with the hypothesis that such horizon ratio information is used, manipulations of false floors (and the like) affect assessments of objective/environmental size (Warren & Whang, 1987; Marks, 1987; Wraga, 1999a, 1999b; Rogers, 2003). It appears that in these cases size assessments do indeed work directly from horizon ratio information (see especially Warren & Whang, 1987; see also Sedgwick, 1986; Wraga, 1999a, 1999b). There is also overwhelming intuitive-phenomenological evidence that the detection of body-scaled size shapes conscious size experience. For example, people often remark that the rooms in which they spent early childhood years look much smaller when revisited later in life. Similarly, in the movie Fantastic Voyage, we readily accept that blood clots appeared enormous to the adventurers, who were miniaturized and sailing through blood vessels. (For more examples and extended discussion, see Bennett, 2010.) Horizon ratio information is not the only route to determining body scaled size (Bennett, 2010), but these observations do fit well with the range of horizon ratio experimental results cited above (once again, see Bennett, 2010, for more discussion).

It would surely be surprising if size-oriented mid-range behavior did not also closely vary with the way size consciously appears (in fine-grained ways that go beyond the selection of action aims and of types of actions). The matter needs to be investigated experimentally. This can be done with the following VR experiment. It is inspired by the Warren and Whang (1987) study, which, of course, did not use VR.

Participants would view a simulated door-like opening, situated in the mid-range, on a textured ground surface. The object might look like the frame of a doorway constructed out of two by fours. Viewing should be in simulated stereo in at least one condition. The task for participants would be to walk through the opening, without touching its sides, on the way to a target on a far wall that aligns with the optically specified horizon. The latter target requirement should keep participant heads roughly level to the ground surface and roughly stable (though running a condition allowing free head movement would also likely be a good idea). Textured side-walls would also be visible to enhance the (implicit) horizon information. The level of the ground surface would then be manipulated, as would where the optically specified implicit horizon cuts the door frame. This should probably be done with a between-subject design so participants would not get disoriented. In one condition, the ground surface would be lowered. Thus, the horizon would cut the frame-object at a higher point, and, as a result, the opening would correspondingly be detected/represented as smaller because a fewer number of eye-level units high and wide would be assessed. In another condition, the ground surface would be raised, and so the frame-object should be gauged to be larger. As in the previous horizon ratio studies cited above, assessments of how wide the opening visually looks should be gauged in a variety of ways (probably with a different group of participants).

Based on previous studies and on the compelling intuitive-phenomenological evidence (both discussed above) we expect that the floor-horizon manipulations will significantly affect how wide the opening looks. At the same time, recorded kinematic data will indicate speed of approach at different positions, how much participants bend their shoulders in relation to their head and heading angle, and when/where they begin to turn their shoulders.[12] We think it likely that participants will approach the openings that look smaller (with the lowering of the ground level) more slowly, anticipating the need to retain proper balance with a sharper upcoming shoulder turn. For openings that appear smaller participants may also begin their shoulder rotation earlier, and perhaps also bend their shoulders more (as it is plausible that participants will begin at least their preparation to so turn their shoulders outside of "personal space"[13]). In other words, we predict an important degree of coupling of the size information that shapes conscious size experience with the just described aspects of action as participants negotiate passable openings.

Our hypotheses here are in line with the interpretations of their data offered by Warren and Whang (1987). They conclude that horizon ratio information likely controls behavior in approaching and passing through openings (and we have noted that horizon ratio information also very likely shapes conscious size awareness). Warren and Whang

---

[12] Subjects would be tracked using a sonic/inertial head/body tracking system; see appendix I below for details.

[13] We are not aware of data bearing specifically on the conjecture here. But evidence from VR studies by Fajen and Warren on obstacle avoidance (cf. Fajen & Warren, 2003) does suggest that at least some overt effects begin at three meters or more on obstacle avoidance routes taken. This may also reflect the results of visual intake that begins even farther out from the obstacle.

could not, of course, without VR, have had participants pass through perceived openings while also manipulating the horizon ratio optical information.

It is possible to shoehorn this case into the customary ("three condition") form described at the end of the previous section. In principle, the stereo geometry does in some sense specify the presence of a door frame of the same size across the ground-level/horizon shifts. Another source of size information, available across conditions, is in the optic flow field, provided that the participant moves towards the opening at a constant known speed.[14] It is possible that the stereo-derived size information (and/or the optic-flow-derived size information) will control approach deceleration/speed (and/or shoulder-bending behavior), while the experience of size is dominated by the size information yielded by the changing horizon ratios (as ground level is manipulated).[15] If so, then Milner and Goodale would be right about this kind of behavior. But given the salience and the likely use of horizon-ratio size information, we join Warren and Whang (1987) in expecting that horizon ratio information dominates the control of behavior in approaching and passing though openings. At the least it is a plausible live possibility that the Milner and Goodale scheme does not apply to the behavior described. The matter should be carefully checked experimentally.

D.F. and Mid-Range Navigation-Locomotion

Goodale and Milner (2004, pp. 18-19, 27, 117) offer informal observations that seem to indicate that their visual form agnosic patient D.F. has intact mid-range navigational abilities. Goodale and Milner (2004, p. 28) also allude to a single small study of D.F.'s stepping procedure (the study is Patla & Goodale, 1996). Both considerations may suggest that the Milner and Goodale scheme applies as well to mid-range, action space, visually guided behavior. We have significant doubts. We first explain why one must be careful in drawing conclusions from informal observations of the sort provided.[16] We then critically assess the stepping study involving D.F.

The anecdotal reports of D. F. successfully navigating rich natural environments should be assessed cautiously. Similar to honeybees (Dyer, 1991), but unlike desert ants (Wolf & Wehner, 2000), there is good evidence that humans prefer to rely on simple landmark image-matching to navigate (Foo, Warren, Duchon, & Tarr, 2005). This is so even after a single trial in a novel environment (Foo, Duchon, Warren, & Tarr, 2007). It

[14] Suppose the instantaneous distance is x, and visual angle alpha. Then the width of the doorway w is approximately: 1.) w = x alpha. Next, take the derivative with respect to time t, on both sides:
2.) 0 = x d(alpha)/dt + alpha dx/dt. Here dx/dt is the constant speed of viewpoint translation (assumed known), d(alpha)/dt is the rate of the angular change, which is measurable, and alpha is also measurable. Therefore, the x can be solved in 2), which is then substituted in 1) to get w. Here for simplicity the small angle assumption is used, tan(alpha) = alpha. But this is not essential. (Thanks to Zili Liu for this analysis.)

[15] In fact, we do not expect the optic flow size information to affect experienced size: Warren and Whang (1987) found no difference in size assessments of an opening with moving as opposed to stationary subjects.

[16] There is some suggestion that Goodale and Milner may not need convincing here (Goodale & Milner, 2004, p. 27). The prominence Goodale and Milner (2004) give the anecdotal reports, however, leaves this somewhat unclear. Either way, our remarks will be a caution to the unwary.

is reported that D.F. has at least some spared perception of color and texture properties (Goodale & Milner, 2004). This could support encoding and matching of "images" of color-texture-defined landmarks. Only carefully controlled experiments can disentangle just what D.F. can do here, and how she does it. These studies have not been done. From all we know right now, D.F. may be quite impaired in negotiating genuinely novel and/or complex environments.[17]

In the Patla and Goodale (1996) study, both D.F. and normal subjects stepped over isolated risers of varying height and location. The risers were situated on a flat, open ground-surface. The trajectory of D.F.'s toe, in stepping over these obstacles, was found to be in line with the toe trajectories of five normals. Nonetheless, D.F.'s conscious perception of the height of the risers was poor, much different from that of normals. No direct data, however, is reported about the speed at which D.F. does the approaching and stepping. Patla and Goodale do say that D.F.'s "ability to deal with obstacles" is "relatively" intact. They don't explain why they implicitly qualify their claim by adding the "relatively." In a recent public talk Goodale did qualify some of his remarks about a new, similar patient, by briefly mentioning that this patient needed more time to accomplish at least some visually guided tasks. The timing data is important to have. It can help in determining which inferences are correct to draw about normals from the behavior of patient with extensive brain damage.

This said, we don't doubt that D.F. can do the stepping task used in Patla and Goodale (1996) at least serviceably well. She might rely on structure information in the optic flow to detect the presence of a riser; in so far as we can tell from Milner and Goodale (2006), MT and surrounding (motion processing) cortical areas appear to be at least importantly spared in D.F. She might even rely on her partially preserved ability to visually determine surface-quality, color-texture (Goodale & Milner, 2004) to tell that there is "something ahead" that stands out (especially after getting initially acquainted with the experimental set up). Either way, perhaps she then, at least in part, uses stereo information to help guide stepping trajectory. Normals may even perform the described stepping task in something broadly like this way, at least in similar (rather simple and uncluttered) settings. All these are open questions requiring more study. We are not committed to the idea that all action with mid-range aims is directly guided by information that is reflected in conscious experience (wholly or in important part).[18]

---

[17] There are further grounds for caution: In her book on the agnosias Farah (2004) writes that visual form agnosics are "unable to negotiate visual environments of any complexity" (p. 41).

Goodale's and Milner's most striking anecdote has D.F. joining in a picnic trek up a mountain path (Goodale & Milner, 2004, pp. 18-19), but they later appear to concede that D.F. may have had previous familiarity with this setting (Goodale & Milner, 2004, p. 27).

[18] Patla, Niechiej, Racco, and Goodale (2002) is a stepping study done entirely with normals that the authors take to support the conclusions drawn from the earlier stepping study involving D.F. Though we just conceded that the Milner and Goodale general picture may apply to stepping behavior, at least in some settings, we are not convinced by the interpretative arguments in this second paper. We provide details in Bennett and Foo (2009).

Concluding Remarks

We have seen that it is an unappreciated open question whether action guided by information specifying mid-range action aims or targets fits the Milner and Goodale scheme. And we have described experiments to test whether this is so for important kinds of mid-range, action space behavior. It is an important supplementary part of our methodological recommendations that the results from the study of one kind of visually guided action may not generalize.

But why might at least much of the visual information that directly guides reaching and grasping not be reflected in conscious experience, while key information that guides and shapes important mid-range action is so reflected (if we are right in our conjectures)? We don't know. We could say more, drawing from recent reflections in a burgeoning literature on the possible nature and role of varieties of conscious awareness. We think the matter might, maybe, turn on the utility of the accessibility to decision and control processes of certain kinds of visual information. That would be another paper, and an inconclusive one at present. We do think that the query here may implicitly and importantly get cart first, while we point to what should be a pulling horse. Answers to the empirical questions we pose about action space behavior will provide important constraints on theorizing about the nature and role of experiential awareness (or the information that shapes such awareness). No one at present has decisive answers to these empirical questions concerning mid-range action. Nonetheless, we have provided empirically based reasons to think that the Milner and Goodale scheme may well not apply to much mid-range, action space behavior. We have described experiments that would decide the matter.

Appendix I. On Virtual Reality (VR)

With the advent in the 1990s of improved display technologies, lightweight movement tracking equipment, and enhanced computer graphics, researchers gained enhanced means to directly test hypotheses about the visual control of human behavior (Tarr & Warren, 2002). In this section, we will review several types of VR now available.

As a starting point, any participant who is able to interact with a video game or a computer system can achieve "desktop VR." A key factor is that the user is included in the control loop in real time via an input device such as a keyboard, joystick, or steering wheel. Here an important factor in deepening the felt immersion is the field of view (FOV). Simply put, bigger FOV is better. This is in keeping with experiences in IMAX theatres, for example.

One way researchers have moved beyond desktop VR is by expanding the viewing area into small (~ 3m x 3m) display rooms like the CAVE^{TM}. Here the participant wears stereoscopic goggles and can navigate the depicted scenes via devices like a tracked glove. In these set-ups, there is compelling immersion in the three dimensional scenes filling the panels making up the cube (typically at least the four walls), especially by comparison to desktop VR. For the types of experiments concerning mid-range action space behavior that we discuss in the current paper, however, larger "free-walking" "immersive" VR is necessary (Tarr & Warren, 2002).

Such free walking, immersive VR laboratories furnish a controlled testing environment that possess even greater perceptual realism. For example, the Virtual Environment and Navigation laboratory (VENLab) at Brown University allows participants to walk in an immersive virtual environment (12m x 12m). Participants do this while wearing a low-latency (50-70ms), wide FOV stereo head-mounted display (60° x 40°) with a sonic/inertial head/body tracking system. Computers generate the appropriate scene given the tracking data. These trackers can be placed on a participant's head, or they might be so placed as to gauge turning of the shoulders (as in our proposed experiment modeled on Warren & Whang, 1987). Audio white-noise and the ability to navigate the virtual environment, simply by walking or even running (see Fink, Foo, & Warren, 2010), help to enhance the illusion. Participants report high levels of immersion in such settings. Comparing walking and obstacle avoidance in such VR set-ups with walking and obstacle avoidance with real targets confirms the efficacy of this technology (Fink, Foo, & Warren, 2007).

VR laboratories have proven useful in vision science by creating convincing testing environments that allow for the careful manipulation of visual variables in ways that were previously impossible to do. For example, optic flow can be dissociated from egocentric direction of travel (cf. Warren et al., 2001). A recent study even changed the gravity-associated acceleration of ballistic objects so that baseballs do not travel in natural parabolas (Fink, Foo, & Warren 2009). One need be limited only by one's imagination.

Appendix II. The View From Neuroscience

We close with some brief remarks concerning the first row (in Table 1 above) in our portrayal of Milner and Goodale's two visual systems hypothesis. An important part of the Milner and Goodale view is that the different psychological-coding distinctions observed in rows 2-4 of Table 1 correspond to anatomical differences in the brain.

We think our hypotheses are compatible with the results of contemporary neuroscience. We lack the space to review the many pieces of evidence from visual neuroscience that Milner and Goodale (1995, 2006, 2008; see also Goodale & Milner, 2004) deploy in developing and defending their two visual systems picture of neural-functional architecture. As we have noted, we think that aspects of their account are compelling. Milner and Goodale certainly also do not deny that there are important interactions between their posited ventral and dorsal streams (cf. the epilogue in Milner & Goodale, 2006). They may still, however, under-estimate how intertwined the processing in the ventral and dorsal "pathways" is.

In detailing one important route that visual information is held to travel, Milner and Goodale (1995, 2006, 2008; see also Goodale & Milner, 2004) describe how visual information first passes from the retina to the lateral geniculate nucleus of the thalamus.

Visual information is then described as traveling through primary visual cortex (V1) before there is a splitting of neural processing streams. Milner and Goodale appear to display some hesitation about where to group mid-level cortical regions like MT and V3A (Milner & Goodale, 2006, p. 204). But the broad picture is that visual processing splits into a ventral and a dorsal stream at least shortly after primary visual cortex (Milner & Goodale, 1995, 2006, 2008, Goodale & Milner, 2004). The ventral stream runs through temporal cortex, while the dorsal stream leads through parietal cortex. D.F.'s damage is complex and somewhat diffuse, but it is concentrated in neural structures associated with parts of this, hypothesized, ventral stream. Milner and Goodale describe other patients ("optic ataxics") whose damage is concentrated in neural structures associated with parts of the hypothesized dorsal stream. These patients appear to show an opposite pattern of deficits (very roughly put, they showed impaired action, but intact conscious experience).

How clean is the distinction between these two, hypothesized, streams of visual processing? We cautiously suspect that the answer is, "not very." We also think, however, that it is unclear at present just what is correct to say.

Present-day knowledge about human visual neuroanatomy results from converging evidence from a number of sources and approaches. Traditionally, one had to wait and perform post-mortem analysis to correlate brain lesions with documented behavioral deficits. Prospects have brightened, however, with the advent/refinement of non-invasive techniques like PET, EEG, MRI and fMRI. So, techniques like fMRI allow for identification of damaged brain areas in living patients like D.F (cf. the epilogue in Milner & Goodale, 2006). Such techniques also provide the basis for inferences about the neural activity present when doing certain targeted experimental tasks with immobile normal participants. As often noted, though, inferences using these techniques are indirect, relying on multiple assumptions about the relation between the measured activity and neural processing. Bartels, Logothetis, and Moutoussis (2008), for example, present a balanced and nuanced discussion of both the significant limitations of the use of fMRI in the study of the human visual system, as well as its great promise. They argue that the fMRI derived BOLD signal may be a coarse and uncertain indicator of spiking activity. Nevertheless, they maintain that this signal might still reflect other physiological activity that importantly shapes visual perception.

In another main traditional method, researchers have long placed microelectrodes deep into the brains of non-human animals to record from individual neurons (Bartels et al., 2008). Macaque monkeys are the most intensively studied animal model for the human visual system. Emerging comparisons between human and macaque visual systems (e.g., using fMRI), however, have revealed that the relations are complex and, in many respects, presently not well understood (Orban, Van Essen, & Vanduffel, 2004; Todd, 2004).

Recent reviews of visual neuroanatomy—working from results obtained using the foregoing and other methods—suggest that the flow of information past V1 may eventually dissociate into two, identifiable, ventral and dorsal processing streams. But these reviews also reveal that the intervening organization is complicated (e.g. areas V2-V4, MT, MSTd, etc). There are dense interconnections, feedback and feedforward connections, seemingly redundant functions, and species-specific details. All this complicates inferences. (For helpful reviews, see Orban, Van Essen, & Vanduffel, 2004; Ungerleider & Pasternak, 2004; VanEssen, 2004).

In sum, this seems to us a time for both excitement and for caution in theorizing about the functional architecture of the visual brain. These are exciting times, surely, with the development of promising new tools for researchers to use and to refine. But these are also early days still in understanding our most complex organ. We expect continued fruitful two-way interactions across disciplines in understanding the functional architecture of the human visual system and in understanding underlying computational mechanisms. We see nothing in what is currently known from visual neuroscience (or elsewhere) about visual system architecture that rules out our hypotheses. We expect that the results of the experiments we propose in the current paper will place important constraints on future theorizing in the visual neurosciences.

Acknowledgments

Thanks to Zili Liu for the help noted in footnote 14. Thanks to Quoc Vuong for an article reference. Thanks to Chris Hill for discussion and for general encouragement.

References

Bartels, A., Logothetis, N.K., & Moutoussis K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends in Neuroscience*, *31*, 444-453.

Bennett, D.J. (2010). How we measure up the world in size experience. *Philosophy and Phenomenological Research* (forthcoming).

Bennett, D.J. & Foo, P.S. (2009). A note on the Patla and Goodale stepping studies. Unpublished manuscript.

Bruggeman, H. Yonas, A., & Konczak J. (2007). The processing of linear perspective and binocular information for action and perception. *Neuropsychologia*, *45,* 1420-1426.

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

Cutting, J. & Vishton P.M. (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, CA: Academic Press.

DeVignemont, F (2009). The coming out of bodily sensations. (ms., draft chapter of a book under preparation.)

Dyer, F. C. (1991). Bees acquire route-based memories but not cognitive maps in a familiar landscape. *Animal Behaviour, 41,* 239–246.

Fajen, B. & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 343-362.

Farah, M. (2004). *Visual agnosia* (2nd ed.). Cambridge, MA: MIT Press.

Fink, P. W., Foo, P. S., & Warren, W. H. (2009). Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of Vision*, *9*(13), 1–8. http://journalofvision.org/9/13/14/, doi:10.1167/9.13.14.

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

Kunz, B.R., Wouters, L., Smith, D., Thompson, W.B., & Creem-Regehr, S.H. (2009). Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics*, *71*, 1284-1293.

Loomis, J.M., Da Silva, J.A, Philbeck, J.W. & Fukusima, S.S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science*, *5,* 72-77.

Loomis, J. M. & Beall, A. C. (1998). Visually-controlled locomotion: Its dependence on optic flow, 3-D space perception, and cognition. *Ecological Psychology*, *10*, 271-285.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30*, 1763–1779.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (TOG), 21*, 645-652.

Milner, A.D. & Goodale, M.A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Milner, A.D. & Goodale, M.A. (2006). *The visual brain in action* (2nd edition with new epilogue). Oxford: Oxford University Press.

Milner, A. D. & Goodale, M.A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46*, 774-785.

Orban, G.A., Van Essen, D., and Vanduffel, W. (2004) Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Science*, *8*, 315-324

Patla, A.E. & Goodale, M.A. (1996). Obstacle avoidance during locomotion is unaffected in a patient with visual form agnosia. *NeuroReport, 8*, 165-168.

Patla, A. E., Niechwiej, E., Racco, V., & Goodale, M.A. (2002). Understanding the contribution of binocular vision to the control of adaptive locomotion. *Experimental Brain Research*, *142*, 551-561.

Philbeck, J.W. & Loomis, J.M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 72-85.

Rogers, S.J. (2003). Truth and meaning in pictorial space. In H. Hecht, R. Schwartz, and M Atherton (Eds.), *Looking in to pictures*. Cambridge, MA: MIT Press.

Sedgwick H A (1986) Space perception. In K.L. Boff, L. Kaufman, and J.P. Thomas (Eds.) *Handbook of perception and human performance, Vol. 1: Sensory processes and perception.* New York: Wiley.

Servos P. & Goodale M A (1994). Binocular vision and the on-line control of human prehension. *Experimental Brain Research*, 98, 119-127.

Tarr, M. J. & Warren, W. H. (2002) Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience*, 5, 1089-1095.

Todd, J.T. (2004). The visual perception of 3D shape. *Trends in Cognitive Science*, 8, 115-121.

Ungerleider, L.G. & Pasternak, T. (2004) Ventral and dorsal cortical processing streams. In L. M. Chalupa and J.S. Werner (Eds.), *The visual neurosciences* (Vol 1). Cambridge, MA: MIT Press.

VanEssen, D.C. (2004). Organization of visual areas in macaque and human cerebral cortex. In L.M. Chalupa and J.S. Werner (Eds.), *The visual neurosciences*. Cambridge, MA: MIT Press.

Warren, W.H. & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 371-183.

Warren, W.H. & Fajen B.R. (2004). Behavioral dynamics of human locomotion. *Ecological Psychology*, *16*, 61-66.

Wolf, H., & Wehner, R. (2000). Pinpointing food sources: Olfactory and anemotactic orientation in desert ants, *Cataglyphis fortis*. *Journal of Experimental Biology, 203*, 857–868.

Wraga, M. (1999a). Using eye height in different postures to scale the heights of objects. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 518-530.

Wraga, M. (1999b). The role of eye-height in perceiving affordances and object dimensions. *Perception & Psychophysics*, *61*, 490-507.