

# VISUAL DEMONSTRATIVES

**Mohan Matthen**

**University of Toronto**

## I. Introduction

Playing a game of squash, my opponent hits a drop shot to the left corner. I run to the front of the court – I do not bump into him or the walls as I do. Then, seeing that the ball has bounced high, I step out on my right foot, and hit the ball high on the front wall for a cross-court lob. What kind of visual information allows me to plan and then execute this complex act of coordination? I do not mean this as a neuroscientific, or even a psychological, question, but as one in philosophy, and though I hope that what I say will be scientifically sound, I shall be as empirically non-specific as I can.

The question can be posed in an old-fashioned framework. There is a process of practical reasoning that leads up to my action. Minimally:

1. I want to strike the ball.
2. The ball is there.
3. So: I must run there . . . etc.

The question is: What premises deriving from visual content must figure in such a process?<sup>1</sup> How do the various terms in the above relate to what we immediately and non-inferentially see?

---

<sup>1</sup> See Matthen 1988 and Burge 2005 for treatments (from somewhat different perspectives) of questions like this. Both papers emphasize how the assignment of intentional content to visual states facilitates accounts of visual data-processing given by cognitive science.

In this paper, I distinguish three sorts of idea that play a role in visually guiding action. My aim is to sketch an account of how these ideas – two types of visual idea and one non-visual (as I shall argue) – interact in visually guided action.

## II. Three Ideas of the Target

For the sake of simplicity, consider just the ball. How is it represented in the above process of reasoning?

1. First of all, I must have a game-related idea of the ball – an idea that gives it a place in the rules and tactics of a squash game – for this governs the formation of intentions such as striking it, doing so before it bounces twice, making it hit the front wall between the tin and the top line, making it difficult for my opponent to strike it, and so on. Speaking more generally, even when one is speaking of the relevance of visual data to behaviour, one needs to bring in non-visual ideas of the objects involved. For when I form the intention to act upon an object, I do so under an idea that fits it into my broader aims. These broader aims are rarely confined to evincing *behaviour* that satisfies certain physical parameters – running in a certain direction at a certain speed, hitting the ball with a certain force at a certain angle, and so on. What I want to achieve is usually comprehensible only in some behaviour-transcending framework. I want to win a point, neutralize the opponent's advantage in court-position, trap him against the back wall, and so on. These aims have to be achieved by my bodily movements, but they go beyond these movements. Moreover, my target cannot be specified just in visual terms. I aim to strike that small black sphere, but only because it is the ball in play. It is because the visual specification is identified with specification of the target under the governing idea that my physical behaviour is launched.

This is true even of animals incapable of explicit reasoning. Consider a dog chasing a ball thrown by its owner. The dog is retrieving for its owner. If its owner threw a Frisbee, the dog would chase after it; if some other person nearby threw a ball, her dog would not chase it (or can be trained not to). In this case too, it is because the dog's visual information is assimilated to its broader aims that its physical behaviour is

launched. Because the point extends in this way to animal actions, it is not restricted to situations in which highly acculturated terms are in play – as they are in my example of the squash game. Target-oriented behaviour, we might say, presupposes a mental representation of the target under which action on that target is chosen. This mental representation transcends behaviour understood in purely physical terms.

Returning to the squash ball, then, let us label the behaviour-transcending idea of it – the idea relevant to my squash game-related intentions –  $[B_i]$ . (When I mean to be talking about the idea, I shall put square brackets around  $B_i$ ; without these brackets, the symbol denotes the ball.)

2. In order to translate my game-related intentions into physical behaviour, I must be able visually to identify the ball. Having formed specific intentions with regard to the ball under the idea  $[B_i]$ , I need to know which object in my vicinity falls under this idea, track where it is, and so on. Thus, I might engage in an implicit mental process something like the following:

I want to strike  $B_i$ .

$B_i$  = the thing that looks like so.

So, I want to strike the thing that looks like so.

Let us say that there is a visual idea, or image, that corresponds to the phrase “the thing that looks like so” above. Call it  $[B_v]$ . In order to know where to direct my intended behaviour, I must visually search for and act on something that satisfies  $[B_v]$ .

3. Lastly, since physical action is in question, I must possess the information needed to control my body relative to the ball. For this, I must be able to locate and track the ball in “egocentric space” – by which I mean real space, measured in a coordinate system in which some point on my body is at the origin,  $\langle 0, 0, 0 \rangle_E$ . (The subscript marks egocentric coordinates; action-relevant representation of position is in polar coordinates, so that the first coordinate represents the target’s distance, the second its azimuth relative to some body-defined direction, and the third its elevation.) At a

## VISUAL DEMONSTRATIVES

certain point I must move my feet in such and such a direction. As I conduct my sequence of actions, I update the position of targets relative to me.

Identifying the ball visually is not sufficient for me to complete my intended course of behaviour. Targeted bodily movement requires me, or my visual system, to compute the position of the ball and other objects relative to my body, i.e., in egocentric space. Once I have done this, I can engage in the following reasoning:

I want to strike  $B_V$ .

$B_V$  = the material object at  $\langle r, \theta, \varphi \rangle_E$ .

So I want to strike the material object at  $\langle r, \theta, \varphi \rangle_E$ .

The upshot of these considerations is that I will have an egocentric location representation for the ball. Call this  $[B_E]$  – let it stand in for “the material object at  $\langle r, \theta, \varphi \rangle_E$ ” above. My bodily-movement schema at any given point of time will make use of  $[B_E]$ .

Empiricist orthodoxy maintains that the ball’s egocentric coordinates are given in visual consciousness – for, according to a philosopher like Hume, the “visual field” is egocentric. This seems to be a mistake. My conscious visual image of the ball and its trajectory is court-centred, not self-centred. I am not consciously aware of its egocentric trajectory – for instance, it does not present the appearance of moving faster in my direction as I accelerate toward it, though in egocentric terms, of course, it is moving faster. I am aware of its speed and my speed relative to the court. I am not conscious of  $[B_E]$ , it seems – at least not fully.

Hume was wrong, then, about the egocentricity of conscious visual representations; nevertheless, animals perform their tasks in ways that indicate that their movements are controlled egocentrically. As R. S. Woodworth (1899) demonstrated in a remarkable Ph.D. thesis, human voluntary movements are astonishingly accurate. For example, he observed some labourers pounding spikes with a sledgehammer for an hour. He estimated that the arc of their swing was about 150

cm; yet their 2 cm wide target was missed only once in an estimated total of 4000 swings among them all. For accuracy of this magnitude, he figured, the mean variance of the point of impact would be much less than 1 cm. Similarly, when we rapidly handwrite a row of letters, the corresponding points of the letters – for instance, the top and middle points of the ‘b’ and of the ‘h’, and the tails of the ‘g’ and ‘y’ – vary in height, Woodworth estimates, by less than 5%. This sort of accuracy in real time demands not only that vision should determine the position of the target, but also that it should provide this information to the efferent systems in the egocentric form that they require. Transforming allocentrically coded conscious vision into egocentric form would be far too slow. [B<sub>E</sub>] must, therefore, be computed independently. I’ll return to this point in section IV, below.

### III. Properly Visual Ideas

Before I can get to the main problem of describing the interaction among these ideas, I’ll need to deal with a preliminary question. Are the three ideas that I have posited really independent of one another, or are they simply three aspects or parts of a unified conscious presentation?<sup>2</sup>

First, I’ll defend the view that [B<sub>i</sub>] is different from [B<sub>v</sub>].

Earlier, I argued that [B<sub>i</sub>] – the idea of the ball that shows the point of the action undertaken – is not, as such, a visual idea. (Here, I am speaking generally: visual recognition tasks and the like might be couched in visual terms.) The ball is acted on under squash-tactical maxims of action. But I need visually to identify the ball in order to put game-related intentions into effect. And even when [B<sub>v</sub>] is on-line and governing my actions, the non-visual idea [B<sub>i</sub>] still figures in my process of reasoning. For suppose that while I am rushing to the ball, I notice that my opponent has retreated to the back

---

<sup>2</sup> See David Milner and Melvyn Goodale 1995, chapter 1, Andy Clark 2000, Scott Glover 2004, Goodale and Milner 2004a and b, and Mohan Matthen 2005, chapter 13, for background discussion relevant to the next two sections.

of the court. Then, I might change my plan. [B<sub>i</sub>] is in constant interaction with visual ideas of the ball. In light of these considerations, it seems that [B<sub>i</sub>] and [B<sub>v</sub>] are distinct and independent.

Now, the distinction between visual and non-visual ideas has come under attack. The ordinary language verb 'see' does not distinguish between purely visual ideas and ideas that have a non-visual component. I can say I see of something that it is *blue* and with equal linguistic propriety I can say that I see of it that it is a *five-dollar bill*. Yet, the latter requires that I subsume what I see under a non-visual concept, while the former is a visual concept that comes pre-packaged with the visual state. This distinction is elided in many philosophical treatments of vision. The notion has become well entrenched in some philosophical circles that there is no structural difference between the two examples given above – that seeing something as *F* is always a matter of having inarticulate visual sensations, which do not in, of, or by themselves present an object as possessing any property. Thus, even seeing that something is blue is taken to require a "view" or theory about blue things. (For a well-presented recent version of this theory see Anil Gupta 2006.)

Though it is not possible to argue the point in detail here, I regard this notion as utterly mistaken. First of all, there is no such thing as an inarticulate visual sensation – a sensation is always a presentation of some object as possessing some sensory property, and it is so in itself, not in virtue of other beliefs the perceiver holds. (See Matthen 2005, especially chapters 1-3.) We respond instinctively to sensations; contrary to Gupta (2006) we do not have to form a "view" about what they mean.

Secondly, it is only in such cases as seeing a dollar bill, and not in such cases as seeing something blue, that a further step of subsuming a visual sensation under a learned or inferred non-visual concept is required. For there are some *properly visual* ideas, and seeing a *V*, where *V* is properly visual, is a direct visual apprehension, for which a subject needs no theory. One simple way of making this point is the following, modelled on a move made by Sydney Shoemaker's (1968) and Gareth Evans's (1982)

discussion of “identification-free” judgements (*ibid* 179-191). Suppose that you identify a piece of paper poking out of somebody’s wallet as a five-dollar bill. There are two ways that your judgement can go wrong. First, you might be wrong about the colour of a Canadian five-dollar bill, and thus you may take a green-looking banknote to be a five-dollar bill. Let’s call this an error of *misconception*. Second, the banknote may look blue to you, and since you (correctly) think that all blue Canadian banknotes are five-dollar bills, you may take it to be a five-dollar bill. You are wrong because it is actually green. Let’s call this *misperception*. The point that I want to make is that with regard to properly visual concepts, error through misconception is impossible. You can’t be led into error by misconceiving blue – by mixing up the visual marks of blue with those of green, for instance. This is the mark of a properly visual idea.

*Conclusion 1* [B<sub>v</sub>] consists of ideas that are immune to misconception by anybody who can perceive them. [B<sub>i</sub>] contains ideas that are not immune to misconception. Thus, [B<sub>i</sub>] ≠ [B<sub>v</sub>].

#### IV. Egocentric Visual Ideas

Now, I’ll explore the character of [B<sub>e</sub>] and show that it is independent of [B<sub>v</sub>].

Some hold that when you look at the squash ball you get an indivisible package of visual information – that it is round, black, small, moving, and *there*. When you want to act physically on something, the argument continues, you search for it under an incomplete visual description – for instance, you might search for a black, round thing. When you make visual contact with this thing, you (ideally) get a much fuller package of visual information. This package includes its location relative to you. It is only by an act of abstraction that you can separate location out as a distinct visual idea of the ball. [B<sub>e</sub>] is an inextricable part of [B<sub>v</sub>].

The intuition that featural and locational information are inextricably linked accords with a notion of eye-hand coordination that some find intuitive. According to this notion, vision provides an image, which the subject must use in order to achieve

## VISUAL DEMONSTRATIVES

contact with external things. Think of a video game in which you simulate flying an airplane. You work by monitoring a television image. This image is not used in a purely egocentric fashion – when you shift about in your seat, the image shifts relative to you, but this makes no difference to your piloting. You act on your controls in such a way as to make the image change in certain ways. When an object gets too close, for instance, the imaged gap between the airplane and the object gets too small; you correct this by acting on your joy-stick in such a way as to make the gap grow bigger. When you want to land the airplane, you manipulate the joy-stick in such a manner as to make the axis of the runway continue on from the axis of the airplane – the closer end touching the nose – and then you make it get larger and larger while maintaining this alignment. Similarly, it seems, a human subject in the real world can translate his movement-plan into an image-manipulation plan. I catch a ball by first moving in such a way as to stabilize the position of the ball-image in my visual field, and then I catch it by reducing the gap between the images of the ball and of my hand. The intuition is that bodily movement is controlled indirectly by manipulating  $[B_V]$ . Let's call this the Act-by-Image-Manipulation model, or AIM. AIM appears to eliminate the need for  $[B_E]$ .

The AIM model of visual guidance is at odds with developments in cognitive neuroscience. There is now ample evidence that there is no smoothly integrated visual image of the sort envisaged in the intuitive picture of the control of action. In the last few years, the debate around this proposition has revolved around the supposed separation of two visual streams. One of these, the so-called ventral stream, is concerned with characteristics that objects possess independently of the illumination and perspective in which they may stand at the moment of viewing – characteristics such as colour, shape, surface texture, trajectory, allocentric position, etc. The other visual stream, sometimes called the dorsal stream, is concerned with positional information relevant to the selection and control of behaviour. The neurological



component of the distinction is irrelevant to my concerns, and I will not rely on it here.<sup>3</sup> What is just about uncontroversial is that visual processing for perceiver-independent object-qualities – I'll call this *descriptive* visual processing – is largely independent of processing for egocentric position, which I shall call *motion-guiding* visual processing. My aim is to argue that [B<sub>V</sub>], which is furnished by descriptive vision, does not substitute for [B<sub>E</sub>], which is furnished by motion-guiding vision. Later, we'll see how this carves out a special semantic role for [B<sub>E</sub>].

Here is a thought-experiment that brings out the independence of [B<sub>V</sub>] and [B<sub>E</sub>]. Think of two actions as follows. (A) Suppose that you are sitting in front of your desk, looking at it. There are several objects on it. You pick up the pencil and write something with it. Then, (B) you turn away from the desk, and recall a visual image of it. This time, you select the eraser. You mime picking it up and erasing something.

According to the AIM model, both actions are guided by a visual image similar to that which is generated by a sophisticated computer game – call this image [Desk-Hand<sub>V</sub>] – though in the case of (B), this image might be somewhat degraded. The idea is that in (A), you act on the pencil indirectly, by bringing about certain changes in [Desk-Hand<sub>V</sub>]. There seems to be no reason why you cannot do the same in (B). Of course, action (B) will not be as fluent and accurate as (A). Assuming that you haven't practised (B), in which case you will not be completely reliant on the image, you will not have the same confidence; your reach will be more tentative; it may not be in exactly the right direction or land at the right height; your grip may not be properly sized to the eraser.

---

<sup>3</sup> In Matthen (2005, chapter 13), I made a function-based distinction between descriptive vision and motion-guiding vision parallel to the seminal distinction between vision-for-perception and vision-for-action made by Milner and Goodale (1995) and Goodale and Milner (2004a). The functional distinction was meant to be independent of the ventral stream/dorsal stream distinction, also employed by Milner and Goodale. The latter is supposed to be the neurological substrate of the former, but it is conceptually distinct. This is not fully appreciated by some critics. Raftopoulos (2009) criticizes my distinction largely on the grounds that ventral and dorsal stream processing are not independent of one another. But I do not intend to make any claim about the anatomical loci of these visual functions. Raftopoulos agrees with my claim that visual reference is determined independently of visual description (*ibid.* 350), and this is the claim I wish to elaborate here.

## VISUAL DEMONSTRATIVES

Why? One reason for this is, of course, that the recalled visual image is not as detailed or accurate as an on-line visual image: this partly accounts for inaccuracies in your action. But there is another important deficiency in simulated action: the feedback that you get from the recalled image does not translate as smoothly into action. This, I believe, is important to understanding the failure to correct for inaccuracies in reaching, and for the slowness and tentativeness of the simulated action.

There are three phases of a *non*-simulated visually guided action such as (A). First, you search for the pencil by its visual characteristics. This first phase is clearly driven by descriptive vision. You use the generic pencil-image that you have stored in memory to select an object on your table by its visual characteristics. Next, your knowledge of the object and your intention determines a motor-schema – this will include trajectory, speed, style of grip, and force. Finally, vision somehow helps you translate the motor-schema into physical behaviour. (As we shall see in a moment, it is not the only guidance unit operating here.)

Now, let's consider the simulated action, (B). In its first phases, this action is very similar. You will select the eraser from the several objects presented by your recalled image, and select a motor-schema that depends on the action you intend to mime. Then, you launch your hand toward the eraser based on its position in the recalled image – presumably, this image is (or can be) arrayed in front of you in the same manner as a sensory image.

It is in this motion phase that the lack of visual feedback seems most debilitating. But why? You have an image of your desk in your mind's eye. That mental image presents the desk spread out in space. What is missing, relative to (A)? Is it that you cannot see the position of your hand relative to the position of the eraser in your mental image, and so cannot monitor the gap between hand and imagined eraser? But why is this a shortcoming? After all, you possess *bodily* awareness of the position of

your hand. And surely this is a large part of the guidance operating in (A).<sup>4</sup> For it is exceedingly unlikely that in the sorts of cases described by Woodworth (1899) – hammering a spike, writing a row of letters – visual feedback is the sole provider of spatial information. Vision seems to provide the position of the spike or the paper, and perhaps some of the last second course corrections, but it is bodily awareness that tells you how the hammer and the pencil are going, and how much force is needed complete the action.<sup>5</sup> The spatial awareness provided by vision and by bodily awareness are integrated. So why not AIM in the integrated image?

My intuition is that AIM is actually a good model of how motor-schemes are executed when vision is off-line. I would contend that the more fluid execution in case (A) shows that a better method is at work there. On-line vision guides the execution of motor-schemes by providing egocentric location information directly to the limbs. Off-line vision is unable to do this. In simulated action, we are therefore forced to AIM – and since this method is indirect, it is slower and more inaccurate. This leads to a distinction that is (perhaps) a bit crude and overstated. Descriptive visual content contains a message about how things are in themselves, independently of the observer’s perspective. This content can be used for movement, but only using AIM. Motion-guiding vision provides the limbs with egocentric coordinates that they can use directly, without the need to translate.

The difference of coding accounts for certain other discrepancies between conscious image and performance in reach-to-grasp manoeuvres. It was shown a while ago that the hand reacts to positional shifts of a target that are not consciously seen (Goodale, Péllison, and Prablanc 1986). When subjects reach for a target that is shifted

---

<sup>4</sup> See Santello, Flanders, and Soechting 2002; Winges, Weber, and Santello 2003 for evidence in support of this claim. Santello and co-workers show that simulated actions are deficient in some ways, but can approach visually guided action in other ways.

<sup>5</sup> I practised writing rows of letters with my eyes both open and shut. The shut-eye efforts were nearly as good with respect to uniformity of height and legibility, but quite a bit worse with respect to alignment with the line.

during a subject eye-saccade (just before the reaching action is completed), they are able to adjust their reach to the new position, even though they report not having seen it. This indicates a non-congruity between the conscious visual image and egocentric positioning. Again, size-contrast illusions – displays in which a target appears larger or smaller than it really is because of contrasting objects in the display – do not much affect grasping. The target may look larger than it is, but when a subject reaches out to grasp it, she sizes her grip appropriately (Aglioti, DeSousa and Goodale 1995). Once again, this shows a mismatch between the conscious visual image and the one determining movement toward the target.

It has been proposed that both discrepancies mentioned in the preceding paragraph arise from the differing functions of descriptive and motor-guiding vision. Descriptive vision is concerned with viewer-independent position, and suppresses random and sudden shifts because in all probability they arise from shifts in viewer position. But as far as motion-guidance is concerned, it is irrelevant whether these shifts are due to target or viewer movement – either way, they are relevant to control of movement, and are not to be suppressed as far as this application is concerned. Similarly, descriptive vision is concerned with object-size, and uses visual context to discount perspectival variation. By contrast, motion-guiding vision is not concerned with the perspective-independent size of the target, but just with where the fingers are relative to the target. To the extent that the *conscious* visual image derives from descriptive vision, it diverges from what is needed by motor-control functions.

*Conclusion 2* [B<sub>V</sub>] and [B<sub>E</sub>] are not artificially distinguished parts of a single image. On-line vision uses [B<sub>E</sub>] in visual guidance of bodily movement.

## V. Visual Objects

I argued earlier that vision directly gives us awareness concerning certain objects. I now want to show that some of these are material objects.

This sounds truistic, but it is actually somewhat controversial. For as David Lewis (1966) once wrote (summarizing a 1949 report by Roderick Firth): “Those in the traditions of British empiricism and introspectionist psychology hold that the content of visual experience is a sensuously given mosaic of colour-spots, together with a mass of interpretive judgements injected by the subject” (357). The idea is that vision presents the perceiving subject with such a “colour-mosaic”, which the subject interprets in order to construct a scene with objects distributed in the external world.

More recently, Austen Clark (2000) has constructed a theory in which the content of visual experience consists of visual features attributed to *places* in a three-dimensional visual field. Clark’s visual features are not restricted to colours, and are not spatially minimal. They include “colour, luminance, relative motion, size, texture, flicker, line orientation” (186). However, the content of visual experience does not include *material* objects, according to Clark – in his view, visual features are attributed to places, not material objects. “The characterization of appearance seems to require reference to phenomenal individuals: the regions or volumes at which qualities seem to be located,” he says (61) – the regions or volumes are subjects and the features are predicates. Clark uses the term “feature-placing” to describe the kind of content he ascribes to visual experience.

So, according to both the British empiricists of the Lewis-Firth report and a sophisticated contemporary philosopher of cognitive science steeped in recent neuropsychology literature, material objects are not delivered by vision.

There are good reasons for thinking that the feature-placing view is mistaken. To start with, it doesn’t make sense to say that visual features are *predicated* of places. Susanna Siegel (2002) puts the point well in a review of Clark (2000):

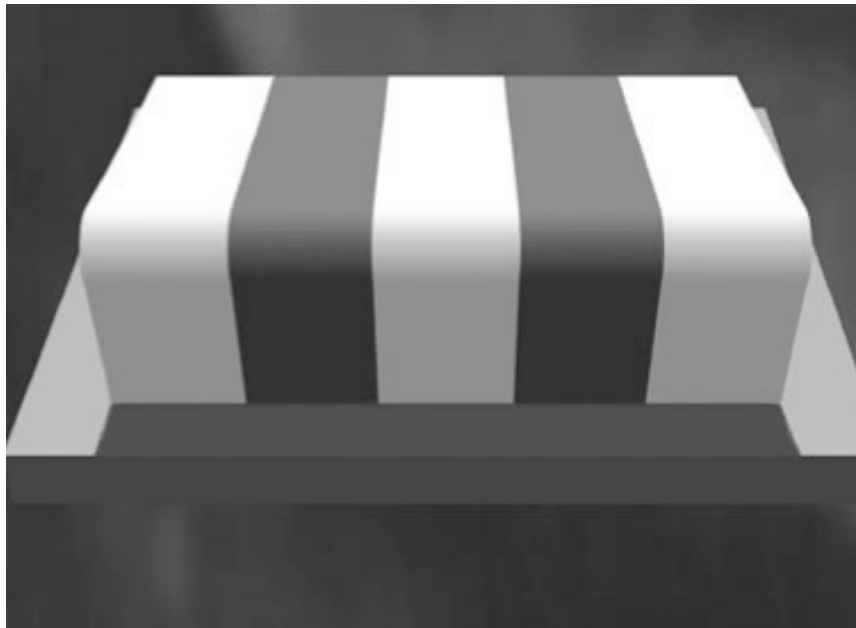
[Clark] repeatedly says that sensory systems attribute qualities to places. For example “The sensation of a red triangle ... picks out places and attributes features to them” (147; cf. 69, 70, 77, 165, 167, 185). Taken literally, these claims seem questionable. If audition told us that it was a place, rather than something at that place, that was cheeping, we would have all sorts of errors to correct in the move from audition to thought. We would

## VISUAL DEMONSTRATIVES

be similarly misled by vision if it told us that a certain region of space was red, while remaining neutral on whether anything occupying that place was red. (137)

Siegel's point is that it is literally false to say that a *place* is coloured, or that it is making a sound (as opposed to saying that there is something coloured in the place or that a sound was emanating from it). It is false, if for no other reason, then because the material object will take its colour and its noises with it when it moves. The colour that resides in a place can change simply because the thing that occupies it is replaced. It is for such reasons that while it may be permissible to say that colours are *in* places, it is not permissible to say that colours are predicated of them.

Another point to consider is that visual data processing employs algorithms that work only because they are applied to material objects. That is, visual data processing would not deliver veridical experience if the world were not a certain way. Consider this display in Dale Purves and Beau Lotto (2003, 57)<sup>6</sup>:



---

<sup>6</sup> This figure appears with the kind permission of Beau Lotto. See:

<http://www.lottolab.org/illusiondemos/Demo%2016.html>.

Light appears to be striking this object from somewhere behind. (Notice the shadow it casts in front, and the shade that envelops its front side.) Each stripe looks uniform in colour, but less brightly illuminated on the front side. Now, it seems clear in the image above that the dark stripes running along the top of the object are considerably darker than the light stripes on the front – the former look a darker grey. In fact, as you will discover if you cover everything else up, the lower parts of the light stripes (the parts that look as if they are in the shade) are exactly the same brightness as the upper parts of the dark stripes.

The illusion is explained by noting that the colour of the stripes is computed against the background of assumptions concerning how opaque objects intercept light. Since the brightness gradient decreases uniformly in a way that indicates indirect lighting, the visual system delivers experience as of stripes of uniform reflectance. There is no corresponding true assumption about *places* – places are not opaque; they do not intercept light. There is no shade or shadow in a world of places.

Vision is adapted, then, to the contingent presence in the world of things of particular sorts. Zenon Pylyshyn (2007) puts it in this way:

The mind has been tuned over its evolutionary history so that it carries out certain functions in a modular fashion, without regard for what an organism knows or believes or desires, but because it is in its nature. (ix)

Looking at the Purves-Lotto display above, it is clear that, whatever one might *believe* about the natural world, the *visual* world is simply not presented as a world of places. And this is because vision computes brightness as if it is dealing with a world that contains opaque material objects. Elizabeth Spelke (1990) has written that vision identifies material objects by their “cohesion, boundedness, rigidity, and no action at a

distance". These conditions are characteristic of material objects, and hence they have come to function as principles for the segmentation of scenes into *visual* objects.<sup>7</sup>

Now consider how motion is perceived. Again, Siegel (2002) puts the point well:

What happens in sensory phenomenology when a subject sees a basketball make its way from the player's hands to the basket? The information that it's one and the same basketball traversing a single path is not given by sentience if sentience is limited to feature-placing. On Clark's view, the information that it's one and the same basketball traversing a single path has to be given non-sensorily. The subject's visual experience stops short. (137)

To emphasize the point, consider the beta-phenomenon.<sup>8</sup> A light flashes to your left, goes off, and then another light flashes somewhat over to the right of the first one. If the interval is quite long – five seconds, say – the two flashes are seen as unconnected – a flash here, another flash there. However, as the interval decreases, the display is seen as a moving light. In fact, the light is seen as traversing the empty space in between the two flashes. What is the *subjective* difference between the two displays? Clearly, just that in the second case there is an illusory appearance of motion. But places do not move: motion consists of the same thing occupying different places at different times. Thus, Clark, who restricts visual ontology to places, cannot explain how vision can

---

<sup>7</sup> This leads to certain oddities of visual ontology: vision renders immaterial things such as images in mirrors, shadows, stains, and patches of light as visual objects. These appear as objects because they approximate Spelke's principles most of the time – though since shadows are cast on the nearest object that the light intercepts, they may suddenly expand or contract, and are non-rigid. Cast shadows look object-like, though they are visually distinguishable from material objects, but shade (as in the sides of objects facing away from illumination) does not look object-like.

<sup>8</sup> The phenomenon I am about to describe was called "beta" by Max Wertheimer, though it is mistakenly called 'phi' in common parlance (and by me in 2005, chapter 12). The rather different phenomenon that Wertheimer called phi is produced by rapid alternating flicker. In phi, we seem to see an occluder moving in front of the flickering lights. This occluder appears to be of the same colour as the background: it is a kind of negative space that appears in front of the light that is flickering "down" – the momentarily dimmer one.

See <http://www1.psych.purdue.edu/Magniphi/PhisNotBeta/index.html> for details.



generate the appearance of motion.<sup>9</sup> Yet, as Pylyshyn has long argued, vision not only detects movement, but also tracks objects and their features through movement.

This is a world in which most surfaces that we see are surfaces of physical objects, so that most of the texture elements we see move coherently as the object moves; almost all elements nearby on the proximal image are at the same distance from the viewer; and, when objects disappear, they often reappear nearby, and often with a particular pattern of occlusion and disocclusion at the edges of the occluding opaque surfaces, and so on. (2007, x)

What we seem to see in beta is a material object in motion; vision finds it through an application of Spelke's principles.

Finally, there is the evidence of visual perception in infants. Elisabeth Spelke, Renée Baillargeon, and others have observed the orderly emergence of object perception in infants as they grow up. They found results like these:

Infants were found to perceive a partly hidden object as a connected unit if the ends of the object moved together behind the occluder. Any unitary translation of the object in three-dimensional space led infants to perceive a continuous object: Vertical translation and translation in depth had the same effect as lateral translation . . . Perception of a moving, center-occluded object was not affected by the object's configurational properties: Infants perceived a connected object just as strongly when the object's visible surfaces when the object's visible surfaces were asymmetric and heterogeneous in texture and color as when they formed a simple shape of a uniform texture and color. (Spelke 1990, 33)

If these principles of object-segmentation were learned, as was assumed in the empiricist tradition, the pace of their emergence in infants could be expected to be

---

<sup>9</sup> Clark (2004, 569) responds: "flow patterns can give a powerful impression of movement (your movement) even though you do not perceive anything to be moving" – and gives the example of a blur or streak created by a rapidly moving object, in which the object itself is not seen. This response misses the point in two ways. First, whether or not we see an object moving in rapid optic flow, it is undeniable that we *do* see an object moving in the beta-phenomenon, and when we look at a ball being thrown. In these cases, we do *not* seem to see a temporal succession of coloured places. Secondly, it is not necessary to see the thing to which motion is attributed in order to see motion attributed to a thing. In the visual blur of a fastball, we see something moving very fast without seeing what it is that is moving very fast.

proportionate to the amount of exposure that a given infant has to the relevant data, and the infant's quickness to generalize. And we would also expect that individual humans would arrive at slightly different (though perhaps broadly accurate) segmentation principles – in the way that they arrive at different principles of, say, parallel parking or differentiating between the music of Mozart and Haydn. But they do not. Object segmentation emerges at more or less the same age in all infants, and the principles are the same from one to another. This is evidence that they are innate, and their emergence a matter of ontogeny and development rather than learning.

*Conclusion 3* Vision delivers direct awareness of material objects.

## VI. Visual Reference

Visual states are about *individual* things. And this creates a puzzle. Suppose I am in a darkened room, looking at an illuminated blue sphere – call it *S1*. Later, I am taken to another darkened room, and I look at another illuminated blue sphere, *S2*. Now it may be that since these two spheres look just the same, I have indistinguishable visual experiences in the two rooms. Yet it seems that in the first room my visual states were targeted on *S1*, while in the second they are targeted on *S2*. How does it come about that subjectively indistinguishable visual states can be directed toward different objects? Note that an image, such as a photograph, does not change its reference in the same way – it is targeted on the same individual regardless of where I look at it.

One possibility is that visual states single out their objects in a purely descriptive way. Suppose I seem to see a red disc at place *p*. According to the descriptivist theory, the thing I see is that which most closely resembles what I seem to see. In other words, the object of my visual state is that which most closely satisfies the descriptive content of my visual state. According to this theory, the content of my two visual experiences is 'blue sphere in front of me' or possibly 'blue sphere in front of me that is causing me to have this experience'. In the two rooms, different objects satisfy this description. The two experiences are the "same" because they both have this content – but the content is satisfied by different referents in different situations.

The descriptivist theory cannot properly accommodate misperception. Let  $x$  be the thing I see. I may misperceive  $x$  – suppose that  $x$  is orange, and that I misperceive it as red. Then, nothing satisfies the descriptive content of my visual state. The descriptivist theory would then imply that I perceive nothing (or, if formulated so that I perceive whatever comes closest, that I see a nearby red thing, even though this other thing has nothing to do with my visual state). But this seems wrong: it is  $x$  I see, even though I misperceive it.

It is worth noting that all of the visual and non-visual ideas that we have been discussing so far are subject to error in this manner. One can be wrong about an object being a squash-ball, one may misperceive its colour and shape, one may be wrong about where it is in terms of its egocentric coordinates.

What makes  $x$  the thing that I see, if it is not that it satisfies the descriptive content of my visual state? In a classic article, H. P. Grice (1961) argued that  $x$  is what I see because it causes my visual state (in the right kind of way). This theory offers us an initially satisfactory result: it allows that even things that are radically misperceived could be the objects of our perceptual states. This result runs counter to descriptivist theories in the right kind of way. However, I am thinking of visual states as reason-conferring states: states that rationally lead to *thoughts* and beliefs. Let us say that:

A visual state  $V$  is about  $x$  if and only if  $V$  *directly and by itself* gives the perceiver grounds for believing something about  $x$  – in brief, if  $x$  is a *direct epistemic target* of  $V$ .

I will argue that Grice's approach does not always give us the right result concerning direct epistemic targets. Grice may be correct in his analysis of the locution 'S sees  $x$ ', but if so, it would follow that what one sees is not always the direct epistemic target of one's visual state.

Under what conditions does a visual state constitute grounds for a perceiver to have a thought about an object? In the case of misperception described above, something looks to me as if it is red. That it looks this way to me gives me a defeasible

## VISUAL DEMONSTRATIVES

reason for thinking that it is a red disc. The visual state that I have described puts me in direct contact with  $x$  for epistemic purposes, though it is not accurate as far as the colour of  $x$  is concerned. And it may be that in this particular case, Grice's theory works – the object to which I gain direct epistemic access by means of my visual state is, as it happens, the thing that I see – that is, the object that caused me to have my visual experience.

But now think of a different kind of case. Suppose that I am looking at a red button and its reflection in a mirror. The image is not a physical object, and it has no causal power. It causes nothing. All the information that my eye receives comes from and is caused by the real button. According to Grice's theory, then, I see only one thing here, the real button, though I see it twice. However – and this is my point – my view of the image does not directly and by itself give me reason to believe of the real button that it is red. It *directly* gives me reason only to believe that the *image* is red. My view of the image gives me reason for believing that the button is red only in the presence of further beliefs about mirrors and images.

Here is another kind of example. You are in a cloud of fruit flies. You see hundreds of little specks. In Grice's view, you see each and every fly, because each causes some part of your visual image. But this visual state does not give you a reason to believe of any particular fly that it has any particular property. Let's suppose that the flies all look yellow. This gives you reason to believe that every fly is yellow. But in my view, this still does not give you a reason to believe of any particular fly that it is yellow. The reason is that you cannot visually single out any particular fly. You can form beliefs about a particular fly on the basis of how things look to you only if you can visually single it out.

Here is a way of thinking about direct epistemic targets that is consonant with these observations.

A visual state  $V$  has  $x$  as its direct epistemic target, only if  $V$  directly and by itself enables the perceiver visually to attend to  $x$ .

One cannot form a perceptual belief about an individual based on a visual state unless one attends to that individual. It follows that taken by itself a visual state can give a perceiver unmediated grounds for believing something about an individual only if it enables the subject to attend to that individual.

This gives the right kind of result in the case of the indistinguishable blue spheres. Each visual state enables me to attend to the sphere that is in fact in front of me, and not the other one. It gives the right result also about misperception: one may well be able to attend to an object despite being mistaken about its colour. Finally, the proposal is designed to deal with the fruit-fly case. You can visually single something out only if you can attend to it. Attention gives the condition under which one can not only receive information from an object, but also use that information to arrive at beliefs about the individual.<sup>10</sup>

Now, being able to attend to something is, among other things, a physical capacity. It depends on the subject's ability to turn his eyes to the thing, fixate it, focus on it, etc. By conclusion 2 above, it follows that vision controls attention through egocentric location coordinates. Of course, it can do this in error; in the case of a stick partly underwater, for instance, vision may direct your attention to a location that the stick does not occupy. But this is immaterial: attention is to the thing, not the location. The point that I find important here is that a first condition for forming beliefs about things on the basis of vision is that one is able physically to react to it.

*Conclusion 4* The direct epistemic target of a visual state,  $X$ , is that to which the egocentric coordinates  $[X_E]$  directs your attention.

## VII. Indexicality

Paul Snowdon (1981, 1990) makes a suggestion about direct epistemic targets that has a great deal in common with the one advanced in the preceding section. Snowdon

---

<sup>10</sup> This thesis is broadly consonant with John Campbell's (2002) treatment of visual reference.

## VISUAL DEMONSTRATIVES

proposes that when you visually perceive something, you are thereby capable of making a demonstrative judgement about it. Vision cannot give you reason to believe something about a particular object, unless it bestows upon you, directly and by itself, the ability to single the thing out and attend to it. This is a *physical* ability cognate with the ability to point to the thing, move toward it, and make a demonstrative judgement about it.

I want to flesh out this suggestion in a way that ties it to the visual ideas discussed earlier. My addendum to Snowdon's suggestion is that vision singles out its object by furnishing the perceiver with an egocentric location for that object. The location is not provided "explicitly" – that is, seeing something does not enable a perceiver to *say* where things are relative to her. Rather, seeing something enables a perceiver to attend to the thing and orient herself relative to it.

Egocentric coordinates are indexical. They determine a particular position in space, given the perceiver's own position. For any object that the perceiver sees, vision specifies egocentric coordinates for the object. These egocentric coordinates uniquely determine which objects the perceiver sees because it directs my attention to it. When I look at my computer, I see *it*. At home, I may have an exactly similar visual experience, because I am editing the same document on an exactly similar computer. Yet it is this computer I now see, not the one at home. This is because the egocentric coordinates that my visual system gives me for the computer are indexed to my current location, not my home location. Thus, visual reference is not purely descriptive – it is indexical. (With a photograph, it is different: images presented in a photograph are not indexed to my current location, and looking at the photograph does not enable me to orient myself with respect to the object depicted by it.)

Now, only objects in real space can be assigned egocentric coordinates. Objects such as the "stars", or phosphenes, that appear when we receive a blow to the head, after-images, etc., have no position in space. Hence, they cannot be demonstrated. An after-image has no position; hence its position cannot be indicated. Even if such things

## VISUAL DEMONSTRATIVES

appear, in some sense, to be in front of one, they do not look as if they are in external space. (See Siegel 2006 for discussion relevant to this point.) I'll summarize this position by saying that these private phenomena have only *phenomenal position*, and no egocentric position in the sense intended. I mean thus to acknowledge the appropriateness of positional relations such as 'to the left' etc. for images, but to distinguish these relations from those that imply location in space. After-images are seen as 'to the left' etc., but not as occupying any position relative to me, having any size relative to the size of my body, not as moving relative to my body, etc.

Since an after-image is not an object that occupies space, it makes no sense to ask whether they are the same as other objects outside of space. Suppose you are suddenly dazzled by a bright light and so are afflicted by an after-image for a few minutes. After a minute or so, somebody asks you: is the pink spot you now see the same as the pink spot you saw a minute ago? There is no good answer to the question as asked – the after-image has not shown spatio-temporal continuity (since it occupies the same position in your visual field despite your own motion), but on the other hand it has, in some sense, persisted. After-images have no location. So though it makes sense to ask whether the disturbance in your visual field is continuous and located in the same visual field-place – it does not make sense to ask whether it is the same *object*. There *is* no object here, and no appearance of one. As Snowdon (1981) says, we cannot demonstratively identify after-images and the like – “only objects, so to speak, in the world can be so identified” (190). With such phenomena, things may look to the perceiver as if there is a spot of light or floating spot in front of him, but there is no object about which he can form a belief, and hence no epistemic target of his visual state.

Only on-line seeing *directly* gives you egocentric location in this sense. As argued in section IV, neither recollection nor mere imaging is capable of guiding bodily motion. This does not mean that only on-line seeing is targeted on objects. If a state is directly descended from or created from a state that directly gives you egocentric location, then it is targeted on the object given in the ancestor state. For example, when I try to

imagine what my daughter would look like in a blue raincoat that I am thinking of getting her as a birthday present, the image I conjure up is targeted on my daughter even though it does not assign egocentric location to her. However, this state does not allow me to attend to my daughter, or gather information about her.

*Conclusion 5* There is an element in on-line seeing (as distinct from recalling, imagining, etc) that indexically links visual states to external objects; this fails for internal objects such as phosphenes and after-images.

### VIII. Disjunctivism

Snowdon (1981, 1990) endorses a view known as disjunctivism on the basis of his view about demonstratives. I have endorsed the position concerning demonstratives. I will conclude with a critique of disjunctivism.

Here is the position that Snowdon advances:

DIS. The best theory for the state of affairs reported by 'I seem to see a flash of light' is that EITHER there is something I can demonstratively identify that looks to me to be a flash of light OR it is to me as if there is something that I can demonstratively identify that looks to me as if it is a flash of light (but there is not).<sup>11</sup>

By saying that this is the 'best theory', Snowdon means to imply that the two disjuncts specifically describe different kinds of situations, each of which would be correctly, but non-specifically, described by 'I seem to see a flash of light'. Thus, neither disjunct can be deleted from the definition without sacrificing completeness.

Snowdon offers us an interesting example (which he takes from J. N. Hinton) to make his point. I am sitting in a darkened room and I seem to see a very brief faint flash of light.

---

<sup>11</sup> This wording is a composite assembled from Snowdon (1981), 184-185.



## VISUAL DEMONSTRATIVES

Consider, first, a case in which:

- a. There is really a light that I see.

In this case, there really is something that looks to me as if it is a flash of light. It has egocentric location, moreover, and I can demonstratively identify it. So, DIS (above) works: the first disjunct is satisfied.

Now consider a different case. Suppose that:

- b. My visual presentation is *as of* a light; i.e., it is to me as if there is a light. However, there is no light there – it is an after-image.

Theorists opposed to disjunctivism think that in b, my visual state is exactly the same as in a where what I see is really was a flash of light. So, they say, this too is a case in which I seem to see a flash of light. And in this case, too, these non-disjunctivists say, there is something that looks to me like a flash of light – namely, the after-image. Thus, the non-disjunctivists argue, the first disjunct is the best theory of both cases, and there is no reason for the second disjunct to be added in.

Snowdon disagrees with this. One simple way of arguing the point (not Snowdon's, but he would agree with the premises) is this. It is not possible, as I argued in the preceding section, demonstratively to identify an after-image. In case b., therefore, there is no possibility of demonstrating the object that looks to me as if it is a flash of light. Thus, Snowdon argues, it is not the same experience as in case a, which *does* support demonstrative identification. This is his reason for thinking that there are two quite different states of affairs in which 'I seem to see a flash of light' is true – one kind supports egocentric location and the other does not. Snowdon says:

The disjunctive picture divides what makes looks ascriptions true into two classes. In cases where there is no sighting they are made true by a state of affairs intrinsically independent of surrounding objects; but in cases of sighting the truth-conferring state of affairs involves the surrounding objects. (1981, 186)

## VISUAL DEMONSTRATIVES

Looks-judgements are made true by two types of occurrence: in hallucinations they are made true by some feature of a (non-object-involving) inner experience, whereas in perceptions they are made true by some feature of a certain relation to an object, a non-inner experience, (which does not involve such an inner experience). (1990, 130)

I agree with Snowdon about a number of aspects of this case. But it seems to me that he is wrong about case b. Translating it into my terms, I would say if the experience in case b is genuinely *as of* a light – if it genuinely is to me that as if there is a flash of light – then it should seem to me as if the light is in a certain position relative to me. In my way of thinking about this matter, it is characteristic of seeming to see an object that that object is endowed with egocentric, not merely phenomenal, location. Consequently, my visual state in b. will *seem* to support pointing, moving towards, etc. – though *in fact* it does not support pointing, moving towards, etc. Thus, my visual state *does* support a demonstrative – that is, it assigns egocentric location to the flash of light – but because my visual state is inaccurate, the demonstrative that it supports is vacuous, and does not single anything out.

It is instructive, here, to consider two further cases:

c. My visual presentation is *as of* an after-image; it is to me as if I am suffering an after-image. However, I am not suffering an after image; it is a real light that I see.

Here, it seems to me, my visual state fails to assign egocentric location coordinates to the thing that I see. If it seems to me as if it is an after-image, then it seems to me as if it isn't a real thing in the external world, and hence it is not sensed as having real location relative to me, only phenomenal location. In this sense, it is experientially like:

d. My visual presentation is *as of* an after-image; it is to me as if I am suffering an after-image. And I am indeed suffering an after-image.

Snowdon maintains (1981, 190) that “a person seeing a light but believing that he is having an after-image *may be allowed to make a demonstrative judgement to the effect that that is an after-image*” (my emphasis). I take it that he means that case c

supports a demonstrative. Thus, Snowdon thinks that in case c that I will have judged of the light that it is an after-image. This is where I disagree: phenomenal location does not even seem to support demonstrative identification. Therefore, I am unable to make a judgement about the light.

Here, then, is the difference between Snowdon's position and mine. Snowdon thinks that demonstrative thought is impossible without something that is demonstrated. Thus, he thinks that the condition under a visual state is demonstrative is an external condition – i.e., whether there is a light there. I think that a visual state is demonstrative is demonstrative if it assigns egocentric coordinates to a thing. This is an internal condition. On my way of thinking, demonstrative thought is vacuous when something seems to possess egocentric location, but does not. Further, Snowdon thinks that after-images appear to have egocentric location, and that demonstration is not only possible but successful when there is something at the egocentric location that the after-image appears to have. I believe, on the contrary, that part of what it is to seem to see an after-image is to seem to see something that has no egocentric location.

In my view, it is that a and b present things as possessing egocentric location, while c and d do not. This view is different from the one that Snowdon is most concerned to oppose. His main target is the view is that there is some core experience that is in common to all four cases. I join him in rejecting that view. On the other hand, I reject the position that, as Snowdon puts it, "it is quite possible for elements (objects, or states of affairs) external to the subject to be ingredients of an experience" (1990, 124). This leads him to the view that a and c are genuine demonstratives, and b and d not. My view, to repeat it once again, is that demonstration requires only something that looks like an object and which vision endows with egocentric coordinates.

## **IX. Conclusion**

I have argued that on-line visual states assign seen objects egocentric locations. It is by means of these location-assignments that perceivers act on these objects quickly and accurately. Egocentric location-assignments also enable the perceiver to attend to the

## *VISUAL DEMONSTRATIVES*

objects, and thus to form beliefs about them on the basis of nothing other than the visual state itself. Off-line visual states such as recalling and imaging do not assign seen objects egocentric locations, and do not directly support physical movement or information gathering about any object. Moreover, subjective visual phenomena such as after-images are also not assigned egocentric coordinates. This is why these phenomena do not seem to be presentations of real external objects.

REFERENCES

- Aglioti, Salvatore, Joseph F. X. DeSousa, and Melvyn A. Goodale. Size Contrasts Deceive the Eye but Not the Hand. *Current Biology* 5: 679-85.
- Burge, Tyler. 2005. Disjunctivism and Perceptual Psychology. *Philosophical Topics* 33: 1-78.
- Campbell, John. 2002. *Reference and Consciousness*. Oxford: Clarendon Press.
- Clark, Andy. 2000. Visual Experience and Motor Action: Are the Bonds Too Tight? *Philosophical Review* 110: 495-519.
- Clark, Austen. 2000. *A Theory of Sentience*. Oxford: Clarendon Press.
- Clark, Austen. 2004. Sensing, Objects, and Awareness: Reply to Commentators. *Philosophical Psychology* 17: 563-589.
- Evans, Gareth. 1982. *The Varieties of Reference* (edited by John McDowell). Oxford: Clarendon Press.
- Firth, Roderick. 1949-50. Sense-Data and the Percept Theory. *Mind* 58: 434-465 and 59: 35-56.
- Gendler, Tamar Szabó and John Hawthorne, eds. 2006. *Perceptual Experience* Oxford: Clarendon Press.
- Glover, Scott. 2004. Separate Visual Representations in the Planning and Control of Action. *Behavioral and Brain Sciences* 27: 3-24.
- Goodale, Melvyn A. and A. David Milner. 2004a. *Sight Unseen: An Exploration Of Conscious And Unconscious Vision*. Oxford: Oxford University Press.
- Goodale, Melvyn A. and A. David Milner. 2004b. Plans for Action. *Behavioral and Brain Sciences* 27: 37-39.

## VISUAL DEMONSTRATIVES

- Goodale, M. A., D. Péllison, and C. Prablanc. 1986. Large Adjustments in Visually Guided Reaching Do Not Depend on Vision of the Hand or Perception of Target Displacement. *Nature* 349: 154-156.
- Grice, H. P. 1961. The Causal Theory of Perception. *Proceedings of the Aristotelian Society* Supplementary Volume 35: 121-152.
- Gupta, Anil. 2006. Experience and Knowledge. In Gendler and Hawthorne: 181-204.
- Lewis, David K. (1966). Percepts and Color Mosaics in Visual Experience. *Philosophical Review* 75: 357-368.
- Matthen, Mohan. 1988. Biological Functions and Perceptual Content. *Journal of Philosophy* 85: 5-27.
- Matthen, Mohan. 2005. *Seeing, Doing, and Knowing: A Philosophical Theory of Sense-Perception*. Oxford: Clarendon Press.
- Milner, A. David and Melvyn A. Goodale. 1995. *The Visual Brain in Action*. New York: Oxford University Press.
- Purves, Dale and R. Beau Lotto. 2003. *Why We See What We Do: An Empirical Theory of Vision*. Sunderland MA: Sinauer.
- Pylyshyn, Zenon W. 2007. *Things and Places: How the Mind Connects with the World*. Cambridge MA: MIT Press.
- Raftopoulos, Athanasios. 2009. Reference, Perception, and Attention. *Philosophical Studies* 144: 339-360.
- Santello, Marco, Martha Flanders, and John F. Soechting. 2002. Patterns of Hand Motion during Grasping and the Influence of Sensory Guidance. *Journal of Neuroscience* 22: 1426-1435.
- Siegel, Susanna. 2002. Review of Austen Clark, *A Theory of Sentience*. *Philosophical Review*, 111: 135-138.

## VISUAL DEMONSTRATIVES

- Siegel, Susanna. 2006. Subject and Object in the Contents of Visual Experience. *Philosophical Review* 115: 355-388.
- Snowdon, Paul. 1981. Perception, Vision and Causation. *Proceedings of the Aristotelian Society, New Series*, 81: 175-192.
- Snowdon, Paul. 1990. The Objects of Perceptual Experience. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 64: 121-166
- Spelke, Elizabeth S. 1990. Principles of Object Perception. *Cognition* 14: 29-56
- Winges, Sara A., Douglas J. Weber, and Marco Santello. 2003. The Role of Vision on Hand Preshaping During Reach to Grasp. *Experimental Brain Research* 152: 489-498.
- Woodworth, R. S. 1899. The Accuracy of Voluntary Movements. *Psychological Review Monographs*