C. Kennard & R.J. Leigh (Eds.)
Progress in Brain Research, Vol. 171
ISSN 0079-6123

CHAPTER 5.5

Eye movements as a probe of attention

Albert Hoang Duc, Paul Bays and Masud Husain*

Institute of Neurology & Institute of Cognitive Neuroscience, UCL, London, UK

Abstract: Most studies of visual attention have used indirect manual reaction time measures to make inferences about where processing priority is allocated in the scene. Eye tracking offers an important opportunity to probe more directly where attention is deployed, in both health and disease. Here we provide a selective overview of oculomotor investigations which have demonstrated how attention can be captured "bottom-up" by conspicuous elements in the scene as well as how it can be allocated "top-down" on the basis of task goals and reward outcomes. We consider the concept of *salience maps* within the brain that might be responsible for computing attentional priorities and saccade targets in the visual scene, discussing the evidence for their existence. Finally, we consider how one supposedly crucial role of attention — that of binding features which belong to an object — might operate so that object representations are veridically maintained and remapped across eye movements.

Keywords: visual salience; oculomotor capture; unilateral neglect; object files

Introduction

Many of the pioneering studies in attention research have focused on visual attention. Almost invariably, early studies used manual reaction time as an index of where attention is deployed. But such measures are quite indirect. Thus, Posner's highly influential spatial cuing paradigm makes inferences which depend crucially upon differences in reaction time in response to a stimulus that appears at either a previously cued or uncued location (Posner, 1980). Tracking eye movements has the potential to provide a more direct measure of where attention is deployed since the direction of gaze is generally considered to be tightly coupled to the orienting of attention — at least, under normal circumstances (Hoffman and Subramaniam, 1995; Deubel and Schneider, 1996). Uncoupling of gaze direction and attention can, of course, occur as Posner's task clearly demonstrates. The value of eye tracking is that in natural scene viewing — where the visual environment is complex compared to many simple experimental situations — it should provide a good guide to the locus of attention. In recent years, researchers have capitalized on this possibility, seeking eventually to understand how attention and gaze are deployed to make sense of the visual world.

Oculomotor capture, attention, and reward

Visual scenes typically contain many objects that compete for the control of attention and eye movements. When an object is intentionally chosen to be the focus of attention, selection is said to occur in a voluntary, goal-directed manner. However, when

---

*Corresponding author. Tel.: +44(0)20 7679 1177;
Fax: +44(0)2079168517; E-mail: m.husain@ion.ucl.ac.uk

DOI: 10.1016/S0079-6123(08)00659-6

specific properties in the visual environment determine what is selected, regardless of the observer's goal or intention, selection is said to occur in an involuntary or stimulus-driven manner. Salient objects that appear suddenly in the visual field, for example, often capture attention even if they are irrelevant for the task at hand (Yantis and Jonides, 1984; Remington and Johnston, 1992); similarly for the oculomotor system.

Theeuwes and his colleagues used a paradigm (Fig. 1) in which observers viewed a display containing six grey circles presented on an imaginary circle around a central fixation point (Theeuwes et al., 1998; Theeuwes and Irwin, 1999). Subsequently all but one of the circles changed to red and participants were explicitly instructed to saccade towards the only grey stimulus remaining. On half the trials, an irrelevant red distractor circle was presented with abrupt onset which observers were told to ignore. When there were no such onsets, participants generated saccades that went directly towards the target. However, on distractor

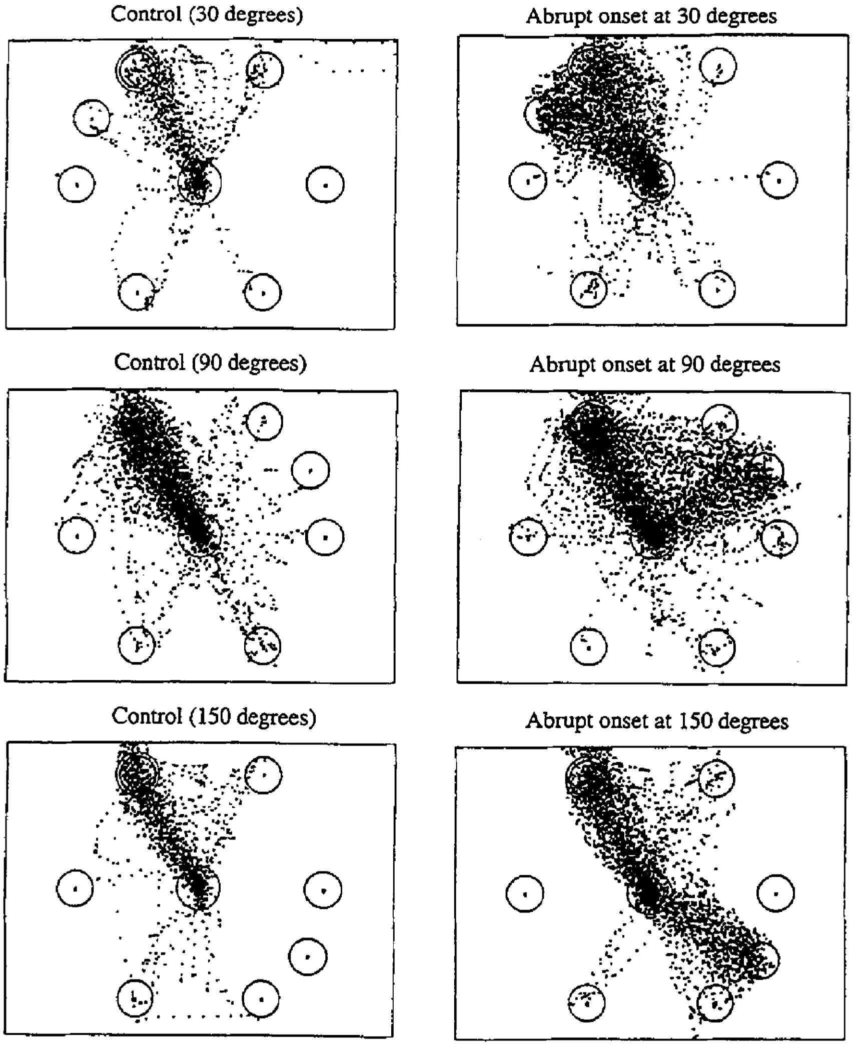

Fig. 1. Oculomotor capture. When an abrupt-onset distractor was presented (right panels), the eyes often went to it first, instead of the target. Left panels depict the control condition without an abrupt-onset distractor. (Adapted from Theeuwes et al., 1998.)

trials, the eyes went in the direction of the irrelevant abrupt-onset stimulus on about 30–40% of these trials, stopped briefly, before proceeding to the target: in other words, gaze was frequently captured by the distractor. When asked, none of the observers was aware that their eye movement behaviour was influenced by the abrupt onset, even though in many cases the eyes went in a direction completely opposite to the target. The presence of such oculomotor capture demonstrates that top-down, or goal-directed control is not always able to override bottom-up capture by abrupt-onset stimuli in the visual scene.

More recently, Milstein and Dorris have used oculomotor capture as a novel behavioural index to investigate the influence of expected value on saccadic control, along with the more conventional measure of saccadic reaction time (Milstein and Dorris, 2007). These authors wanted to investigate whether saccadic preparation toward a particular target was influenced by its expected reward value, defined as the product of reward probability and reward magnitude associated with making a saccade to that particular location. Participants were instructed to fixate centrally and then make a saccade to a red target, presented either to the left or right after a 400 ms warning period. The magnitude of the reward subjects would obtain depended on target location. The influence of expected reward value on saccadic preparation was probed occasionally (30% of trials) by flashing a green distractor during the warning period. Distractors sometimes erroneously triggered saccades to them instead of the target.

The question here was whether the frequency of such oculomotor captures was modulated by the location occupied by the distractor. Would more erroneous saccades be made to a distractor occupying a location associated with a high target value? The investigators found that the greater the reward associated with a target location, the shorter the reaction time for correct saccades. Importantly, distractors presented towards the side of higher-valued targets induced more oculomotor captures than distractors presented toward the side of lower-valued targets. Moreover, the higher the expected value of a target, the higher the proportion of oculomotor captures to distractors flashed nearby those targets. Attention to a spatial location therefore appears to be modulated by the reward associated with that position, as well as its salience.

Salience or priority maps in the brain

What about the allocation of attention in more real-world scenes? In recent years, the concept of a *salience map* has been put forward to account for how visual items are selected for attention (Itti et al., 1998; Itti and Koch, 2001). In such schemes (Koch and Ullman, 1985), visual input is first broken down into a set of topographic maps coding image dimensions such as colour, intensity, or contrast (Fig. 2). Different spatial locations then compete for saliency within each dimension map, such that only locations which locally stand out from their surroundings can persist. The maps generated for each image dimension are then combined to create a single master "saliency map," which codes for local conspicuity over the entire visual scene. Competition among locations in this map gives rise to a single winning location that corresponds to the next saccade target. This location is then suppressed, and a new winner-takes-all process selects the following target, and so on. Thus, the salience map approach can provide an explicit model that generates precise quantitative predictions about the spatial distribution of fixation locations and their sequence.

Parkhurst and his co-workers have examined the degree to which stimulus salience guides the allocation of attention in natural scene viewing (Parkhurst et al., 2002). Human eye movements were recorded while participants viewed a series of images of complex natural and artificial scenes. The correlation between stimulus salience (computed on a similar basis to Fig. 2) and fixation location was found to be significantly greater than that expected by chance alone. Moreover, attention was significantly more stimulus-driven just after stimulus onset when top-down influences are presumably weakest. The results indicate that stimulus-driven mechanisms contribute significantly to attentional guidance under natural

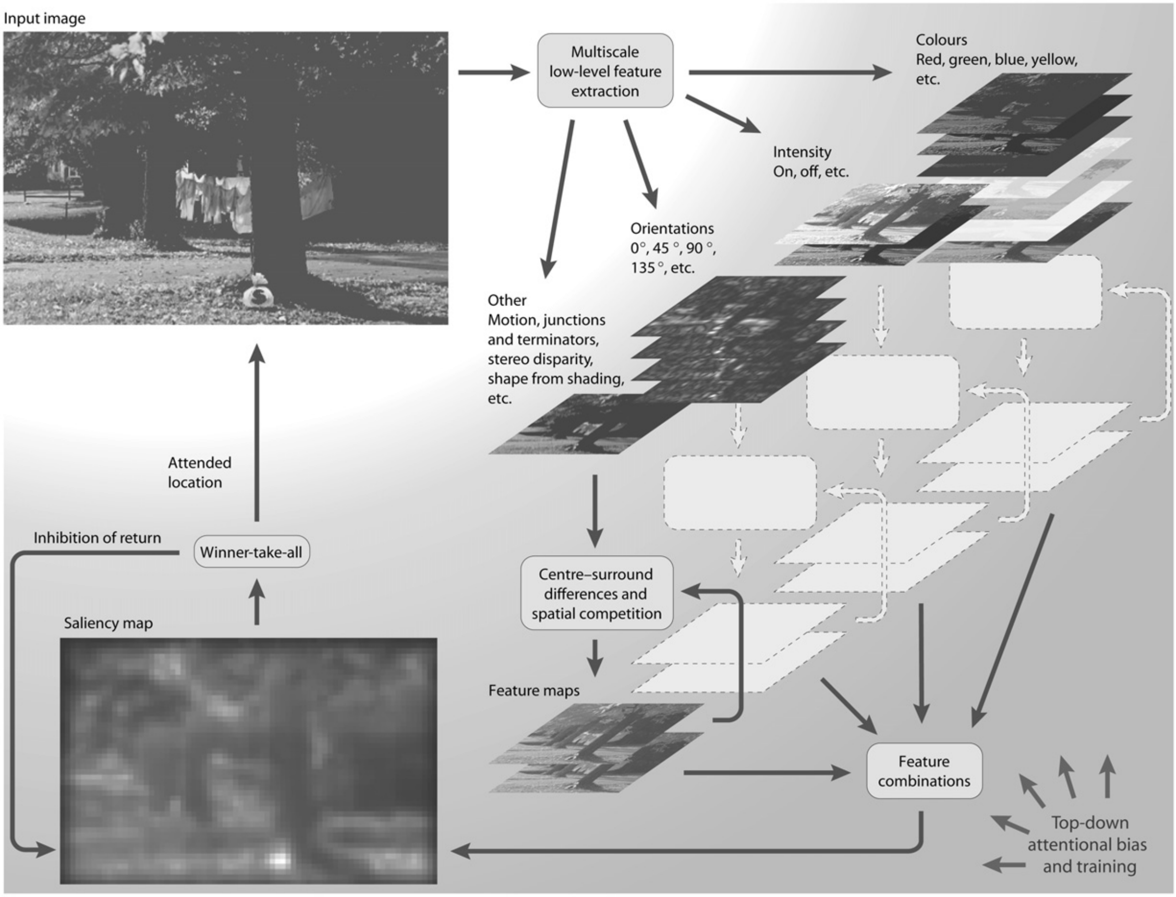

Fig. 2. Modelling the bottom-up control of attention. According to Koch and Ullman (1985), shifts of attention can be predicted by a master "salience map." The input image is decomposed into several pre-attentive feature detection mechanisms, which operate in parallel over the entire visual scene. Neurons in the feature maps then encode for spatial contrast in each of those feature channels. In addition, neurons in each feature map spatially compete for salience. After competition, the feature maps are combined into a unique saliency map, which topographically encodes for salience irrespective of the feature channel in which stimuli appeared salient. Adapted with permission from Itti and Koch (2001).

viewing conditions. However, the focus on free-viewing of static images may underestimate the importance of top-down control on eye movements, and it is unlikely that human gaze-control can be predicted solely on the basis of image characteristics (Henderson, 2007).

Although numerous models of attentional orienting have been proposed, none of these models specified the neural mechanisms or the exact underlying cortical areas responsible for encoding salience. Studies of neglect patients (Rafal, 1994; Friedrich et al., 1998; Mesulam, 1999; Husain and Rorden, 2003) suggest that posterior parietal cortex (PPC) might play a role in the ability to orient visual attention. Single neuron recording studies in monkeys have brought deeper insight into the neurophysiology of this region. Gottlieb and her colleagues hypothesized that the lateral intraparietal area (LIP) could provide a selective spatial representation of objects that are likely to attract attention, either by their salience or task relevance (Gottlieb et al., 1998; Goldberg et al., 2006). LIP is thought to participate in covert orienting of attention and in saccade planning

(Colby and Goldberg, 1999; Snyder et al., 2000; Goldberg et al., 2002) and LIP neurons have been shown to have visual on-responses to abrupt onsets in their receptive fields, as well as sustained activity preceding saccades to the receptive field (Barash et al., 1991). Results from Gottlieb's study showed that the visual representation in LIP is sparse, with only the *most salient* or behaviourally relevant objects represented (Gottlieb et al., 1998).

A different study has explored how the PPC responds to more complex visual displays (Constantinidis and Steinmetz, 2001). The underlying hypothesis was that activity of PPC neurons selectively represents the spatial location of the stimulus that attracts attention. In order to address this issue, single neurons in area 7a of the PPC were recorded in monkeys performing a spatial version of a match-to-sample task. The task required them to locate salient stimuli in multiple-stimulus displays and release a lever after a subsequent stimulus appeared at the same location. Neurons responded preferentially to the appearance of salient stimuli inside their receptive fields. The presence of multiple stimuli did not change the overall activity for the majority of neurons or the population code for the location of the salient stimulus. These results suggest that area 7a neurons represent the location of the stimulus that attracts animal's attention and could provide the spatial information required for directing attention to a salient stimulus in a complex scene. However, several other areas of the brain have also been proposed as the locus of the theoretical salience map, including the frontal eye fields (Thompson and Bichot, 2005), primary visual cortex (Li, 2002), and regions in the ventral visual pathway (Mazer and Gallant, 2003). It may be that the principle of a topographically organized map of stimulus priority is a common mechanism to be found in multiple brain regions.

Breakdown of such priority or salience maps might be one contributory factor to disorders associated with difficulty in directing attention to certain parts of space, such as in the syndrome of unilateral neglect. Patients with this disorder, most prominently associated with right PPC lesions (Mort et al., 2003), often fail to direct their attention and gaze to stimuli to the left (Mannan et al., 2005), although the degree of neglect may be modulated by the attentional demands of the task (Kaplan et al., 1991; Mannan et al., 2005). Thus, the degree of inattention is not simply for one sector of space, regardless of the contents of the visual scene. Neglect is more severe in densely-cluttered environments where distractor stimuli may serve to capture attention more frequently (Kaplan et al., 1991). Moreover, neglect patients often re-fixate items they have already fixated, sometimes erroneously considering them to be novel items that they have not previously inspected (Husain et al., 2001; Mannan et al., 2005).

Attended objects and their representations across saccades

So far we have been interested in how eye movements may be used as a surrogate marker for the locus of attention. In the real world, a key issue is how representations of objects selected for attention are maintained across the potentially disruptive influence of eye movements. Treisman and Gelade (1980) had originally proposed in their highly influential feature integration theory that a key role of spatial attention is to bind features (colour, form, etc.) together. Later, Kahneman, Treisman, and Gibbs developed a theory of *object files* — episodic representations that gather and maintain information about objects they represent *over time* (Kahneman et al., 1992). An object file contains information about features of the object, such as its colour, shape, and luminance. It integrates the features that belong to an object and therefore might be considered as a representation in which features are kept bound. Importantly, it also serves as a representation in which information is updated over time, as objects move or change. Crucially, spatial location is used to address the file (Treisman, 2006). But what happens when the eyes move? How do spatial locations of objects keep updated and how does the brain keep track of object files across saccades?

We decided to investigate this issue using the object reviewing paradigm (Kahneman et al., 1992). In this task subjects view two objects (e.g., squares) each of which has a letter within it. After a short period of viewing (the ‘‘previewing’’ phase) both letters are removed, and the objects move to new positions. A single letter is then presented in one of the two objects, and the subject’s task is to name it. Typically, this response is faster if the letter appears in the same object that it had been in, than if it appears in the other object. The benefit in response time is referred to as an *Object-Specific Preview Benefit* (OSPB), because it is derived from previewing a letter within the same object across time.

What are the effects of eye movements on object feature integration or ‘‘object files’’ created by attending to an object? We used a modified version of Kahneman’s paradigm to address this question (Kahneman et al., 1992). First, Kahneman and his colleagues, and subsequent studies on object files (Henderson and Anes, 1994; Noles et al., 2005) used letters as stimuli. However, we reasoned that it might be more appropriate to use more complex stimuli to test the fidelity of information stored in object files. Moreover, single feature objects are uncommon in real environments. Consequently, we used a combination of a colour and texture as stimuli. Second, in order to achieve a more robust OSPB, we asked participants to press a key on each trial to indicate that the final stimulus matched one of the initially presented stimuli, and to withhold any key press if not. This response, unlike voice-key naming, requires participants to attend to the initially presented stimuli, but still allows us to measure OSPBs. Third, in addition to the OSPB measure, we used a novel probe of the quality of object file representations by introducing irrelevant feature conjunctions in our identification task. We reasoned that the maintenance of faithful object files should be associated with a low number of false alarms even when irrelevant feature conjunctions are presented.

In our experiment, the initial display consisted of two empty boxes, one centred above fixation and the other below (Fig. 3). Eye position was recorded continuously. After fixation had been maintained for 1 s, a stimulus made of a colour and a texture was presented for 500 ms within each box. There was then a delay of 2 s. During this period the fixation point could either remain on at the centre (*fixation condition*) or shift to the periphery and then back centrally (*saccade condition*), with the subject in this condition having to make an outward saccade to the new peripheral location of the fixation cross and back again to the centre. Finally, a single stimulus was presented in one of the two boxes for 500 ms. The final stimulus could be identical to one of the preview stimuli and be presented in the same box (termed *match congruent* trial) or in the other box (*match incongruent* trial). Alternatively, it could be completely different with no features in common with either preview stimulus (*non-match* trial), or different but share one feature with one of the initial stimuli (*non-match conjunction* trial). The subject’s task was to judge whether the final stimulus was the same or different from the initial stimuli by pressing a key as fast as possible.

The OSPB is calculated as the difference in response times between match incongruent and match congruent trials: that is, the reaction time advantage in recognizing a final stimulus if it had been previewed *within the same object*. In this experiment we found that in the fixation condition, the OSPB was 18.9 ms, compared to 28.93 ms in the saccade condition (Fig. 4). A two-way repeated measure ANOVA revealed a main effect of congruency (congruent vs. incongruent; $p<0.01$) but no interaction with condition (saccade vs. fixation). Thus, the result here would suggest that there is a significant OSPB, which is invariant across eye movements, perhaps indicating that object-specific information is indeed maintained equally well in both fixation and saccade conditions. However, this measure alone does not interrogate the fidelity of feature binding. To do this we also need to examine accuracy across all types of trial.

While subjects made almost no errors on non-match trials, their performance dropped on *non-match conjunction* trials, where the final stimulus shared a single feature with one of the preview stimuli (81.3% accurate, SE = 3.2%). In order to quantify subjects’ ability to discriminate between non-match conjunction trials and match trials, we calculated perceptual sensitivity (d') by running

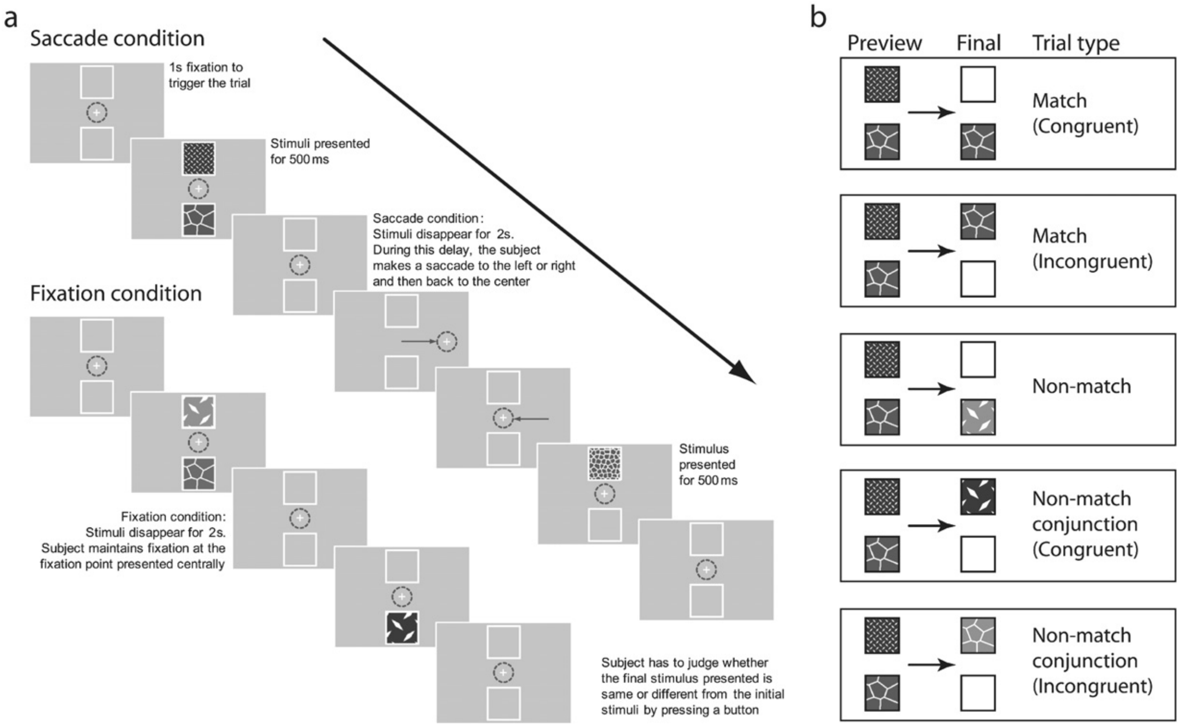

Fig. 3. An experiment to study the fidelity of object feature binding across saccades. (a) The task was to determine whether the final stimulus was same or different from the initial (preview) stimuli by pressing a key as fast as possible. The figure shows both the fixation and saccade condition. Stimuli were composed of a colour and a texture. The dashed circle shows where fixation should be maintained at each time epoch. (b) Examples of the different trial types. Subjects were instructed to press the button only on *match* trials, in which the final stimulus was identical to one of the preview stimuli.

a signal detection analysis (Fig. 4). The results revealed that participants were better in discriminating the final stimulus in the fixation condition ($d' = 2.86$, SE = 0.17) than the saccade condition ($d' = 2.58$, SE = 0.14). A two-way ANOVA showed a significant effect of condition (fixation vs. saccade; $p < 0.02$) but not congruency. Importantly, however, there was a significant interaction between condition and congruency ($p < 0.01$).

In the fixation condition, perceptual sensitivity rose from 2.70 (SE = 0.14) on incongruent trials to 2.93 (SE = 0.18) on the congruent trials. In other words, participants were better at discriminating a final stimulus when it was presented in the same box as the matching preview stimulus — analogous to the reaction time difference indexed by the OSPB. However, in the saccade condition, subjects were no better in discriminating that difference, as revealed by equal d' values for congruent and incongruent trials ($d' = 2.55$, STE = 0.15 and $d' = 2.59$, STE = 0.13, respectively; Fig. 4). Thus, the object-specific advantage of discriminating stimuli in the congruent condition (identical stimulus in same box) compared to the incongruent condition (identical stimulus but in different box) is lost with intervening saccades.

These results suggest that, although an object file is not destroyed by eye movements as demonstrated by a positive OSPB in the saccade condition, the fidelity of information carried over across eye movements is not as well maintained as when the eyes are kept static. In order to rule out the possibility that subjects had a lower perceptual sensitivity in the saccade condition due to a faster reaction time (speed-accuracy trade off), a statistical correlation between perceptual

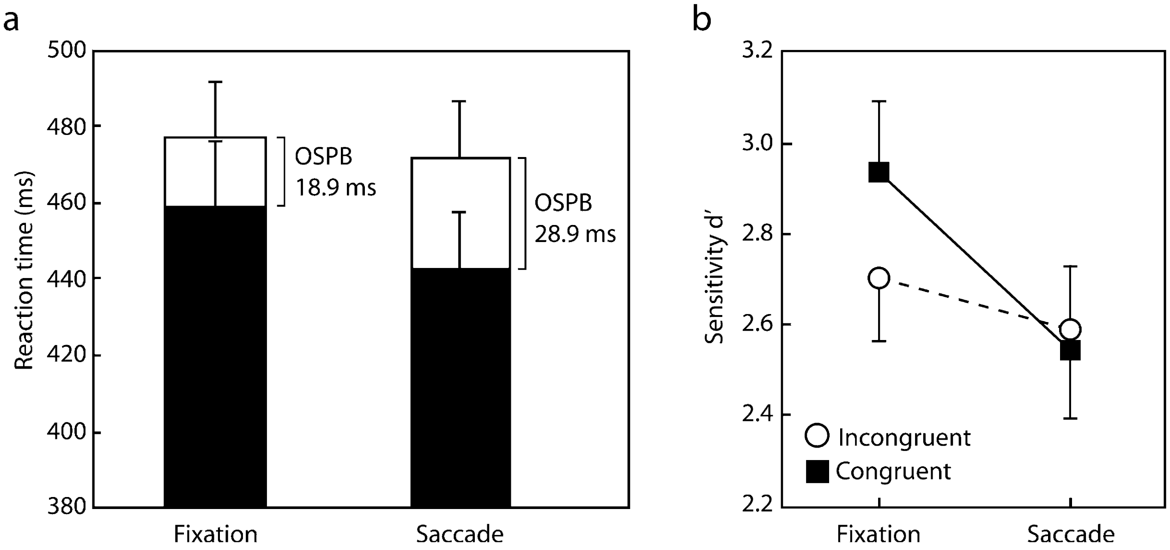

Fig. 4. Performance on congruent and incongruent trials in the fixation and saccade conditions. (a) Median reaction time on congruent (black) and incongruent (white) trials in each condition. (b) Perceptual sensitivity measured by *d'*. A value of 0 indicates inability to judge whether the probe stimulus is same or different from previously seen targets, whereas higher values mean better sensitivity. Note the discrimination advantage for congruent trials is lost when there is an intervening saccade. Error bars indicate ±1 SE.

sensitivity and reaction times was performed. This showed no significant effect ($r = 0.28$, $p > 0.1$). The present findings provide the first evidence that eye movements can interfere with the maintenance of object representations and may therefore constrain theories of feature integration across gaze shifts when viewing natural scenes.

Conclusions

The study of eye movements has much to offer for improving our understanding of visual attention. By measuring eye position we can obtain a far better index of the allocation of attention than more indirect manual reaction time measures. Eye tracking has revealed how attention can be captured by abrupt-onset stimuli or salient items in the visual scene. Such studies have also led to the concept that there might be salience maps in the brain, which guide the deployment of attention. When such representations are damaged, for example, after injury to the PPC, there may be florid inattention of one sector of space, as in the neglect syndrome. Finally, the investigation of eye movements may also constrain models of visual attention. The integration of features into object representations or files appears to be degraded to some extent by gaze shifts and this has implications for the fidelity of such representations when subjects view real-world scenes.

References

Barash, S., Bracewell, R.M., Fogassi, L., Gnadt, J.W. and Andersen, R.A. (1991) Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. J. Neurophysiol., 66(3): 1095–1108.

Colby, C. and Goldberg, M.E. (1999) Space and attention in parietal cortex. Annu. Rev. Neurosci., 22: 319–349.

Constantinidis, C. and Steinmetz, M.A. (2001) Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. Cereb. Cortex, 11: 581–591.

Deubel, H. and Schneider, W.X. (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. Vision Res., 36(12): 1827–1837.

Friedrich, F.J., Egly, R., Rafal, R.D. and Beck, D. (1998) Spatial attention deficits in humans: a comparison of superior parietal and temporal-parietal junction lesions. Neuropsychology, 12(2): 193–207.

Goldberg, M.E., Bisley, J., Powell, K.D., Gottlieb, J. and Kusunoki, M. (2002) The role of the lateral intraparietal area of the monkey in the generation of saccades and visuospatial attention. Ann. N.Y. Acad. Sci., 956: 205–215.

Goldberg, M.E., Bisley, J.W., Powell, K.D. and Gottlieb, J. (2006) Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. Prog. Brain Res., 155: 157–175.

Gottlieb, J.P., Kusunoki, M. and Goldberg, M.E. (1998) The representation of visual salience in monkey parietal cortex. Nature, 391: 481–484.

Henderson, J. (2007) Regarding scenes. Curr. Dir. Psychol. Sci., 16: 219–222.

Henderson, J. and Anes, M. (1994) Roles of object-file review and type priming in visual identification within and across eye fixations. J. Exp. Psychol. Hum. Percept. Perform., 20(4): 826–839.

Hoffman, J.E. and Subramaniam, B. (1995) The role of visual attention in saccadic eye movements. Percept. Psychophys., 57(6): 787–795.

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J. and Kennard, C. (2001) Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. Brain, 124(Pt 5): 941–952.

Husain, M. and Rorden, C. (2003) Non-spatially lateralized mechanisms in hemispatial neglect. Nat. Rev. Neurosci., 4(1): 26–36.

Itti, L. and Koch, C. (2001) Computational modelling of visual attention. Nat. Rev. Neurosci., 2(3): 194–203.

Itti, L., Koch, C. and Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell., 20: 1254–1259.

Kahneman, D., Treisman, A. and Gibbs, B. (1992) The reviewing object files: object-specific integration of information. Cogn. Psychol., 24: 175–219.

Kaplan, R.F., Verfaellie, M., Meadows, M.E., Caplan, L.R., Pessin, M.S. and DeWitt, L.D. (1991) Changing attentional demands in left hemispatial neglect. Arch. Neurol., 48(12): 1263–1266.

Koch, C. and Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol., 4(4): 219–227.

Li, Z. (2002) A saliency map in primary visual cortex. Trends Cogn. Sci., 6: 9–16.

Mannan, S.K., Mort, D.J., Hodgson, T.L., Driver, J., Kennard, C. and Husain, M. (2005) Revisiting previously searched locations in visual neglect: role of right parietal and frontal lesions in misjudging old locations as new. J. Cogn. Neurosci., 17(2): 340–354.

Mazer, J.A. and Gallant, J.L. (2003) Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron, 40(6): 1241–1250.

Mesulam, M.M. (1999) Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. Philos. Trans. R. Soc. Lond. B Biol. Sci., 354(1387): 1325–1346.

Milstein, D.M. and Dorris, M.C. (2007) The influence of expected value on saccadic preparation. J. Neurosci., 27(18): 4810–4818.

Mort, D.J., Malhotra, P., Mannan, S.K., Rorden, C., Pambakian, A., Kennard, C., et al. (2003) The anatomy of visual neglect. Brain, 126(Pt 9): 1986–1997.

Noles, N.S., Scholl, B.J. and Mitroff, S.R. (2005) The persistence of object file representations. Percept. Psychophys., 67(2): 324–334.

Parkhurst, D., Law, K. and Niebur, E. (2002) Modeling the role of salience in the allocation of overt visual attention. Vision Res., 42: 107–123.

Posner, M. (1980) Orienting of attention. Q. J. Exp. Psychol., 32: 3–25.

Rafal, R.D. (1994) Neglect. Curr. Opin. Neurobiol., 4: 231–236.

Remington, R.W. and Johnston, J.C. (1992) Involuntary attentional capture by abrupt onsets. Percept. Psychophys., 51(3): 279–290.

Snyder, L.H., Batista, A.P. and Andersen, R.A. (2000) Intention-related activity in the posterior parietal cortex: a review. Vision Res., 40: 1433–1441.

Theeuwes, J. and Irwin, D. (1999) Influence of attentional capture on oculomotor control. J. Exp. Psychol., 25(6): 1595–1608.

Theeuwes, J., Kramer, A., Hahn, S. and Irwin, D. (1998) Exogenous and endogenous control of attention: the effect of visual onsets and offsets. Psychol. Sci., 9(5): 379–486.

Thompson, K.G. and Bichot, N.P. (2005) A visual salience map in the primate frontal eye field. Prog. Brain Res., 147: 251–262.

Treisman, A. (2006) Object tokens, binding and visual memory. In: Zimmer H., Mecklinger A. and Lindenberger U. (Eds.), Handbook of binding and memory. Perspectives from cognitive neuroscience. Oxford University Press, Oxford, pp. 315–338.

Treisman, A. and Gelade, G. (1980) A feature integration theory of attention. Cogn. Psychol., 12: 97–136.

Yantis, S. and Jonides, J. (1984) Abrupt visual onsets and selective attention: evidence from visual search. J. Exp. Psychol. Hum. Percept. Perform., 10: 601–621.