The Two-Visual-Systems Hypothesis and the Perspectival Features of Visual Experience

Robert T. Foley¹,²,³, Robert L. Whitwell⁴, and Melvyn A. Goodale³,⁵,⁶

¹ The Rotman Institute of Philosophy, The University of Western Ontario.
² The Department of Philosophy, The University of Western Ontario.
³ The Brain and Mind Institute, The University of Western Ontario.
⁴ The Department of Psychology, The University of British Columbia.
⁵ The Department of Psychology, The University of Western Ontario.
⁶ The Department of Physiology and Pharmacology, The University of Western Ontario.

Keywords: The two-visual-systems hypothesis; ventral and dorsal processing; egocentric and allocentric frames of reference; perspectival accounts of visual experience.

This is a final draft, please quote from print version.

Corresponding author:

Robert Foley,
The University of Western Ontario,
Stevenson Hall, 2157,
London, Ontario, Canada N6A 5B8
Email: rfoley5@uwo.ca
Abstract

Some critics of the two-visual-systems hypothesis (TVSH) argue that it is incompatible with the fundamentally egocentric nature of visual experience (what we call the ‘perspectival account’). The TVSH proposes that the ventral stream, which delivers up our visual experience of the world, works in an allocentric frame of reference, whereas the dorsal stream, which mediates the visual control of action, uses egocentric frames of reference. Given that the TVSH is also committed to the claim that dorsal-stream processing does not contribute to the contents of visual experience, it has been argued that the TVSH cannot account for the egocentric features of our visual experience. This argument, however, rests on a misunderstanding about how the operations mediating action and the operations mediating perception are specified in the TVSH. In this article, we emphasize the importance of the ‘outputs’ of the two-systems to the specification of their respective operations. We argue that once this point is appreciated, it becomes evident that the TVSH is entirely compatible with a perspectival account of visual experience.
1. Introduction

An influential account of the functional organization of the human visual system has argued that the visual operations mediating visual perception are distinct from those mediating the control of action (Goodale & Milner, 1992). This two-visual-systems hypothesis (TVSH) proposes that the ventral stream of visual pathways in the cerebral cortex, which delivers up our visual experience of the world, works in an allocentric frame of reference, whereas the dorsal stream, which mediates the visual control of action, uses egocentric frames of reference. A central claim of the TVSH is that dorsal-stream processing for action does not contribute to the contents of visual experience and is largely isolated from ventral-stream processing.

This provocative account of how vision works has captured the attention of a number of philosophers interested in how we represent the visual world. Much of the philosophical discussion has focused on whether or not the contents of visual experience contribute to the control of action (See for example Briscoe & Schwenkler, n.d.; Campbell, 2002; Clark, 2010; Clark, 2001; Kozuch, n.d.; Mole, 2009; Shepherd, 2015; Wallhagen, 2007; Wu, 2013). But there is another equally contentious issue about whether or not Goodale and Milner’s conception of the TVSH can accommodate the fundamentally egocentric nature of visual experience (what we call the ‘perspectival account’). It is this issue that we focus on here. We argue that the contradiction between the TVSH and perspectival accounts are more apparent than real, arising from a basic misunderstanding of how the operations mediating perception and the operations mediating action are specified in the TVSH.
Central to this misunderstanding is the failure to appreciate that the operations are specified in terms of the outputs of the 'two systems'. Once this point is clarified it becomes evident that the TVSH is entirely compatible with a perspectival account of visual experience.

2. The Perspectival Account

According to perspectival accounts of visual experience, objects are represented in egocentric space. Discussions about the egocentric features of visual experience can focus on different aspects of visual experience. For the purpose of this article, we concern ourselves with two features of visual experience that motivate the perspectival account. Both of which arise from reflections on the visual array and how objects in the visual array, and their relationships to us, change as we shift our perspective.

One source of motivation for the perspectival account is grounded in the (immediately evident and intuitively undeniable) fact that we perceive and can readily consider a spatial relationship between any given object in the visual array.

---

1 There are many different positions that develop the claim that there are egocentric components to our visual experience. These positions involve complex issues about the necessity of egocentric processing for visual experience and develop the claim in multiple different ways. Philosophers have argued: that action is a necessary condition of our visual experience being structured the way it is (cf., Burge, 2010; Evans, 1982; Grush, 2007; Merleau-Ponty, 2012; Peacocke, 1992; Strawson, 1959 for very different versions of this approach); that the contents of our visual experience are constituted out of sensorimotor contingencies (e.g., O’Regan & Noë, 2001); and that the location of the perceiver is represented in visual experience (e.g., Cassam, 1994; Husserl, 1990). These complexities do not concern us here. Instead, this paper is concerned with an apparent clash between perspectival accounts in general and the TVSH. The simple sketch set out in this section should suffice to motivate the topic of this paper.
and ourselves. This variety of perspectival account simply amounts to an acknowledgement that the observer’s perspective defines which objects in the visual array are to their left or right, which are up or down, and which is nearer or further away. A second source of motivation for the perspectival account is grounded in the (perhaps somewhat less evident) fact that we can perceive a feature (e.g., shape) of a given object in the visual array as being primarily a function of its projection on the retina. The projection of the object on the retina, of course, depends on the viewpoint of the observer. An example or two will hopefully make this point clearer: although we see a ring as being circular irrespective of our viewpoint, we can also perceive it as ovoid when we perceive it from a particular angle. By the same token, looking at your office door while standing to its side readily yields the perception of a normal-looking rectangular-shaped door.

Nevertheless, this viewpoint can also yield the impression (with some effort) that you are looking at a trapezoidal-shaped door. The fact that the ring can appear to be ovoid and the door can appear trapezoidal reflects their particular projection on the retina.² The point here being that the relation between the observer and the object partially determines how the object is represented in visual experience.

It is clear that these intuitive and obvious perspectival features of visual experience (location and shape) depend on the egocentric relation between the observer and

² It may be the case that the ability to perceive both the projected shape of objects on the retina and their ‘real shape’ is largely restricted to objects with relatively simple geometric shapes, such as rings and doors. It should be noted, however, that the ability to perceive the projected shape of objects may depend on expertise, and that training means that visual artists, such as painters, can easily adopt such viewpoints when making realistic paintings or drawing of a scene. In addition, it may be the case that different instructions can lead subjects to make different judgements about features such as size (see (Gilinsky, 1955).
the environment. Indeed, it seems undeniable that there are such perspectival features of our visual experience. Some of the proponents of the perspectival account have argued, however, that the two visual systems hypothesis (TVSH) as proposed by Goodale and Milner (1992) is at odds with the perspectival account (see for example, Briscoe & Schwenkler, n.d.; Briscoe, 2009; Brogaard, 2012). According to this ‘perspectival challenge’, the TVSH, which would appear to attribute egocentric processing entirely to the dorsal stream, cannot accommodate the egocentric nature of visual experience, and as a result, should be rejected. In order to understand the perspectival challenge, and where it goes wrong, we must first look in a bit more detail at how the concepts of egocentric and allocentric frames of references have been characterized.

### 3. Different Frames of Reference for Perception and Action

The TVSH describes a broad functional and anatomical division in the primate visual system. According to the TVSH, the ventral and dorsal visual streams that arise from early visual areas in the primate cerebral cortex serve different functions. The ventral stream, which projects to the temporal lobe, processes visual information to inform perceptual and cognitive tasks such as memory, recognition, identification, visual imagery and categorization (‘vision for perception’). The dorsal stream, which projects to the parietal lobe, mediates the visual control of skilled goal-directed action (‘vision for action’) (Goodale & Milner, 1992). A fundamental insight of the TVSH is that perceiving the environment and acting on it each present unique
problems for the brain to solve. The successful execution of their separate functions demands that the ventral and dorsal streams utilize radically different kinds of processing.

One way of describing the different kinds of processing that are carried out in the ventral and dorsal streams is in terms of two different frames of reference (Milner & Goodale, 2006). A frame of reference is a way of representing the locations of entities in space. The dorsal stream is often described as processing visual information in an egocentric (observer-based) frame of reference, whereas the ventral stream is described as processing visual information in an allocentric (scene-based) frame of reference. One has to be careful here, however, because the terms egocentric and allocentric are not precisely defined. Moreover, their use varies across different literatures. In the spatial cognition literature, for example, an egocentric frame of reference is one which specifies the locations of objects in a scene with respect to the viewpoint of the observer, whereas an allocentric frame of reference is one in which the locations of objects are represented relative to other objects in the scene independently of the viewpoint of the observer (see Klatzky, 1998). Importantly, these definitions are used in the context of spatial navigation and memory tasks in which the person moves from a current location to another in their environment.

---

3 This idea resonates to some degree with J.J. Gibson’s (1979) notion of “direct perception.” It is important to note, however, that Gibson did not make a distinction between the visual processing mediating perception and that underlying the control of actions. For a discussion of this issue, see Goodale and Humphrey (1998)
In the TVSH literature, on the other hand, the terms egocentric and allocentric are largely applied to the visual processing of the current visual array. In this case, the concern is with how people make judgements about the locations and features of objects in the visual array or how they interact with particular objects in that array. When mapped onto the functional profiles of the ventral and dorsal streams, the use of the terms allocentric and egocentric has not always been as precise as it could have been. In fact, it was never Goodale and Milner’s intention to exclude any and all egocentric-based operations from the ventral stream. Rather, in discussing ventral-stream operations in terms of allocentric (or scene-based) frames of reference, they sought to highlight the importance of scene-based processing to visual perception. Likewise, in discussing dorsal-stream operations in terms of egocentric (or effector-centred) frames of reference, Goodale and Milner simply sought to highlight the importance of spatial relationship between the goal object and the relevant effector (points, which we will expand on below). This ambiguity in how the terms allocentric and egocentric have been applied may have misled certain critics when formulating the perspectival challenge to the TVSH.

4. The Perspectival Challenge

As we have noted, the TVSH holds that there is a functional distinction between the visual operations mediating visual perception and those mediating the control of action. Importantly, the vision-for-perception and vision-for-action streams have been characterised as operating on visual information within egocentric and
allocentric frames of reference respectively. Furthermore, the TVSH is committed to the claim that the operations of the dorsal stream do not contribute to the contents of visual experience. According to proponents of the perspectival challenge, these two claims mean that the TVSH is incompatible with the perspectival account of visual experience. To some, this supposed consequence of the TVSH is patently absurd. Briscoe (2009), for example, argues that we could not even imagine what it would be like for our visual experience to be purely in an allocentric frame of reference:

What would it be, one might ask, to have the abilities needed perceptually to identify the locations of objects relative to other objects in one’s field of view—e.g., the location of one candlestick relative to another on the dining table—but not those needed ever to identify the location of objects relative to one’s own body? Is such non-perspectival, i.e., purely allocentric, experience of space possible? (446)

According to this criticism, the TVSH commits us to an account of visual experience in which the contents of the observer’s visual experience represents no perspectival features of the scene at all. If this were the case then there would be no right or left, no up or down, no near or far since such spatial relations are defined relative to the perspective of the observer. Somewhat like Kant’s hapless astronomer (Kant, 1992), who has a perfect map of the stars but no way of orienting it relative to their own position, our visual experience of the world would somehow be situated as a view from nowhere. This version of the perspectival challenge highlights the notion of viewpoint independence that we saw in the definition of allocentric processing given above and argues that we cannot even imagine what it would be like for our visual experience to be genuinely viewpoint independent.
Brogaard (2012), on the other hand, argues that the TVSH cannot account for the representation of specific features of objects in visual experience (such as the ovoid shape of the ring). On this account, since the ventral stream processes visual information in an allocentric frame of reference, it cannot deliver up the egocentric representation of perspectival properties of the object (i.e., a representation of shape derived from the retinal projection of the object). As a result, ventral stream processing cannot account for the perspectival representation of intrinsic properties of objects (such as shape) in visual experience.

Both of the above versions of the perspectival challenge argue that introspection on the nature of visual experience warrants the claim that there are features of visual experience that are incompatible with the TVSH. We can rationally reconstruct the perspectival challenge as relying on the following argument.

The TVSH is committed to the following claims:

A1: The dorsal stream only processes visual information in an egocentric frame of reference.
A2: The ventral stream only processes visual information in an allocentric frame of reference.
A3: The contents of ventral stream processing contribute to the contents of visual experience.
A4: The contents of dorsal stream processing do not contribute to the contents of visual experience.

However, according to the perspectival account:

A5: We sometimes experience objects and scenes relative to our perspective (i.e., in an egocentric frame of reference).

Therefore:

C: Either A5 is false or at least one of A2 and A4 is false.
Proponents of the perspectival challenge have concluded that because A5 is so intuitively obvious: either A2 is false and the ventral stream processes visual information in an egocentric frame of reference (Briscoe, 2009; Briscoe and Schwenkler, forthcoming); or A4 is false and the dorsal stream contributes to the contents of visual experience (Brogaard, 2012). Either way, the TVSH is said to be undermined by the perspectival challenge.

Of course it is always an open possibility that empirical research could undermine an intuitive claim such as A5. Nevertheless, in what follows we assume that there are aspects of our visual experience that are egocentrically structured.\(^4\) Granting this claim, we argue that the features of visual experience the perspectival account is concerned with are perfectly compatible with the TVSH. Thus the perspectival challenge does not undermine the TVSH. Before addressing this central issue however, we consider a brief clarificatory point.

5. 'Contributions' to Visual Experience

It is important to be clear on the sense in which 'contribute' is being used when we ask whether the dorsal stream contributes to the contents of visual experience.

\(^4\) It may be the case that perspectival features of visual experience can be realised in purely allocentric frames of reference. On this account the subject is represented as another object in a scene-based frame of reference. Other’s have argued that such perspectival representation requires “genuinely egocentric” processing (See Campbell, 2002; Jacob & Jeannerod, 2007; Jacob & Vignemont, 2010; Schwenkler, 2014; and Wu, 2014 for some discussion of these issues in the context of the TVSH). In what follows we do not take a position on this issue, and instead simply acknowledge that, whichever account may be right, there are perspectival features of visual experience.
There is an obvious sense in which the dorsal stream does contribute to the contents of visual experience. As Goodale and Milner (2006) have pointed out, visuomotor mechanisms in the dorsal stream mediate shifts of gaze and/or covert attention from one location in the visual scene to another. Presumably this dorsal-stream area must interact with ventral-stream perceptual processes when shifts of gaze and attention are executed. Such mechanisms are integral to the realisation of visual experience.

Our concern here, however, is not with whether dorsal stream processing is necessary for visual experience; rather it is whether some specific contents of visual experience critically depend on the contents of dorsal-stream operations. To use the example discussed earlier: Does dorsal processing of a ring in an egocentric frame of reference contribute to the observer’s conscious visual experience of that ring? The short answer is no: The TVSH is committed to the claim that dorsally mediated coding of objects for visually guided action does not contribute to how those objects or scenes are represented in conscious visual experience. Nonetheless, the processing of object location in the dorsal stream that is required for shifting gaze or attention to that object is necessary for the normal unfolding of visual experience.

We now turn to the main focus of this article – what is meant by the claim that the dorsal and ventral streams process visual information in egocentric and allocentric frames of reference respectively.
6. Allocentric and Egocentric Processing in the TVSH

As we saw above, the perspectival challenge claims that TVSH’s characterisation of dorsal and ventral processing as involving egocentric and allocentric processing respectively means that it cannot accommodate the egocentric features of visual experience. But the tension here is only apparent. According to the TVSH, recognizing an object and acting on that object require very different kinds of processing of the visual array. A central insight of the TVSH is that one cannot properly define the functional role of the two streams solely in terms of their inputs. Instead the TVSH conceives of the functional role of the two streams as being largely defined in terms of their outputs to other regions of the brain, what we might call the ‘consumers’ of those outputs, and the tasks those consumers serve.

As Milner and Goodale (2006) put it “...our perspective on the division of labour between the dorsal and ventral streams is to place less emphasis on input distinctions (for example, object location versus intrinsic object qualities) and to take more account of the output characteristics of the two cortical systems.” (40) This was a significant departure from previous accounts of the division between the two streams, such as Ungerleider and Mishkin’s (1982), which hypothesised that the dorsal and ventral streams served the coding of spatial and object features of the distal environment respectively. It is our contention that the perspectival challenge is based on a failure to appreciate the significance of the ‘consumer-based’ approach to characterising the functional profiles of the two visual streams.
6.1 Egocentric Frames of Reference and Visual Processing for Action

The first way that this failure manifests itself is by conflating two different senses of ‘egocentric.’ The meaning of the term 'egocentric frame of reference' as it is being used in A1 and A5 is radically different. The standard definition of an egocentric frame of reference is of a frame of reference in which locations, objects, object relations and other aspects of the scene are represented with respect to the perspective of the observer (Klatzky, 1998). This definition captures the meaning of the term as used in A5 (and in perspectival accounts) well. In contrast, however, this standard definition does not capture the way in which the term egocentric frame of reference is being used in A1 to refer to dorsal-stream processing for visually guided action. This misunderstanding follows from the failure to appreciate the move from an input-based to a consumer-based model of dorsal-stream processing. It is our contention that an appreciation of how the two senses of egocentric processing used in A1 and A5 differ from one another will help to clarify why the TVSH is not at odds with the perspectival account. In order to develop this argument, we need to delve into a bit of detail about the anatomical and functional specification of the dorsal stream.

The dorsal stream is an occipito-parietal circuit, which consists of multiple distinct projections from the occipital cortex to the posterior parietal cortex – a large and complex brain region. In addition to projections to areas in the midbrain and brainstem that mediate relatively ancient visuomotor mechanisms, the posterior parietal cortex has interconnections with at least three major cortical regions: 1)
pre-frontal cortex, which is implicated primarily in spatial working memory; 2) premotor cortex, which is implicated primarily in visually guided action; and 3) medial temporal cortex, which is implicated primarily in spatial navigation (see Kravitz, Saleem, Baker, & Mishkin, 2011 for a review). Because of their central role in the visual control of action, it is the interconnections of the posterior parietal cortex with the premotor cortex, midbrain, and brainstem that have been the major focus of the TVSH. It is in this context that the cortical and subcortical motor areas are said to be the consumers of information coded in egocentric frames of reference.

Accurate visuomotor performance requires that dorsal processing of visual information for action is carried out just-in-time (occurring only after the appropriate action plan has been selected) and is updated online (as the action is being performed). By virtue of its intimate interconnections with premotor cortex and subcortical motor areas, the dorsal stream provides the necessary information to these consumers in metrically accurate egocentric (i.e., effector-based) frames of reference (see Goodale & Haffenden, 1998 for a review). A paradigmatic example of the challenges facing the system in performing visually guided tasks is the accurate grasping of an object in reachable space. The act of grasping an object involves solving different problems such as transporting the hand to the target’s location and configuring the posture of the hand and wrist in a way that suits not only the geometric properties of the object itself but also the actor’s desired end or goal state. Doing this entails coordinating the timing and amplitude of muscle extensions, flexions, and rotations, all of which must work together to achieve the intended action.
Imagine you are sitting at your desk facing a computer screen with a cup of coffee to your right. You look at the cup and reach to pick it up, sip from it, and return it to the same spot as before. You then go back to looking at your computer screen. A few seconds later you reach out again for your cup only this time you maintain your fixation on the screen, seeing the cup only in the periphery of your visual field. These simple, mundane actions require your visual system to perform quite complex computational processes. In particular, in the first and second reaching actions your hand has to reach toward the same absolute spatial location, even though the image of the cup on the retina is different in each case. As this example highlights, the visual system has to solve the problem of reaching to the same location on the basis of visual information that has a different retinal position. In other instances, it has to solve the problem of reaching to different locations given the same retinal position of the cup. In fact, these examples can be extended to include different positions of the head, shoulder, limb, and so on down to the hand itself.

The dorsal stream is critically implicated in the transformation of visual information about the location of the cup in retinocentric coordinates into hand-based (i.e., effector-based) coordinates for visually guided action. As mentioned earlier, this transformation occurs at the moment the action is about to be initiated. In the case of picking up the cup, the just-in-time computation that occurs could involve a conversion from retino-centric to hand-centred coordinates or perhaps the selection of values in a look-up table (cf. Crawford, 2011; Thaler & Goodale, 2010). Whatever is the case, it is important to note that there is no one unified
representation of egocentric space in which all of the features of the scene are processed and are then used to guide action. After all, computing the effector-based coordinates for every possible action towards every possible target in the current visual array would be computationally intractable. Finally, it is important to note that even this account is an enormous oversimplification. In addition to the hand, the eyes, the arm, and the body also typically move when we reach out and pick up an object under visual control.

Different visually guided actions, which could range from grasping a cup to kicking a ball, each demand their own particular combination of visuomotor transformations. In fact, anatomically distinct circuits within the posterior parietal cortex are responsible for mediating different classes of visually guided actions from changing gaze direction to grasping an object. Moreover, the effector-based visual information for these different actions must be combined with vestibular and somatosensory information if the action is to be successfully performed. As a consequence, different parieto-premotor networks and different subcortical motor areas will be invoked for different actions. In other words, each action and action component is mediated by its own effector-specific network (for review, see Milner & Goodale, 2006).

To summarize, the notion of an egocentric frame of reference, as it is applied to dorsal stream processing for visually guided action, is essentially defined in terms the coordinates required by the consumers of dorsal outputs – those motor areas that mediate the myriad of actions that humans can perform under visual control.
Dorsal processing of visual information is characterized as being in an egocentric framework in virtue of providing accurate, metrical information. This information provides the precise location and orientation of the action relevant features of the visual field, in an effector specific framework, to enable the proper functioning of the effector specific consumers of that information.

Unlike the notion of egocentric processing outlined in our discussion of the perspectival account (section 2), where visual information is organised relative to the perspective of the observer, dorsal processing does not invoke a single, stable, and unified representation of the features of the distal scene relative to the observer. Instead dorsal processing is characterized as egocentric in the TVSH because it makes use of visual information about the location and disposition of objects relative to effector specific frameworks in order to guide skilled, target-directed actions. Egocentric processing for visually guided action in the dorsal stream seems best described as a buzzing web of effector-specific computations that change radically with every task-relevant alteration in the gaze position of the observer and/or the orientation of the relevant limb for the intended action. Moreover, these computations are carried out only when necessary, i.e., when the action is about to be performed. In short, the kind of egocentric processing that the TVSH proposes the dorsal stream performs is quite different from the notion of egocentric that is being appealed to in the perspectival account. In fact, we would argue that the egocentric properties of visual experience have far more in common with what is typically referred to as allocentric processing.
6.2 Allocentric Frames of Reference and Visual Processing for Perception

As we saw earlier, the claim that ventral processing of visual information for perception is allocentric has been taken to conflict with the perspectival account. However, there is an ambiguity here in the use of the term “allocentric”. As discussed above, one feature of an allocentric frame of reference, as it is defined in the spatial cognition literature, is that it is viewpoint independent. Obviously, if the TVSH were committed to the claim that all visual processing in the ventral stream were entirely viewpoint independent, it would be at odds with the perspectival account. That is, if the TVSH held that processing in the ventral stream were exclusively correlated with conscious visual experience – and if said processing were entirely viewpoint independent – then it would be unable to account for the perspectival (egocentric) features of visual experience discussed above. But again, the apparent tension between the TVSH and the perspectival account rests on the same misunderstanding of how proponents of the TVSH identify the functional role of the two streams. As with egocentric processing in the dorsal stream, allocentric processing in the ventral stream is defined relative to the consumers of the outputs of the ventral stream. Clarifying how the role of ventral processing is defined on the TVSH should reveal its compatibility with the claim that there are perspectival features of visual experience.

The ventral stream is an occipito-temporal network linking visual cortical areas and the anterior inferotemporal and prefrontal cortex along multiple routes of relatively
specialized systems for visual information processing. It projects to subcortical and cortical areas that are implicated in various cognitive tasks, including habit formation, emotion, long- and short-term memory, reward, and value. The main role of visual processing in the ventral stream then is to provide rich representations of features of the environment that can be used for recognition, encoded in memory, imbued with emotional meaning, or associated with reward, value, and habitual learning (see Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013 for a review). In fact, the conscious percept that we typically associate with our visual experience of the world is undoubtedly a product not only of ventrally-processed visual inputs but also of top-down expectations and conceptual knowledge. Thus the role of the ventral stream could be characterized as contributing visual information to a broad range of perceptually mediated cognitive operations. In doing so, we can consider some of the basic requirements placed on ventral processing by the kinds of consumers that demand ventral representations for the performance of different cognitive tasks.

One of the essential challenges the system faces in performing these cognitive tasks is to ensure that a particular cognitive operation (performed by a particular consumer module or set of modules) can transcend particular views or exemplars of the same object, object kind, scene, or scene kind. The classic case is object constancy, where an object is recognized as being the same object over a potentially infinite number of viewpoints and viewing distances (see Milner & Goodale, 2006, Chapter 2 especially). Although there is considerable debate about whether this involves the creation of a canonical object representation or instead a series of
interpolations over a set of different views (see Goodale & Haffenden, 1998), there needs to be some sort of object constancy for many cognitive operations (such as recognizing your house from all the others on the street). Moreover, when dealing with object perception, an essential task faced by the system is to transform the incredibly complex and continuously shifting retinal image into a stable representation that admits of re-identification, comparison between object instances and kinds, and encoding in memory (to name but a few tasks). This has to be achieved over long time-delays, from multiple viewing angles, and without being affected by extraneous contextual factors. Because of this, the central demand of many of the consumers of the outputs of ventral processing is for representations of objects and features of scenes that admit of re-identification and comparison among salient features, even when the position of the observer might change. In this sense, the representational output of the ventral stream may be said to be viewpoint independent, at least in the case of producing object representations for certain consumers.

Achieving such stable representations no doubt involves multiple different processes. But one way that the ventral stream contributes to the solution of this problem is by producing representations of object features in object-based frameworks and representations of objects in scene-based frameworks. Thus, as we move through the world, the perceived location of object features and objects remain constant with respect to the world even though their position with respect to us will change. It is this kind of scene- and object-centred reference frame that the TVSH is principally referring to with its use of the term allocentric processing. In
order to achieve object constancy, the visual system must somehow integrate information about our movements and our gaze position with information arising from the retina. The sources of information about our changing viewpoint are incredibly diverse: they range from proprioceptive cues about the position of our eyes, head, and body to much more abstract computations that allow us to, for example, perceive constancies of a scene unfolding on a movie screen. Importantly, however, in addition to perceiving these constancies we also perceive the position of objects and their locations with respect to us – even when our position is somewhat notional as is the case when we are looking at a picture or watching television. Notice that this kind of egocentric (perspectival) perception is completely different from the kind of egocentric (effector-based) computations carried out by the dorsal stream.

Although the allocentric nature of perception is important for understanding the world, the egocentric component is equally important. Indeed, one might argue that the perspectival or egocentric nature of perception plays a critical role in offline action planning. We would not, for example, reach out and attempt to pick up our coffee cup if it were resting on a table on the other side of the room. Nor would we turn left rather than right if our target location were to our right. Notice that these are examples of action planning: a cognitive process that depends on some sort of egocentric perspective, not the actual programming of the action where effector-based computations are required.
To recap, contrary to the characterisation in the perspectival challenge, the TVSH is not committed to the claim that all processing that occurs in the ventral stream is entirely viewpoint independent. Nor is the TVSH committed to the claim that observer-relative features of objects and object locations cannot be processed in the ventral stream. What the TVSH is committed too, however, is the claim that the perspectival features of perception do not arise from the effector-based computations of the dorsal stream that mediate visually guided action.

Although the TVSH has emphasized the important role of object- or scene-based representations in ventral-stream processing, nothing about this characterization precludes the notion of perspectival or egocentric perception also being realized by the ventral stream. As we have argued here: Some consumers of ventral stream processing – such as those involved in object recognition – demand viewpoint-independent computations; others – such as offline action planning and goal selection – require perspectival (egocentric) computations. Again, all of this is easy to understand, if one adopts a consumer-based account of ventral stream (and indeed, dorsal stream) processing.

7 The Psychophysical Basis of the Egocentric-Allocentric Distinction

We saw in the previous section that the TVSH is not committed to any strong claim about the viewpoint independence of the contents of visual experience. Instead, the claim that ventral processing is allocentric is largely based on the observation that
visual processing for perception relies to a large degree on scene-based coordinates. Importantly, much of the evidence for this claim comes from empirical work showing that judgements about visual stimuli rely to a large degree on contextual features, such as the scene-based spatial relationships of objects in the visual array. Such experiments have also shown that target-directed actions remain relatively immune to scene-based spatial relations (see for example, Ganel, Tanzer, & Goodale, 2008; Stöttinger et al., 2012; Stöttinger, Soder, Pfusterschmied, Wagner, & Perner, 2010). The implications of such psychophysical studies have been hotly debated over the last two decades (see Goodale, 2011 for a recent discussion of these issues). We have no intention of defending the conclusions that have been drawn from these studies here. Instead, we simply want to emphasise the fact that such psychophysical studies, which form a major plank of the empirical evidence for the TVSH, do not focus on the notions of viewpoint-dependent or -independent processing.

In an early experiment that formed part of the evidential basis for Goodale and Milner’s original hypothesis (Goodale & Milner, 1992), for example, Wong and Mack (1981) showed that while saccades to visible targets depend on the oculocentric (i.e., effector-based) location of those targets, saccades to a remembered location depend on the scene-based perceptual location of the target. In this study, the experimental subject was presented with a small target in a large surrounding frame. The frame and target then disappeared for 500ms before reappearing. While the target remained in the same location as before, the frame had been moved a few degrees to the right or the left. In a second experiment, the target moved in the same
direction as the frame, but only a third of the distance that the frame moved. In both cases, the subject perceived the frame as remaining relatively constant and the dot as changing its position. In fact, the experimenter-manipulated changes in the frame and the dot could be arranged such that the subject would perceive the dot as moving in the opposite direction to the actual displacement of the frame. Despite this illusory perceptual effect, subjects consistently directed their saccades toward the actual position of the target (its location in retinal, not perceptual, coordinates). However, when subjects were asked in similar conditions to perform a look-back task – in which they were instructed to look back to the original location of the target after having made a saccade to its new location – their memory-based saccades were determined by the relative location of the target to the frame, not by its actual retinal location. That is, while the subjects’ target directed saccades were not influenced (in this case) by scene-based cues about location that induce a perceptual illusion, memory-based saccades were. This study, and others like it (e.g., Bridgeman, Peery, & Anand, 1997; Goodale, Pélisson, & Prablanc, 1986), motivated the claim that the kinds of visual processing guiding the subjects’ performance of the two tasks differed.

While such studies focussed on dissociations in the coding of target location, more recent studies have looked at whether or not the dorsal and ventral streams process information about objects differently. For example, Ganel and colleagues (2008) ran a series of experiments in which two objects of different sizes were displayed within a Ponzo-type figure (the familiar railway-track illusion), creating the perceptual
illusion that the small object was actually larger than the large object. The results showed
that despite judging that the shorter object was longer (or vice versa), subjects’ in-flight
grip aperture reflected the real size of the target objects. However, when subjects were
asked to manually estimate the size of the target objects, rather than pick them up, their
manual estimates reflected the illusory size of the targets. In other words, despite the fact
that the illusion (by definition) affected the participants’ perception of target bar length,
the movement of the hand reflected the real lengths of the target bars.

Significantly for our discussion here, the literature on the resistance of grip scaling
to pictorial illusion has not, in general, related the findings to the perspectival
features of conscious visual experience. Instead of couching the discussion in terms
of the viewpoint-dependence or –independence of perceptual objects and scenes,
the illusory perceptual features of the studies are discussed in terms of scene-based
processing. As was argued in section 6.2, a scene-based reference frame can
incorporate perspectival features of visual experience. Thus, even when an
experimental result or effect is couched in terms of scene-based processing, this
effect might well be intimately associated with or even depend on perspectival
features of the scene. Indeed, many pictorial size-contrast illusions may depend on
features of the picture that contribute to perspectival features of conscious visual
experience (e.g., some elements in the Ponzo illusion may appear to be closer to the
observer than others). Nonetheless, the real core of the distinction between
egocentric and allocentric processing in the TVSH lies in the following claim: Ventral
processing for perception produces outputs that are rely on scene-based cues and
relative metrics, whereas dorsal processing for action produces outputs that are based on absolute metrics in an effector specific framework. As such, there is no immediate tension between the TVSH and the claim that there are perspectival features of visual experience.

**Conclusion**

The TVSH identifies a division between dorsal and ventral stream processing in terms of absolute metrical processing in an effector-specific frame of reference for visually guided action, and relative metrical processing in a scene-based frame of reference for perception. This division is characterised by the demands of the consumers of the outputs of dorsal and ventral processing. We have argued that the kinds of representations that proponents of the perspectival account are concerned with have little to do with the buzzing network of effector specific representations that the dorsal stream produces, and far more in common with the stable representations produced by the ventral stream. Thus it is far more likely that accounting for the perspectival features of visual experience will centre on the question of whether ventral processing can accommodate such features. This would be a problem for proponents of the TVSH only if their notion of allocentric processing could not accommodate such features. However, as we have argued, the functional description of ventral processing given in the TVSH in no way precludes such representations.
How the perspectival features of visual experience are achieved is an important question. It is possible that such features do require the processing of visual information in terms of absolute metrics in an observer-based frame of reference. It is also possible that the perspectival features of visual experience could be realised using relative metrics in a scene-based frame of reference. This matter has been the subject for discussion in a number of recent articles (Bermúdez, 2007; Jacob & Jeannerod, 2007; Kelly, 2008; Kiverstein, 2010; Noë, 2012; Schwenkler, 2014; Wu, 2014). Importantly, however, whatever turns out to be the case, it is highly likely that such perceptual features will be the product of ventral stream processing and therefore will not undermine the TVSH. In addition, the TVSH is not committed to any particular position with regard to how the perspectival features of visual experience are realised. Thus there is no simple argument from the perspectival nature of visual experience to the claim that the TVSH is false.

Acknowledgements

We would like to thank Chris Mole, Chris Viger, Jim O'Shea and the members of the lab associates group at the Rotman Institute of Philosophy for their helpful discussions, questions and advice on earlier versions of this paper.
Bibliography


