Shadows remain segmented as selectable regions in object-based attention paradigms

Lee de-Wit
Laboratory of Experimental Psychology, University of Leuven, Belgium; e-mail: Lee.deWit@ppw.kuleuven.be
David Milner
Department of Psychology, University of Durham, UK; e-mail: a.d.milner@durham.ac.uk
Robert Kentridge
Department of Psychology, University of Durham, UK; e-mail: robert.kentridge@durham.ac.uk

Abstract. It is unclear how shadows are processed in the visual system. Whilst shadows are clearly used as an important cue to localise the objects that cast them, there is mixed evidence regarding the extent to which shadows influence the recognition of those objects. Furthermore experiments exploring the perception of shadows per se have provided evidence that the visual system has less efficient access to the detailed form of a region if it is interpreted as a shadow. The current study sought to clarify our understanding of the manner in which shadows are represented by the visual system by exploring how they influence attention in two different object-based attention paradigms. The results provide evidence that cues to interpret a region as a shadow do not reduce the extent to which that region will result in a within-'object' processing advantage. Thus, whilst there is evidence that shadows are processed differently at higher stages of object perception, the present result shows that they are still represented as distinctly segmented regions as far as the allocation of attention is concerned. This result is consistent with the idea that object-based attention phenomena result from region-based scene segmentation rather than from the representations of objects per se.

Keywords: objects, attention, vision, shadows, segmentation, shape.

1 Introduction

The pattern of stimulation that falls upon the retina is not equally informative. The limited capacity of the visual system can be employed most effectively if it is focused on those aspects of visual structure that most effectively enable us to recognise objects and to act upon them. Shadows pose an interesting representational challenge in this context because on the one hand, they do not represent inherent structure in the environment, but on the other, they are potentially informative with respect to the objects that cast them. Reviewing the status of shadows in the human visual system reveals a somewhat mixed picture, in which inconsistencies in the pattern of illumination implied by shadows are hard to identify (Ostrovsky et al 2005) and in some contexts have no measurable influence on object recognition (Braje et al 2000; Bonfiglioli et al 2004), but in other contexts shadows clearly aid and/or interfere with object recognition (Tarr et al 1998). Moreover, the effect of illumination on the shading *within* an object clearly has a direct influence on shape perception (Ramachandran 1988). Whilst shadows thus have a mixed role in object recognition, they do seem to play an important role in computing the location (Imura et al 2006; Yonas and Granrud 2006) and movement profile (Kersten et al 1996) of the objects that cast them. Most critically for the current study, however, is evidence that access to the visual form of a region is influenced by whether or not it can be interpreted as a shadow (Rensink and Cavanagh 2004; Lovell et al 2009). This article seeks to build on the fact that the interpretation of a region as a shadow can alter the accessibility of the form of that region by testing whether cues that alter the interpretability of a region as a shadow influence the allocation of attention to that region, using 'object'-based attention paradigms.

The concept of an object is potentially controversial in the context of attention (see Driver et al 2001 for an insightful critique of the use of the concept of an object in the context of attention, a topic to which we will return in the Discussion); however, it is generally acknowledged that some form of perceptual organisation influences the allocation of attention (Scholl 2001). Furthermore this perceptual organisation that leads to the objects selected by attention cannot be understood simply in terms of contrast boundaries; rather, it is the manner in which those boundaries are interpreted that is critical (Anstis 1990; Naber et al 2011; Albrecht et al 2008; Tadin et al 2002; Scholl et al 2001). A compelling example of this (although not typically cited in the context of the object-based attention literature) is provided by Anstis (1990), who shows that it is essentially impossible to track the intersection between two lines when those lines appear to be two separately moving items. The same physical intersection is, however, trivial to track when those lines can be interpreted as moving together (as one, cross like, shape). More recently the role of higher level perceptual organisation has been demonstrated in paradigms that exploit the 'within-vs-between object' advantage (cf Egly et al 1994). Classically, these paradigms demonstrate that a target is more easily detected or discriminated when it is preceded by a cue or paired with a comparison target within the same, rather than in a different, object. Albrecht et al (2008) recently showed that whilst this within-vs-between object advantage was elicited by a set of contrast boundaries forming a figure on top of a surface, the same contrast boundaries would not elicit this effect when perceived as a 'hole' cut into that surface. In another recent example Naber et al (2011) exploited a bi-stable grouping stimulus (cf Lorenceau and Shiffrar 1992) to show that the facilitation effect seen in comparing two targets on the same object is contingent not just on the sensory input but also on whether that input is perceptually grouped or not. The fact that higher level factors in perceptual organisation can influence the extent to which a set of closed contrast boundaries can be selected as a unit of attention opens the question as to whether the interpretation of a given region as a shadow would influence the extent to which that region could be selected by attention.

The idea that the same physical shape can be represented in a perceptually less salient manner when interpreted as a shadow was suggested by work, mentioned above, by Rensink and Cavanagh (2004). More specifically, Rensink and Cavanagh (2004) asked participants to search for a target that had a different orientation to other items in a display. Rensink and Cavanagh found that this visual search, or oddity detection, task was harder to perform when the items in the display could be interpreted as shadows. The interpretability of the stimuli as shadows was manipulated in a number of ways; one clearly effective method (also employed in the current set of experiments) simply required the removal of the object that cast the shadow.

To review, shadows have a rather mixed status in influencing different elements of visual perception, they play a clear role in influencing the perception of the location of the objects casting them, they play a more ambiguous role in aiding the recognition of the objects that cast them, and of particular importance for the present study, there is evidence that the interpretation of a region as a shadow appears to lead to that region being processed differently. It is in this context that the current project set out to test the status of shadows in two different object-based attention paradigms, in which cue-target pairings (Egly et al 1994) and the comparison of two targets (Ben-Shahar et al 2007) are facilitated when presented on the same object. Given previous evidence that the manner in which a scene is organised will impact the extent to which regions of that scene will show 'within-object' advantages, this research seeks to exploit the techniques developed by Rensink and Cavanagh for manipulating 'shadow-hood' to ask whether the same region

will be differentially selectable as a unit of attention when that region can or cannot be interpreted as a shadow.

2 Experiment 1

2.1 Participants

Twenty-five participants were recruited in exchange for course credits. All participants were first-year psychology undergraduates with normal or corrected to normal vision. Participants were naive as to the aims of the study. Participants' ages ranged from 18 to 38 (mean 20). There were 8 males and 17 females.

2.2 Materials

The experiment was programmed in Borland C++, with the use of DirectX to ensure accurate timings. Stimuli were constructed off-line before the experiment using the 3-D rendering package 'POV-Ray'. This programme uses accurate ray tracing in order to visualise the way in which light will disperse in 3-D scenes, and thereby provides an optimal image-rendering environment in which to generate veridical shadows. The stimuli were presented on 17 inch PC monitors with a screen resolution of 1024 by 768 pixels.

2.3 Design

The experiment employed a 3 by 2 within-participants design. The first factor, Cue Type, had three levels: valid (cue appears in same location as the target), invalid-within (cue appears at a different location but on the same region), and invalid-between (cue appears at a different location on a different region), as shown in Figure 1. There were equal numbers of each cue type, and these were randomly distributed across trials. The second factor, Shadow Type, had two levels: either the shaded regions could or could not be interpreted as shadows. In line with Rensink and Cavanagh's (2004) study this was achieved by removing the objects that cast the shadows. Shadow or Non-shadow-like stimuli were presented in separate blocks, the order of which was counterbalanced across participants.

One of the critical factors in the original Egly et al (1994) paradigm pertains to the distance between the cues and the targets. In order to establish the effect of objects upon attention, it was critical to equate the distance between and within objects, such that any differences could not be explained in spatial terms. The Egly et al result and the many replications that have followed have clearly demonstrated that this influence of object structure upon attention occurs in addition to that accounted for by the spatial distances between cues and targets. The fact that our stimuli are rendered in a 3-D perspective (in order to enhance the perception of our stimuli as shadows) complicates this aspect of control. This complication arises because it is evident that attention is influenced by 'size constancy' such that attention operates in perceived rather than retinal space (Robertson and Kim 1999). It is therefore more appropriate to align the distances between the stimuli in terms of the 3-D environment rather than the 2-D distances that will hit the retina. Thus, whilst the distance between the four cue/target locations was set to be equal in the reference frame of the coordinates of the surface of the rendered environment, this resulted in unequal distances on the retinal image. More specifically, the two horizontal distances are not identical on the screen (front = 7.0 deg, back = 4.4 deg) and the vertical distances between the target locations was 4 degrees of visual angle. While highlighting this difference in visual angle, it should be borne in mind that it was present in the Shadow and Non-Shadow conditions.

3 Procedure

Each participant completed two blocks of 32 practise trials and then 192 actual trials. Each block contained equal numbers of valid, invalid-within, and invalid-between trials. On each

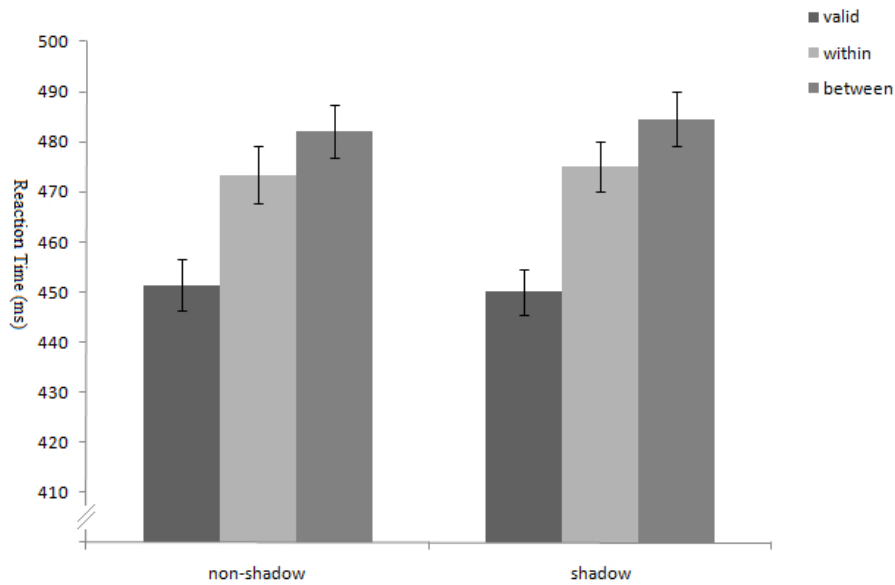|	Non-shadow condition: Region without casting objects	Shadow condition: Region with casting objects
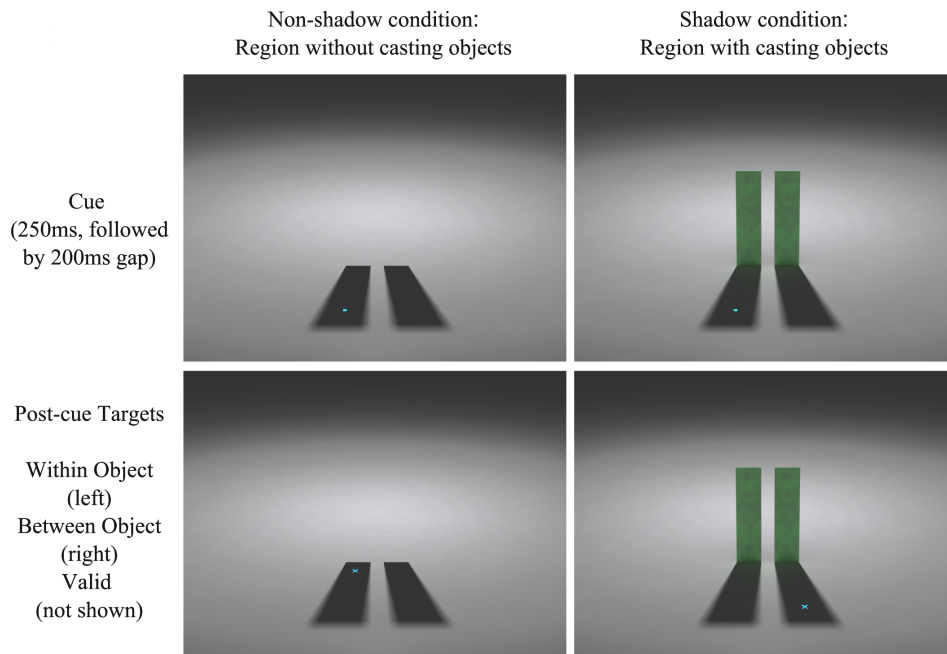
Figure 1. On the left are the shaded regions without casting objects; on the right are the shaded regions with casting objects. The relationship between cues and targets is depicted on the left for within-object cuing and on the right for between-object cuing. In the valid condition (not shown) a target would be presented in the same location as the cue.

trial the participant would be presented with a cue for 250 ms. The cue was a small square appearing in one of the four possible target locations. The use of this cue type was chosen because Rensink and Cavanagh (2004) showed that contrast outlines around the edge of a bounded region (like those typically used as cues in the Egly et al paradigm) can disrupt the interpretation of that region as a shadow. After the cue there was a 200 ms gap before the participants would be presented with one of two targets (an X or an N). The participant simply had to report (using the X and N keys) which target had been presented. There was then a 500 ms pause before the next trial started during which the scene remained visible. The participant was not required to maintain fixation.

4 Results

The results were analysed using a two factor repeated measures ANOVA. The reaction time data showed a clear effect of Cue Type ($F(2,24) = 70.728$, $p < 0.0001$). Comparing only the within-vs-between object Cue Types yielded a significant difference ($F(1,24) = 16.99$, $p < 0.001$) reflecting that targets presented on the same object as the cue were detected more rapidly. This within-vs-between object advantage did not interact with Shadow Type ($F(1,24) = 0.026$, $p = 0.874$), and Shadow Type had no overall influence on the reaction time ($F(1,24) = 0.164$, $p = 0.689$).

The accuracy data revealed no significant main effect of Cue Type ($F < 1$), Shadow Type ($F < 1$) or Cue x Shadow interaction ($F(1,24) = 1.326$, $p = 0.261$). The accuracies for the 'valid', 'invalid within', and 'invalid between' were 94.6%, 94%, and 95.1% for the non-shadow and 93.6%, 94%, and 94.6% for the shadow condition, and their associated standard errors were 0.89, 0.84, 0.64, 0.99, 0.88, and 0.85, respectively.

Figure 2. Reaction time data for the valid, invalid-within, and invalid-between conditions in non-shadow (regions without casting objects) and shadow (regions with casting objects) conditions (error bars represent standard mean error).

5 Interim discussion

The results of Experiment 1 revealed a clear replication of Egly et al's (1994) demonstration of the influence of perceptual organisation upon attention, indicating that shadows are represented by the visual system as 'objects' in this context. That is, the results provide no evidence that shadows are treated as less salient aspects of visual structure than other bounded surfaces with respect to object-based attention. The Egly et al paradigm is, however, just one of a range of tasks that can be employed to illustrate the influence of visual structure upon attention (Duncan 1984; Driver and Baylis 1989; see Scholl 2001 for a review). Ben-Shadar et al (2007) have, for instance, employed a two-item comparison task in which the participant has to report whether two items (which can either be located on the same or on different objects) are the same or different. They found that when briefly presented targets were presented on different 'objects', participants were less accurate in making such judgements. Experiment 2 therefore seeks to explore whether the effects found in Experiment 1 with the Egly et al cuing paradigm will be replicated with this 'divided attention' object-based paradigm. Finally, because shadows come in numerous potential forms, a different shadow type was used, to generalise the result to a shadow cast onto a different surface than that to which the casting objects are attached (see Figure 3).

6 Experiment 2

6.1 Participants

Thirty participants were recruited in exchange for payment or course credits. All participants had normal or corrected to normal vision and were naive to the aims of the study. Participants' ages ranged from 18 to 57 (mean 22: 29 of the participants were aged 18 to 27). There were 12 males and 18 females.

6.2 Materials

Identical to Experiment 1.

6.3 Design

The experiment employed a 2 by 2 within-participants design. The first factor, Object Type, had two levels, within vs between object. On within-object trials the two to-be-compared letters appeared on the same shape, whereas on between-object trials the two letters appeared on different shapes. There were equal numbers of each level, and these were randomly distributed across trials. The second factor, Shadow Type, had two levels: the regions either could or could not be readily interpreted as shadows. In line with Rensink and Cavanagh (2004), this was achieved by removing the objects that cast the shadows. In contrast to Experiment 1, however, the shadows were cast from an object to which they were not attached, and the light source for the scene was visible (see Figure 3). In the non-shadow-like condition, in addition to the removal of the objects responsible for casting the shadow, all other shadows in the scene were removed, to further reduce the likelihood of the shapes on which the targets were presented being interpreted as shadows. Shadow or non-shadow-like stimuli were presented in separate blocks, the order of which was counterbalanced across participants.

In the context of previous data highlighting that attention moves in perceived rather than retinal space (Robertson and Kim 1999) we were again faced with the problem of how to control for the spatial separation between the targets. In Experiment Two a more stringent criterion was adopted such that the within-object distances were longer than the between-object distances. Given the potential ambiguity over how to control for the distance in these 3-D displays, this change ensured that both in retinal and perceived spatial distance the within-object distances were in fact further apart, thus tending to reduce any object-based advantage. The vertical (within object) separation between the targets was therefore fixed at 5.75 degrees whilst the horizontal distances were 3.8 degrees at the top and 4.9 degrees at the bottom.
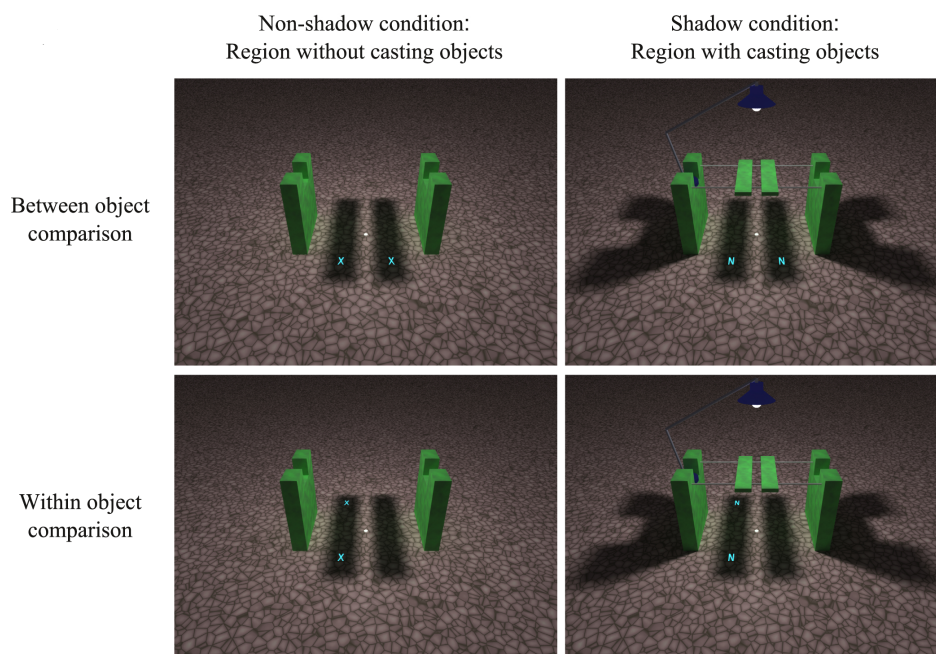

Figure 3. Illustrating the means by which the identification of the regions as shadows was manipulated using the casting objects, a light source, and other shadows in the scene.

6.4 Procedure

Each participant completed two blocks with 4 practise trials and then 192 actual trials. Each block contained equal numbers of within- and between-object trials. On each trial the participant would be presented with two letters ('X X', 'N X', or 'N N') that could either both be the same or different. The participant had to report as quickly and accurately as possible whether the letters were the same or different by pressing the 'S' key if they were the same and the 'K' key if they were different. The two letters would either both be on the same or on different shaded regions. Participants maintained fixation on a small white dot in the centre of the display. There was a 500 ms gap between trials, during which the scene remained visible.

6.5 Results

A 2 by 2 repeated measures ANOVA on the accuracy data revealed a significant effect of Object Type ($F(1,29) = 4.56$, $p = 0.041$).[1] Object Type, however, did not interact with Shadow Type ($F = 0.851$). There was also no effect of Shadow Type on the accuracy data ($F < 1$). The Reaction Time data revealed no significant effects, (Shadow Type, $F < 1$; Object Type, $F < 1$). The Shadow Type x Object Type reaction time interaction ($F (1,29) = 1.037$, $p = 0.317$) was not only non-significant but in fact showed a trend in the opposite direction to that expected if shadows were ignored. The reaction times in the within and between conditions were 609 ms and 601 ms in the non-shadow and 604 ms and 605 ms in the shadow condition; their associated standard errors were 15 ms, 15 ms, 16 ms, and 14ms, respectively.
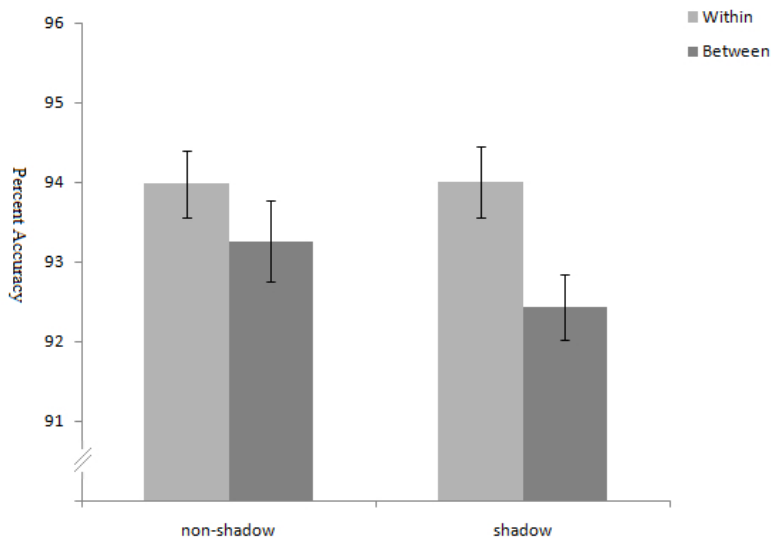
Figure 4. Accuracy data for within- and between-object target pairs presented on shaded regions with casting objects (shadows) or without casting objects (non-shadows).

---

[1] It is apparent that the magnitude of the accuracy difference reported here is smaller than that previously reported. For instance Ben-Shahar et al (2007, Experiment 1) found effect sizes in the order of 3–4%, whereas the difference across the two conditions here is only 1.15%. This difference could have been caused by three factors. First, the 'objects' used here are fuzzy and semi-transparent with the background. Second, the spatial distances within the objects were made larger than those between objects to ensure that any within-vs-between differences definitely reflected object-based effects. Third, the targets in Ben-Shahar et al were presented for very brief periods of time, followed by a mask, leading to much lower accuracies in general (86–76%, depending on their condition). This added difficulty in processing the targets may well bring into clearer light the advantages accruing to target pairings on the same objects and explain why the effect size here is so much smaller. A post-target mask was not employed here simply because it is uncertain how this might affect the interpretation of the regions as shadows.

6.6 Discussion

The representational status of shadows was explored using two object-based attention paradigms. The results provide consistent evidence that increasing the extent to which a region could be interpreted as a shadow had no effect on the extent to which that region could be selected as an 'object' of attention. This study was motivated by the mixed picture regarding the role of shadows, which are often actively utilized, in the visual system in the context of localising the objects that cast them but play a somewhat ambiguous role in influencing the recognition of the objects casting them. More specifically, this study was motivated by visual search data highlighting that access to form information is less efficient when that form can be interpreted as a shadow (Rensink and Cavanagh 2004).

At first glance the current result and that of Rensink and Cavanagh (2004) could appear in contradiction—in that whilst their result suggests that shadows have a lower representational status within the visual system, the current result suggests that the interpretation of a stimulus as a shadow does not influence the preferential allocation of attention within that region. These two findings may, however, tap very different levels of representation. It may be that shadows have to be segmented as distinct regions of space at the early level at which within-vs-between object advantages operate, in order that visual form discrimination mechanisms can then tag what should be 'discounted' from further processing.

The possibility that shadows are accessible as units of attentional selection exactly because their differential status requires that they have to be segmented as distinct areas of space is also consistent with Lovell et al's (2009) argument that shadows are not 'discounted' per se but rather that shadows are represented at distinct coarse spatial scale. Lovell et al's interpretation is based on the fact that although they could replicate Rensink and Cavanagh's finding of less efficient visual search for an 'odd one out' shadow when the difference between this and the other shadows was more subtle, they actually found more rapid visual search performance for stimuli interpretable as shadows when the difference was larger (for example, a 90 degree, rather than just a 30 degree, difference in orientation). Lovell et al argue that visual search is faster for shadows with large differences because they are rapidly identified and segmented but that less efficient search occurs for more subtle discriminations because the tagging of an area as a shadow results in it being represented in a coarse manner.

If shadows can influence attention but are represented in a distinct manner to other objects (either in terms of an active 'discounting' or being represented at a coarser spatial scale), then this clearly returns us to the question of whether it is appropriate to talk about 'object'-based attention effects. This issue was raised most clearly by Jon Driver and colleagues more than 10 years ago (Driver et al 2001) when they argued that so called object-based attention phenomena in fact reflected the influence of segmentation upon attention rather than an influence of objects per se. This distinction (between segmentation and objecthood) itself, however, raises the question of how exactly the segmentation and grouping of sensory input can be defined and understood separately from the construction and recognition of objects. Without attempting to answer this question in full, the current result does potentially offer an illustrative example regarding how this distinction could become manifest. To reiterate, whilst regions identified as shadows do not seem to have the same representational status as other objects (cf Rensink and Cavanagh 2004; Lovell et al 2009), they seem to nevertheless remain segmented as distinct regions such that within-vs-between 'object' attentional effects are still evident. A final important fact to keep in mind is that the two object-based attention paradigms used here do not cover the full range of paradigms operationalized as measures of object-based attention. Multiple object tracking is an important case in point, and it remains an open question in general whether these

different object-based phenomena actually reflect a common form of attentional selection, and whether this selection operates upon the same objects.

References

Albrecht A, List A, Robertson L, 2008 "Attentional selection and the representation of holes and objects" *Journal of Vision* **8** 13–13 ◄

Anstis S, 1990 "Imperceptible intersections: The chopstick illusion", in *AI and the Eye* Eds A Blake, T Troscianko (Chichester, UK: John Wiley & Sons) ◄

Ben-Shahar O, Scholl B J, Zucker S W, 2007 "Attention, segregation, and textons: bridging the gap between object-based attention and texton-based segregation" *Vision Research* **47** 845–860 ◄

Bonfiglioli C, Pavani F, Castiello U, 2004 "Differential effects of cast shadows on perception and action" *Perception* **33** 1291–1304 doi:10.1068/p5325 ◄

Braje W L, Legge G E, Kersten D, 2000 "Invariant recognition of natural objects in the presence of shadows" *Perception* **29** 383–398 doi:10.1068/p3051 ◄

Driver J, Baylis G C, 1989 "Movement and visual attention: The spotlight metaphor breaks down" *Journal of Experimental Psychology: Human Perception and Performance* **15** 448–456 doi:10.1037/0096-1523.15.3.448 ◄

Driver J, Davis G, Russell C, Turatto M, Freeman E, 2001 "Segmentation, attention and phenomenal visual objects" *Cognition* **80** 61–95 doi:10.1016/S0010-0277(00)00151-7 ◄

Duncan J, 1984 "Selective attention and the organization of visual information" *Journal of Experimental Psychology: General* **113** 501–517 doi:10.1037/0096-3445.113.4.501 ◄

Egly R, Driver J, Rafal R D, 1994 "Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects" *Journal of Experimental Psychology General* **123** 161–177 ◄

Kersten D, Knill D C, Mamassian P, Bulthoff I, 1996 "Illusory motion from shadows" *Nature* **379** 31–31 doi:10.1038/379031a0 ◄

Lorenceau J, Shiffrar M, 1992 "The influence of terminators on motion integration across space" *Vision Research* **32** 263–273 doi:10.1016/0042-6989(92)90137-8 ◄

Lovell P G, Gilchrist I D, Tolhurst D J, Troscianko T, 2009 "Search for gross illumination discrepancies in images of natural objects" *Journal of Vision* **9** 1–1 doi:10.1167/9.1.37 ◄

Naber M, Carlson T A, Verstraten F A J, Einhäuser W, 2011 "Perceptual benefits of objecthood" *Journal of Vision* **11** 4–4 doi:10.1167/11.4.8 ◄

Ostrovsky Y, Cavanagh P, Sinha P, 2005 "Perceiving illumination inconsistencies in scenes" *Perception* **34** 1301–1314 doi:10.1068/p5418 ◄

Ramachandran V S, 1988 "Perception of shape from shading" *Nature* **331** 163–166 doi:10.1038/3311-63a0 ◄

Rensink R A, Cavanagh P, 2004 "The influence of cast shadows on visual search" *Perception* **33** 1339–1358 doi:10.1068/p5322 ◄

Robertson L C, Kim M-S, 1999 "Effects of perceived space on spatial attention" *Psychological Science* **10** 76–79 doi:10.1111/1467-9280.00110 ◄

Scholl B, 2001 "Objects and attention: the state of the art" *Cognition* **80** 1–46 doi:10.1016/S0010-0277(00)00152-9 ◄

Scholl B, Pylyshyn Z, Feldman J, 2001 "What is a visual object? Evidence from target merging in multiple object tracking" *Cognition* **80** 159–177 doi:10.1016/S0010-0277(00)00157-8 ◄

Tadin D, Lappin J S, Blake R, Grossman E D, 2002 "What constitutes an efficient reference frame for vision?" *Nature Neuroscience* **5** 1010–1015 doi:10.1038/nn914 ◄

Tarr M J, Kersten D, Bülthoff H H, 1998 "Why the visual recognition system might encode the effects of illumination" *Vision Research* **38** 2259–2275 doi:10.1016/S0042-6989(98)00041-8 ◄

Yamaguchi M K, Kanazawa S, Shirai N, Otsuka Y, Tomonaga M, Yagi A, Imura T, 2006 "Perception of motion trajectory of object from the moving cast shadow in infants" *Vision Research* **46** 652–657 doi:10.1016/j.visres.2005.07.028 ◄

Yonas A, Granrud C E, 2006 "Infants' perception of depth from cast shadows" *Perception & Psychophysics* **68** 154–160 doi:10.3758/BF03193665 ◄