

## Control of Object-based Attention in Human Cortex

John T. Serences<sup>1</sup>, Jens Schwarzbach<sup>2,3</sup>, Susan M. Courtney<sup>1,4,5</sup>, Xavier Golay<sup>4,6</sup> and Steven Yantis<sup>1,5</sup>

<sup>1</sup>Department of Psychological and Brain Sciences, Johns Hopkins University (JHU), Baltimore, MD 21218, USA,

<sup>2</sup>Department of Neurocognition, Universiteit Maastricht, The Netherlands, <sup>3</sup>F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands, <sup>4</sup>F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD 21205, USA, <sup>5</sup>Department of Neuroscience, JHU School of Medicine, Baltimore, MD 21205, USA and <sup>6</sup>Department of Radiology, JHU School of Medicine, Baltimore, MD 21205, USA

**Visual attention is a mechanism by which observers select relevant or important information from the current visual array. Previous investigations have focused primarily on the ability to select a region of space for further visual analysis. These studies have revealed a distributed frontoparietal circuit that is responsible for the control of spatial attention. However, vision must ultimately represent objects and in real scenes objects often overlap spatially; thus attention must be capable of selecting objects and their properties nonspatially. Little is known about the neural basis of object-based attentional control. In two experiments, human observers shifted attention between spatially superimposed faces and houses. Event-related functional magnetic resonance imaging (fMRI) revealed attentional modulation of activity in face- and house-selective cortical regions. Posterior parietal and frontal regions were transiently active when attention was shifted between spatially superimposed perceptual objects. The timecourse of activity provides insight into the functional role that these brain regions play in attentional control processes.**

**Keywords:** attentional control, fMRI, object-based attention, parietal cortex

### Introduction

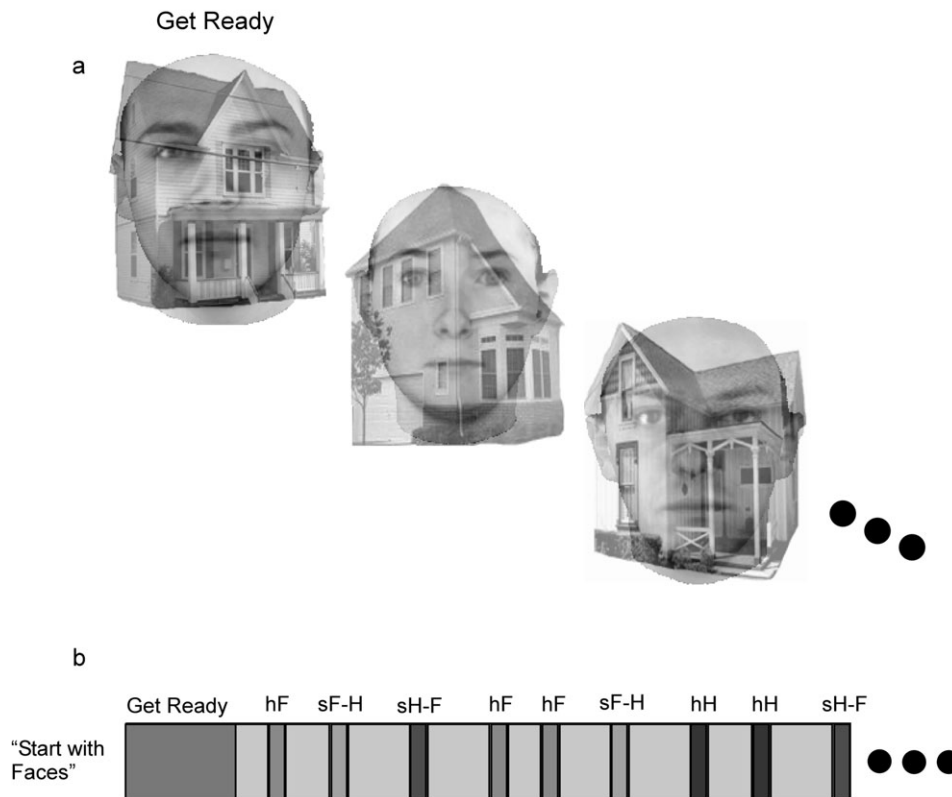
Conscious awareness of complex visual scenes is limited to only a small subset of the available image information at any one time; selective attention is the mechanism that controls access to awareness (Desimone and Duncan, 1995; Yantis, 1998). Two distinct aspects of selective attention can be distinguished: the effects of attention on the strength of early sensory representations in occipital and ventral temporal cortex and the source of attentional control signals in parietal and frontal cortex (Kanwisher and Wojciulik, 2000; Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002). Furthermore, three domains of selective attention have been documented: Visual attention can be deployed to a particular location in space (space-based attention), to visual features such as color or motion (feature-based attention), or to perceptual object representations in which all features of a segmented object are selected at once (object-based attention). The present study investigates the neural correlates of object-based attentional control.

Studies of the effects of attention have shown both behavioral facilitation (e.g. faster and more accurate responses) and enhanced cortical responses to attended locations or features (Posner, 1980; Egeth *et al.*, 1984; Moran and Desimone, 1985; O'Craven *et al.*, 1997; Chawla *et al.*, 1999; Treue and Martinez Trujillo, 1999; Saenz *et al.*, 2002). Because visual systems must

represent objects that are often spatially overlapping and partly occluded, a selective mechanism that operates on perceptual objects, and not just spatial locations, is required for natural vision. Behavioral and neurophysiological evidence has confirmed that object-based selection plays a central role in primate vision (Rock and Gutman, 1981; Duncan, 1984; Roelfsema *et al.*, 1998; O'Craven *et al.*, 1999).

Complementing these studies of the effects of attention have been investigations of the sources of attentional control signals. These studies have implicated a network of areas in posterior parietal and frontal cortex as crucial for the control of spatial attention (Mesulam, 1981; Posner *et al.*, 1984; Wojciulik and Kanwisher, 1999; Corbetta *et al.*, 2000; Hopfinger *et al.*, 2000; Beauchamp *et al.*, 2001; Vandenberghe *et al.*, 2001; Yantis *et al.*, 2002; Bisley and Goldberg, 2003) and feature based attention (Le *et al.*, 1998; Shulman *et al.*, 2002; Giesbrecht *et al.*, 2003; Liu *et al.*, 2003). However, little is known about the control of object-based attention. Although two studies have investigated how objects in the scene can guide the selection of a region of space (Fink *et al.*, 1997; Arrington *et al.*, 2000), there has been no investigation of the control of object-based attention using, for example, spatially superimposed objects (as in Fig. 1) where spatial selection is not possible (Duncan, 1984; O'Craven *et al.*, 1999). This type of selection is distinct from non-spatial feature-based attention examined in previous work (Le *et al.*, 1998; Liu *et al.*, 2003), where shifts of attention between feature dimensions (e.g. color and motion), not segmented object representations, were examined. This difference is important because in object-based attention, all the attributes of an attended object are bound together into a unitary representation (O'Craven *et al.*, 1999).

Here we used blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI) to investigate the neural mechanisms of object-based attentional control. Face and house stimuli, which maximally activate distinct anatomical foci in ventral visual cortex, permitted us to dissociate areas whose activity reflect the effects of attentional deployment from areas that may serve as the source of object-based attentional control signals. The stimuli were spatially superimposed to ensure that subjects were directing attention to objects and not to locations, an approach successfully employed previously to reveal the effects of object-based attention (O'Craven *et al.*, 1999). Observers were required to attend one of two spatially superimposed sequential streams changing at a rate of 1/s for occasional house or face targets. The face and house stimuli were morphed over time to maintain spatio-



**Figure 1.** Behavioral task used in the study. (a) Examples of superimposed face/house pairs (b) The timing of events at the beginning of a typical sequence. The subject first heard a verbal command to start the run by attending to either houses or faces. Each run began with the presentation of a face/house pair together with the words 'Get Ready' for 8 s. Each face spatially morphed into the next face and (simultaneously) each house morphed into the next house, at a rate of 1 morph/s (the morphing lasted for 240 ms, followed by a static face-house pair for 760 ms). Subjects were instructed to maintain attention on the currently attended object stream until they detected a switch target. hF, face hold target; hH, house hold target; sF-H, shift face-to-house target; sH-F, shift house-to-face target; light gray patches, intervals of time during which the subject was attending to faces or houses and nontarget stimuli were present. The interval between targets was randomly jittered (3, 4 or 5 s). Face images developed by the Max-Planck Institute for Biological Cybernetics in Tübingen, Germany.

temporal continuity within each stream and thereby to ensure that the different faces and houses were perceived as a single continuous (though changing) perceptual object throughout the duration of the run (Kahneman and Henik, 1981; Yantis, 1992; Yantis and Gibson, 1994).

Embedded within the morphing stimulus stream were targets that directed the subject to shift attention between the face and house stream or to maintain attention on the currently attended stream. A cortical area that is selective for a particular object type (e.g. faces) should become selectively more active when attention is directed to the preferred object stream. In contrast, an area controlling shifts of attention between objects should yield a pattern of increasing activity following a shift of attention, regardless of the direction of the shift (Yantis *et al.*, 2002).

## Experiment 1

### Methods

#### Subjects

Fifteen neurologically intact young adults (10 females), age 23–33 years, gave written informed consent to participate in the study, which was approved by the Johns Hopkins University institutional review boards. Data from one of the subjects were discarded because of abnormally low accuracy on the behavioral task during scanning (>2 SD from the group mean).

#### Behavioral Tasks

We first defined cortical regions of interest (ROIs) exhibiting greater activity to pictures of faces than places and vice versa (Kanwisher *et al.*, 1997; Aguirre *et al.*, 1998; Epstein *et al.*, 1999). In this localizer task, subjects passively viewed pictures of faces, places, and common objects in 30s blocks, interleaved with 20s of fixation (Kanwisher *et al.*, 1997; Epstein *et al.*, 1999). The stimulation procedures approximated those of Kanwisher *et al.* (1997) except that each stimulus was presented for 750 ms followed by a blank interstimulus interval of 250 ms and there was an 8 s ready period at the beginning of each run.

Next, the same set of subjects completed eight event-related fMRI runs while performing the attention-switching task (Fig. 1a,b). Before scanning began, subjects memorized two face targets and two house targets and then completed a series of practice runs. The specific target items differed from one subject to the next. One face and one house were designated 'hold' targets, and the other face and house were designated 'shift' targets. Each run began with a superimposed face and house in the center of the screen, and the participant was instructed to begin the run by attending to either the faces or the houses. When the run began, each face spatially morphed into another face and each house spatially morphed into another house. Each morph lasted for 240 ms, followed by the static superimposed stimulus pair for 760 ms, yielding one morph per second (spatial morphing was implemented using MorphMan 2000, Stoik Software). The order of target events was randomized within an experimental run and targets were separated in time by 3, 4 or 5 s (equally often) to introduce temporal jittering (Burock *et al.*, 1998). In each of the eight experimental runs, there were nine instances of each target type (targets that were missed by subjects were not included in the analysis). Targets never appeared in the unattended object stream.

The subjects were to press buttons held in both hands whenever any shift or hold target was detected. 'Hold' targets required that attention remain on the currently attended object stream (e.g. faces); 'shift' targets required that attention be shifted to the other stream (e.g. from faces to houses). No fixation point was provided for the subjects, nor were any instructions given to maintain a specific gaze position. We chose to allow free viewing because under most circumstances, the locus of eye-gaze and the locus of spatial attention are yoked; therefore, we could observe changes in the locus of spatial attention indirectly by monitoring eye-gaze position (see Methods and Results concerning eye-tracking below).

The 15 face stimuli were selected from the face database developed by the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. The 15 house stimuli were randomly chosen photographs of single-family homes. All stimuli subtended a visual angle of  $\sim 5^\circ$  at a viewing distance of  $\sim 65$  cm and were grayscale bitmaps. Each stimulus from a set could spatially morph into any of the other 14 stimuli from that set.

#### Data Acquisition and Analysis

MRI scanning was carried out with a Philips Gyroscan 1.5 T scanner in the F.M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute, Baltimore, MD. Anatomical images were acquired using an MP-RAGE  $T_1$ -weighted sequence that yielded images with a 1 mm isotropic voxel resolution ( $T_R = 8.1$  ms,  $T_E = 3.7$  ms, flip angle =  $8^\circ$ , time between inversions = 3 s, inversion time = 748 ms). Whole brain echoplanar functional images (EPI) were acquired with a Philips quadrature head coil in 21 transverse slices ( $T_R = 2000$  ms,  $T_E = 35$  ms, flip angle =  $90^\circ$ , matrix =  $64 \times 64$ , FOV = 240 mm, slice thickness = 6 mm, no gap). The same EPI parameters were used for both the localizer and attention scans.

BrainVoyager software (Brain Innovation, Maastricht, The Netherlands) was used for the fMRI analyses. Images from each data collection run were slice-time and motion corrected before high pass temporal frequency filtering was applied to the functional time series to remove components that occurred three or fewer times per run. No spatial smoothing was performed on the functional data. The images were then transformed into Talairach space (Talairach and Tournoux, 1988) and resampled into 3 mm isotropic voxels. To correct for motion across the scans, each EPI volume for a subject was co-registered to the first volume in the fourth (middle) functional scan acquired for that subject.

Functional localizer data were collected in three runs that consisted of 164 time points each. A group random effects general linear model (GLM) was performed with regressors specified for each of the three stimulus types (faces, places and common objects). Each regressor consisted of a boxcar model of each respective stimulation epoch convolved with a gamma function ( $\delta = 2.5$  s,  $\tau = 1.25$  s; Boynton *et al.*, 1996). The resulting regression vector was cross correlated with the BOLD time series, yielding scalar beta weights corresponding to the relative changes in signal strength associated with that particular stimulus type. ROIs responding more strongly to faces were defined by contrasting the beta weights for the face stimuli with the weighted sum of the beta weights for the common objects and places. ROIs responding more strongly to places were defined by contrasting the beta weights for the place stimuli with the weighted sum of the beta weights for the face and the common object stimuli (Kanwisher *et al.*, 1997; Epstein *et al.*, 1999). All ROIs were defined based on a threshold of  $t(13) = 3.0$ ,  $P < 0.01$  for individual voxels.

Data from the attention task were collected in eight runs conducted in a single session, with 86 time points in each run. To evaluate the BOLD response to each of the four event types (face hold, shift house-to-face, hold house, shift face-to-house), we created separate boxcar regressors for each of the 12 time points following each stimulus type and estimated the least squares fit at each time point; thus the magnitude of the beta weight associated with each time point reflects the relative change in the BOLD signal at that time point following each event (B.D. Ward, <http://afni.nimh.nih.gov/afni/docpdf/3ddeconvolve.pdf>). This approach has the advantage that the data can determine (within limits) the functional form of the estimated hemodynamic response function (HRF); no a priori assumptions about the

shape and parameters of the HRF are required. We then pooled the estimated beta weights across stimulus identity (face, house) and performed a repeated measures t-test comparing the BOLD response to shift and hold events collapsed across timepoints 2-5 (4-10s) post-stimulus. This t-test identified voxels exhibiting a main effect of shift versus hold events on the BOLD response over this time period. The single voxel threshold in the group data was set at  $t(13) = 3.0$ ,  $P < 0.01$ . A minimum cluster size of 590 ml (7 voxels in the original acquisition space) was adopted to correct for multiple comparisons, yielding a corrected statistical threshold of  $P < 0.004$  [determined using the program AlphaSim (B.D. Ward, <http://afni.nimh.nih.gov/afni/docpdf/ALPHASim.pdf>) which was used to run 2000 Monte Carlo simulations that took into account the entire EPI acquisition matrix].

We next computed event-related averages (e.g. Figs 2e-g, 3g-j and 4c,d) of the BOLD signal from each activated cluster by defining a temporal window around each target event (extending from 6 s before to 16 s after event onset) and averaging the BOLD signal within the window for all events of that type across all subjects. The baseline, or 0% signal change, for the event-related average plots was defined as the average BOLD signal during the 6 s preceding each target event. Because the 6 s preceding each event could contain both shift and hold events, negative deflections in the event-related averages following hold (or shift) events should be interpreted as relative decreases in activity, not as absolute inhibition. Error bars on all event-related average plots represent the between subject error at each timepoint.

After the localizer ROIs and the activation clusters from the attention task were determined using the preceding methods, all further analyses were carried out on the event related averages extracted from these regions. We analyze and present the event-related averages because this permits us to show activation levels before time 0. This is crucial to achieve a complete understanding of the theoretically important characteristics of the timecourses extracted from the medial and lateral fusiform regions depicted in Figure 2. The event-related averages were subjected to a three-way repeated-measures analysis of variance (ANOVA) with time (timepoints 4-10 s post-stimulus), stimulus identity (face, house) and target type (shift, hold) as factors. For the purposes of the ANOVA, shift house-to-face targets and hold face targets were considered 'face targets' and shift face-to-house targets and hold house targets were considered 'house' targets.

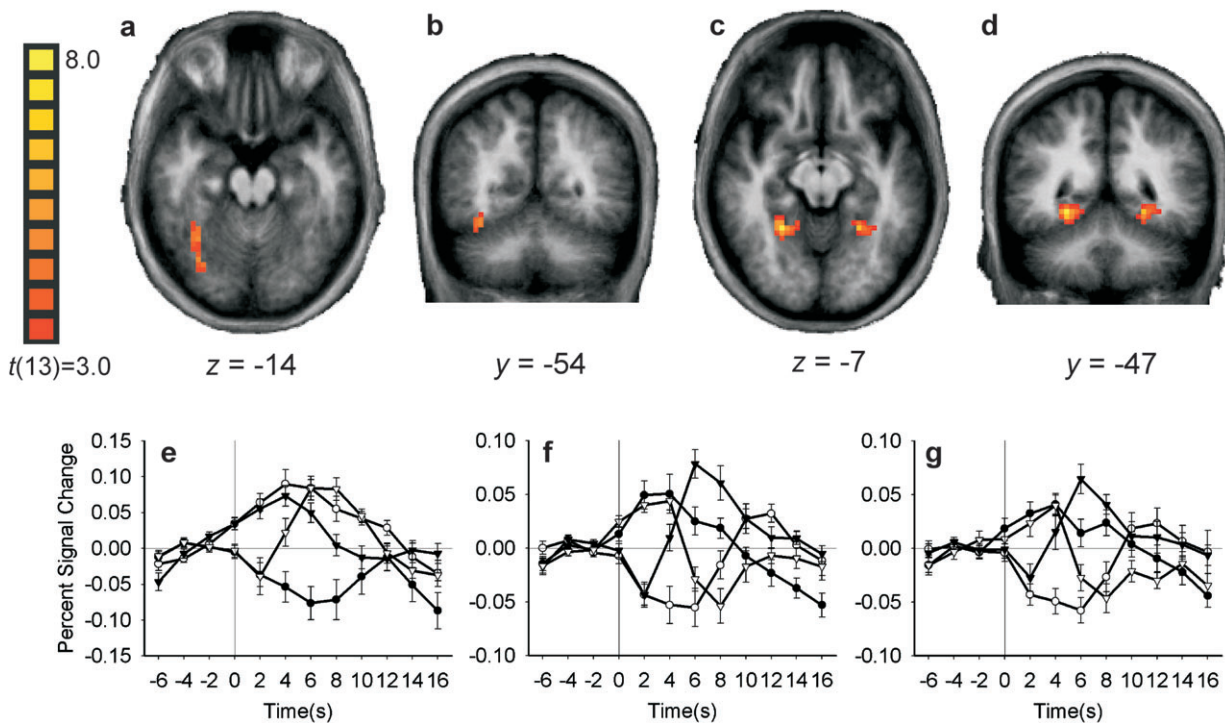
#### Eye Position Monitoring

To evaluate the possibility that switch-related activity was related to overt or covert shifts of spatial attention (rather than shifts of object-based attention), six of the subjects completed eight runs of the attention switching task outside of the scanner while their eye position was monitored at 250 Hz with an EyeLink video-based eye tracking system (SMI, Teltow, Germany). An eye movement was registered if the velocity of the eyes exceeded  $30^\circ/\text{s}$  or the acceleration of the eyes exceeded  $9600^\circ/\text{s}^2$  for  $>8$  ms and the total distance traversed was  $>0.25^\circ$  (excluding blinks) during a 1.5 s temporal window following the onset of each target stimulus. The end of an eye-movement was registered when the velocity and acceleration fell below threshold for at least five contiguous samples (20 ms). In addition to computing the number of eye-movements following each target type, we also calculated the mean Cartesian coordinates of eye-gaze during temporal epochs in which the subject was attending to faces or houses; these epochs were defined as all timepoints during which the subject was attending to a particular stimulus-type, not just the 1.5 s following the presentation of a target. Paired t-tests were then used to compare the mean X and Y eye-gaze position while subjects were attending to faces and houses.

## Results

#### Behavioral Data

A two-way repeated measures ANOVA (stimulus identity  $\times$  target type) revealed no significant differences in the behavioral detection rates for the shift and hold targets [percentage detected  $\pm$  SE:  $94 \pm 1.9$ ,  $91 \pm 2.4$ ,  $87 \pm 3.5$  and  $94 \pm 1.8$  for hold



**Figure 2.** Object-based attentional modulation in ventral extrastriate cortex. All brain images in this and subsequent figures are shown right side on left and depict group activations overlaid on an average of the individual subjects Talairach transformed brains. (a, b) Region of right LatFus gyrus that exhibited an increased BOLD response when subjects viewed pictures of faces. (c, d) Bilateral regions of MedFus that exhibited an increased BOLD response when subjects viewed pictures of places. (e–g) Group event related averages from the attention task computed from the right LatFus, right MedFus and left MedFus, respectively (see Methods). Open circle, hold face; filled circle, hold house; filled triangle, shift face-to-house; open triangle, shift house-to-face. Error bars are  $\pm$ SEM.

house, hold face, shift face-to-house, and shift house-to-face events, respectively,  $F(1,13) = 2.9$ ,  $P > 0.10$ ] or in response times [in ms  $\pm$  SE:  $681 \pm 22.5$ ,  $735 \pm 15.9$ ,  $753 \pm 20$  and  $701 \pm 28$  for hold house, hold face, shift face-to-house and shift house-to-face events, respectively,  $F(1,13) = 2.1$ ,  $P > 0.15$ ]. Note that we did not emphasize response speed, so these response times should be interpreted with caution.

#### Attention Effects in Localizer ROIs

As shown in previous studies, a region of right lateral fusiform gyrus (LatFus, Fig. 2a,b) was more active when subjects viewed faces than when they viewed objects or places and bilateral regions of the medial fusiform gyrus (MedFus) and the parahippocampal gyrus (Fig. 2c,d) were more active when they viewed places than when they viewed faces or objects (Kanwisher *et al.*, 1997; Aguirre *et al.*, 1998; Epstein *et al.*, 1999). In the following text, we sometimes refer to these regions as being ‘face selective’ or ‘house selective’; by this we mean only that they respond more strongly to faces or to houses, respectively. In fact, these regions respond in a characteristic way to many classes of visual stimuli (Gauthier *et al.*, 1999; Haxby *et al.*, 2001).

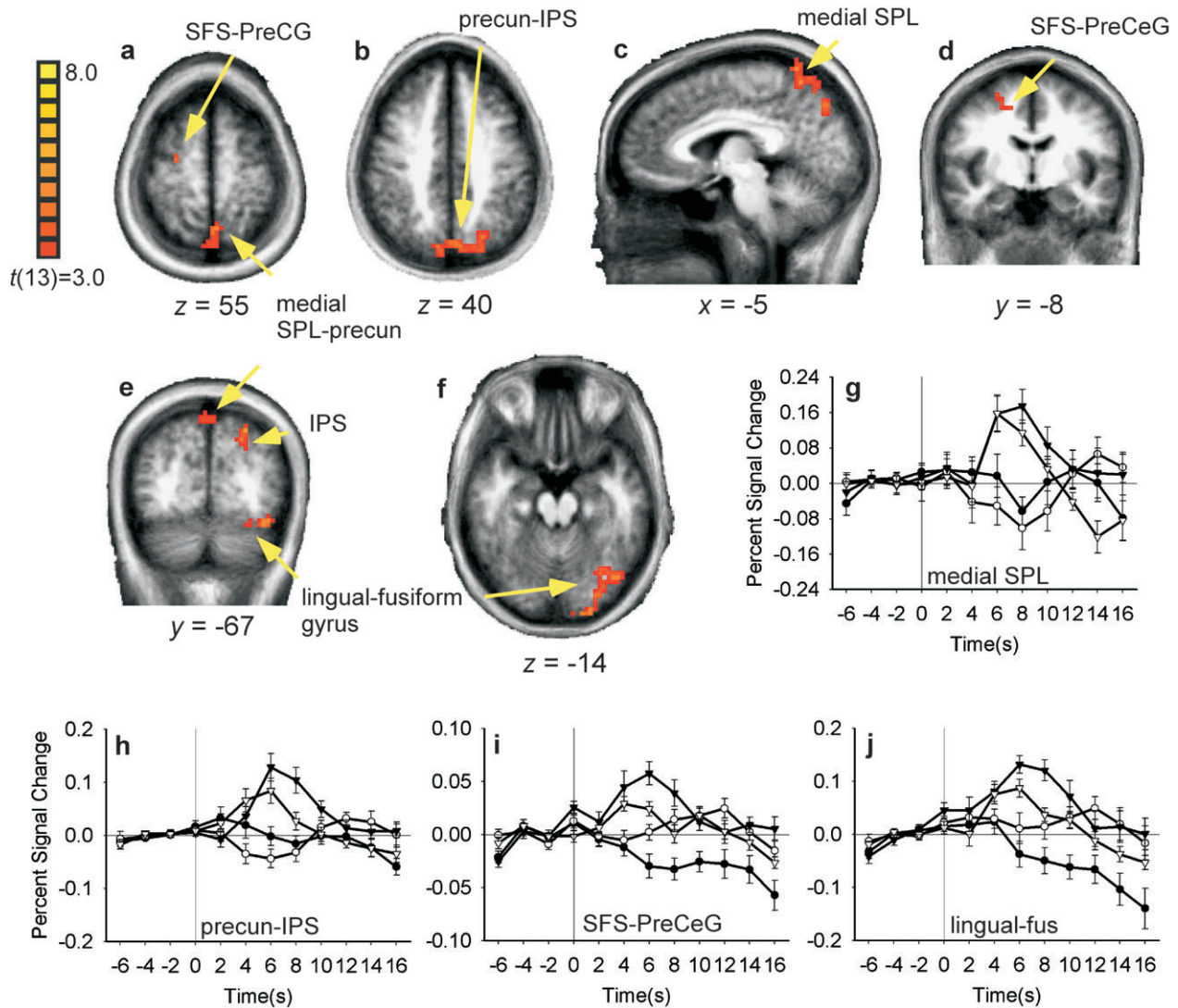
Figure 2e–g show the group mean event-related timecourses of the BOLD signal in the localizer-defined right LatFus and the bilateral MedFus for the four event types in the attention task (hold house, hold face, shift face-to-house and shift house-to-face). A repeated measures ANOVA (time points 4–10 s  $\times$  stimulus identity  $\times$  target type) was used to evaluate the event-related timecourses from each area (see Table 1).

Figure 2e shows that activity in right LatFus was greater when attention was maintained on faces or shifted to faces

(open circles and triangles, respectively) than maintained on houses or shifted to houses (closed circles and triangles, respectively) during the attention task [ $F(1,13) = 20.6$ ,  $P < 0.001$ ]. Figure 2f,g show that activity in the right and the left MedFus regions exhibited the complementary pattern [right MedFus,  $F(1,13) = 14.5$ ,  $P < 0.005$ ; left MedFus,  $F(1,13) = 31.9$ ,  $P < 0.001$ ]. All three of these areas also exhibited an interaction of stimulus identity and time; the preferred stimulus identity came to dominate the BOLD response in these regions as attention was sustained on that stimulus (see Table 1).

The event related average plots for the face and house hold events (open and closed circles, respectively) indicate that the BOLD signal following these events was already different at the moment the events occurred (i.e. time zero in the figure). This is because subjects were already attending either to houses or to faces at time zero. For instance, a face hold target was necessarily preceded by a shift target that instructed subjects to direct attention to the face object stream. Paired one-tailed  $t$ -tests on the timecourse data at time 0 confirmed the early onset of the attention effect in these areas [right LatFus,  $t(13) = 2.67$ ,  $P < 0.01$ ; right MedFus,  $t(13) = -1.8$ ,  $P < 0.05$ ; left MedFus,  $t(13) = -1.9$ ,  $P < 0.05$ ].

The BOLD timecourse following shift targets reveals a markedly different temporal pattern: in right LatFus, for example, a shift from (nonpreferred) houses to (preferred) faces was associated with a relative increase in the BOLD signal and a shift from faces to houses was associated with a relative decrease in the BOLD signal (Fig. 2e, open and closed triangles, respectively). A complementary pattern of activation was observed in right and left MedFus (Fig. 2f,g). The crossover pattern following shift events and the sustained activity following hold



**Figure 3.** Regions exhibiting a greater BOLD response to shift as compared to hold targets (shown right on left). (a) Axial slice showing right SFS–PreCeG and medial SPL–precuneus. (b) Bilateral precuneus activation that extended into the left IPS. (c) Sagittal slice showing medial SPL–precuneus activation and the more ventral bilateral precuneus activation shown in panel b. (d) Coronal slice showing the activation in right SFS–PreCeG. (e) Coronal slice showing medial SPL–precuneus and the portion of the bilateral precuneus activation that extended into left IPS. (f) Axial slice showing the activation in the left lingual–fusiform gyrus. (g–j) Group event related averages from the medial SPL–precuneus, bilateral precuneus–IPS, SFS–PreCeG and the lingual–fusiform gyrus, respectively. Open circle, hold face; filled circle, hold house; filled triangle, shift face-to-house; open triangle, shift house-to-face. Error bars are  $\pm$ SEM.

events corroborates previous reports (O’Craven *et al.*, 1999) that the deployment of attention to one of two spatially superimposed objects can significantly modulate the strength of the sensory representation in areas that respond more strongly to different object types.

#### Shift-related Activity

Figure 3a–f shows cortical regions exhibiting a stronger BOLD response following shift than following hold targets (Table 2, top). In frontal cortex, shift-related activity was observed near the junction of the right superior frontal sulcus (SFS) and the precentral gyrus (PreCeG, Fig. 3a,d), near the putative human homologue of the frontal eye field. In parietal cortex, two major activation foci were observed: the first was in the medial aspect of the superior parietal lobule (SPL, including the precuneus, Fig. 3a,c,e) and a second more ventral activation that extended bilaterally through precuneus cortex into the left intraparietal sulcus (IPS, Fig. 3b,c,e). In addition, an area of

the left occipital pole (not shown, see Table 2) and an activation in the left lingual and fusiform gyri exhibited a main effect of shifting attention (Fig. 3f).

Figure 3g–j shows the group mean event-related BOLD time-course from the medial SPL (Fig. 3g), precuneus–IPS (Fig. 3h), right SFS–PreCeG (Fig. 3i) and the left lingual-fusiform gyri (Fig. 3j). The timecourse of the BOLD signal in these regions contrasts sharply with the timecourse observed in LatFus and MedFus cortex (Fig. 2e–g). First, there is a marked increase in activity following shift events and little change in activity following hold events. Secondly, there was no difference between the hold face and hold house timecourses at time zero in any of these regions. This key result shows that this shift-related activity is transient and not sustained.

Figure 4 shows areas exhibiting a greater BOLD response following hold as compared to shift events: a region of left superior frontal gyrus (SFG, Fig. 4a,c) and a region near the left intraparietal sulcus (IPS, Fig. 4b,d).



ANOVAs on the event-related averages revealed that none of the areas exhibiting a main effect of shift versus hold showed a main effect of stimulus identity. However, some regions exhibited a larger difference in the BOLD response to shift face-to-house and hold house events than to shift house-to-face and hold face events. This difference was manifested in an interaction between stimulus type and stimulus identity (see Table 2) and is most graphically exemplified in Figure 4c,d.

### Eye Movements

A two-way ANOVA (stimulus identity  $\times$  target type) revealed no differences in the mean number of eye movements occurring in a 1.5 s window following the onset of shift and hold targets, [mean number of eye movements  $\pm$  SE:  $1.1 \pm 0.1$ ,  $1.0 \pm 0.17$ ,  $1.1 \pm 0.17$  and  $1.2 \pm 0.18$  for hold house, hold face, shift face-to-house and shift house-to-face events, respectively,  $F(1,5) = 2.7, P > 0.16$ ]. The main effect of stimulus identity and the interaction term also failed to reach significance ( $P_s > 0.39$ ). As an additional check, we examined the number of eye movements made during the temporal interval when the target stimuli were physically present on the display screen (1 s): no significant differences were observed (main effect of shift versus

hold,  $F(1,5) = 0.04, P > 0.84$ ). Finally, we examined the spatial distribution of the subject's gaze position during epochs of attention to faces and houses; paired  $t$ -tests were used to compare the mean Cartesian coordinates of eye gaze and no differences were found [ $t(5) = 0.83, P > 0.4$ ,  $t(5) = 1.75, P > 0.14$  for the comparisons of  $X$  and  $Y$  mean eye position, respectively].

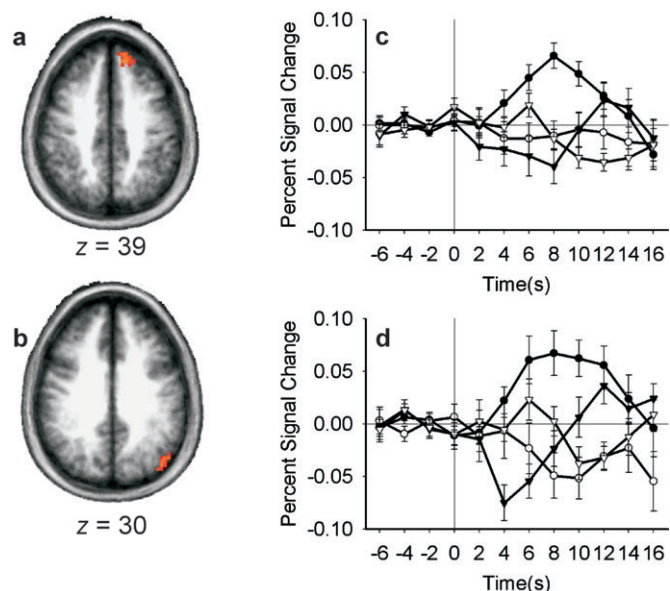
### Discussion

An instruction to shift attention led to systematic modulation of activity in ventral visual areas that respond differentially to face and house stimuli, corroborating a previous report (O'Craven *et al.*, 1999). In addition, regions of posterior parietal and dorsal frontal cortex are transiently more active when attention is shifted between segmented perceptual objects, suggesting a role for these regions in object-based attentional control.

We also observed shift-related activity in occipital and ventral temporal visual regions (e.g. Fig. 3f) that are generally believed to be the recipients of attentional biasing signals (Desimone and Duncan, 1995; Kastner and Ungerleider, 2000) and not the source of such signals. We speculate that accompanying each attention switch was a brief interval of object segmentation during which the motion of low-level visual features caused by the spatial morphing of the to-be-attended object stream may have been particularly salient. This may in turn have evoked heightened responses in early sensory areas following switch signals compared to hold signals. Such responses to low-level features would not be expected to occur in parietal and frontal regions. Further measurements will be required to test this possibility.

Finally, the difference between the BOLD response to shift and hold targets was larger in some regions when attention was switched to houses compared to faces (see Table 2). In addition, two areas were identified that responded strongly to hold house stimuli, leading to a main effect of hold greater than shift (Fig. 4). While hold house targets evoked the strongest responses in these regions, the response profiles were at least nominally similar to the patterns observed in MedFus (e.g. Fig. 2f,g). That is, there was a stronger response to hold house targets, and an increasing response to shift face-to-house targets (with the complementary pattern observed for face hold and shift stimuli). These findings echo previous reports showing that house stimuli evoke a strong response in parietal and superior frontal regions (Ishai *et al.*, 2000; Sala *et al.*, 2003).

An alternative account of the present results might assert that the switch-related activity observed in frontal and parietal areas is a manifestation of spatial shifts of attention to different regions of the face and the house stimuli, rather than to



**Figure 4.** Regions exhibiting a greater BOLD response to hold as compared to shift targets [shown right on left,  $t(13) = 3.0$ – $8.0$ ]. (a) Axial slice showing left SFG activation. (b) Axial slice showing left IPS activation. (c, d) Group event related averages from the left SFG and the left IPS, respectively. Open circle, hold face; filled circle, hold house; filled triangle, shift face-to-house; open triangle, shift house-to-face. Error bars are  $\pm$ SEM.

**Table 1**

Brain regions identified in the independent functional localizers as responding more strongly to pictures of faces or places, respectively

Localizer ROIs	(x, y, z)	Volume (ml)	Diff. at time 0, $t(13)$ , one-tailed	Stim. identity, $F(1,13)$	Time $\times$ identity, $F(3,39)$
Right LatFus	33, -57, -14	1.16	2.67, $P < 0.01^*$	20.6, $P < 0.001^*$	6.5, $P < 0.001^*$
Right MedFus	522, -45, -7	1.81	-1.8, $P < 0.05^*$	14.5, $P < 0.002^*$	17.3, $P < 0.001^*$
Left MedFus	-22, -45, -5	0.648	-1.9, $P < 0.05^*$	31.9, $P < 0.001^*$	8.6, $P < 0.001^*$

Coordinates from Talairach and Tournoux (1988). Statistical tests were performed on the event-related averages from each of these regions during the attention switching task (see Methods).

**Table 2**

Brain areas exhibiting a main effect of switching attention versus holding attention

	(x, y, z)	Volume (ml)	t(13)	Stim. type × stim. ID, F(1,13)
<b>Shift &gt; hold</b>				
Right SFS–PreCeG	19, -7, 49	0.594	3.52	0.09, $P > 0.96$
Medial SPL–precun.	0, -59, 56	1.54	3.45	0.34, $P > 0.57$
Precun.–left IPS	-9, -71, 40	2.56	3.53	3.9, $P < 0.05^*$
Left Occ. pole	-9, -89, -5	1.10	4.0	6.4, $P < 0.05^*$
Left Ling–Fus	-29, -70, -15	1.67	3.64	2.6, $P > 0.13$
<b>Hold &gt; shift</b>				
Left SFG	-10, 40, 39	0.864	-3.64	10.1, $P < 0.01^*$
Left IPS	-41, -70, 28	0.675	-3.59	6.8, $P < 0.05^*$

Abbreviations: SPL, superior parietal lobule; IPS, intraparietal sulcus; Occ, occipital; Ling, lingual; Fus, fusiform; PreCeG, pre-central gyrus; Precun, precuneus; SFG, superior frontal gyrus. Coordinates from Talairach and Tournoux (1988). Average  $t$ - and  $F$ -values computed across all voxels in activated clusters. Areas above the horizontal line appear in Figure 3 and those below the line appear in Figure 4.

nonspatial object-based shifts of attention. For instance, subjects may have been attending to the ‘roof-line’ of the house stimuli and to the eyes–nose–mouth region of the face stimuli, and this would have led to a discrete shift of spatial attention following each shift target. Such a shift would be expected to activate the very regions observed here (i.e. posterior parietal and dorsal prefrontal cortices).

To evaluate this hypothesis, six of the original subjects completed eight runs of the attention switching task outside of the scanner and while their eye position was monitored. During free viewing, the locus of spatial attention is closely yoked to the locus of eye position (Hoffman and Subramaniam, 1995). Therefore, in free viewing, one can effectively monitor the locus of spatial attention by monitoring the position of overt eye movements. Subjects did make eye movements during the task; however, there was no significant difference in the mean number of eye movements made in the 1.5 s following shift and hold targets. In addition, no systematic differences were found in the Cartesian coordinates of gaze position during epochs of attention to faces and houses. Based on these results, we conclude that changes in the spatial extent of gaze position (and by implication, spatial attention) cannot account for the observed switch related activity in the present study.

While the eye-tracking data suggest that subjects were not attending to different spatial locations during epochs of attention to faces and houses, and were not systematically shifting the locus of spatial attention following shift targets, a more subtle spatial attention account of the present results cannot be ruled out solely with eye movement data. For instance, subjects may attend only to the central regions of the display when attending to faces (e.g. the eyes–nose–mouth region) and they may attend to the peripheral parts of the stimulus when attending to houses. This alternative account echoes recent evidence suggesting that the lateral and medial fusiform face-selective and house-selective regions of ventral visual cortex may be retinotopically organized visual areas corresponding to the center and the periphery of the visual field, respectively (Levy *et al.*, 2001; Malach *et al.*, 2002). The eye

movement data do not bear on this spatial ‘zoom-lens’ account of the present results because the center of gaze could remain constant under a spatial zooming account; only the extent of spatial attention should covertly vary while attending faces and houses, respectively.

To test this alternative account, we designed a second experiment in which the house stimuli overlapped with the central portion of the face stimuli (Fig. 5). This stimulus arrangement would render an attentional zooming mechanism ineffective because the relevant object features are now spatially overlapping. If the results of Experiment 1 were due to attentional zooming, then a different temporal and spatial BOLD activation pattern should be observed here.

## Experiment 2

### Methods

#### Subjects

Eight neurologically intact young adults (five females), age 19–30 years, gave written informed consent to participate in the study, which was approved by the Johns Hopkins University institutional review boards. Three of the subjects from Experiment 1 also participated in Experiment 2; the other five subjects had no prior experience with the paradigm. There were no systematic differences between the experienced and naïve subjects.

#### Behavioral Tasks

The functional localizer task used to identify regions of lateral and medial fusiform gyrus was identical to that used in Experiment 1. The object-based attention task was identical to Experiment 1, except that the house stimuli were reduced in size to fit within the central 3° of the face stimuli (see Fig. 5). On average, the eyes–nose–mouth region of the faces subtended 3° vertically, thus the house stimuli overlapped with these foveal features of the face stimuli.

#### Data Acquisition and Analysis

All data acquisition and analysis methods were identical to Experiment 1 except where noted. MRI scanning was carried out with a Philips Intera 3 T scanner (rather than the 1.5 T scanner used in Experiment 1) in the F.M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute, Baltimore, MD. Anatomical images were acquired using an MP-RAGE  $T_1$ -weighted sequence and a SENSE (MRI Devices Inc., Waukesha, WI) head coil ( $T_R = 8.2$  ms,  $T_E = 3.7$  ms, flip angle = 8°, pre-pulse  $T_1$  delay = 852.5 ms, SENSE factor = 2, scan time = 385 s). Whole brain echoplanar functional images (EPI) were acquired in 26 transverse slices ( $T_R = 2000$  ms,  $T_E = 30$  ms, flip angle = 70°, matrix = 80 × 80, FOV = 240 mm, slice thickness = 3 mm, SENSE factor = 2, 1 mm gap). The same EPI parameters were used for both the functional localizer and attention scans. Two runs of functional localizer data were collected and each subject participated in 9 or 10 runs of the experimental task. Two experimental runs were discarded from one subject and one functional localizer run was discarded from another subject due to excessive head movement.

The EPI images from the functional localizer task were spatially smoothed with a 6 mm FWHM Gaussian kernel and the data from the experimental attention task were smoothed with a 4 mm FWHM Gaussian kernel before statistical maps were computed. We opted to spatially smooth the data in Experiment 2 (and not in Experiment 1) to help offset the power loss that accompanies a smaller sample size. A group fixed effects GLM was performed on the data from the functional localizer runs to identify face and house selective ROIs in lateral and medial fusiform gyrus, respectively [single voxel threshold of  $t(2421) = 5.0$ ,  $P < 0.00001$ ].

A separate group random effects GLM was used to identify brain areas that exhibited heightened responses to shift versus hold events during the object-based attention task [single voxel threshold of  $t(7) = 3.5$ ,  $P < 0.01$ ]. A minimum cluster size of 405 ml (15 voxels in the original acquisition space) was adopted to correct for multiple compari-



**Figure 5.** Sample face-house stimuli from Experiment 2. The procedures followed those of Experiment 1 with the exception that the house stimuli now occupied the same region of space as the eyes–nose–mouth region of the face stimuli. Face images developed by the Max-Planck Institute for Biological Cybernetics in Tübingen, Germany.

sons, yielding a corrected statistical threshold of  $P < 0.05$  [determined using the program AlphaSim, (B.D. Ward, <http://afni.nimh.nih.gov/afni/docpdf/alphasim.pdf>) which was used to run 2000 Monte Carlo simulations that took into account the entire EPI acquisition matrix and a 4 mm FWHM Gaussian smoothing kernel].

#### *Eye Position Monitoring*

Eye movements were monitored for six of the eight subjects during fMRI data acquisition using a custom built 30 Hz MRI-compatible video camera that provided input to ViewPoint EyeTracker software (Arrington Research Inc., Scottsdale, AZ). Due to difficulties maintaining stable eye-tracker calibration over the course of the scanning session (because repositioning the subject in the middle of a session is undesirable), a variable amount of eye-tracking data was collected from each subject. For two subjects, three runs of data were collected; four, six, seven and nine runs of data were collected from the remaining subjects, respectively. Because the eye tracker in the scanner had a lower temporal resolution (30 Hz) than the eye tracker we used outside of the scanner in Experiment 1 (250 Hz), the criterion for determining an eye movement was changed. In Experiment 2, an eye movement was registered if the eyes moved  $>0.25^\circ$  between successive video frames, which were acquired every 33 ms, during a 1.5 s temporal interval following the onset of each hold and shift target (excluding blinks). The end of an eye movement was registered when the eye gaze fell below the displacement threshold for at least two sequential video frames. As in Experiment 1, we also computed the mean Cartesian coordinates of the eyes during epochs of attention to faces and houses (these epochs included all timepoints in which attention was focused on faces or houses, not just the 1.5 s following the presentation of a target).

## **Results**

### *Behavioral Data*

A two-way repeated measures ANOVA (stimulus identity  $\times$  target type) revealed that detection accuracy was higher for shift versus hold targets [percentage detected  $\pm$  SE:  $84 \pm 4.3$ ,

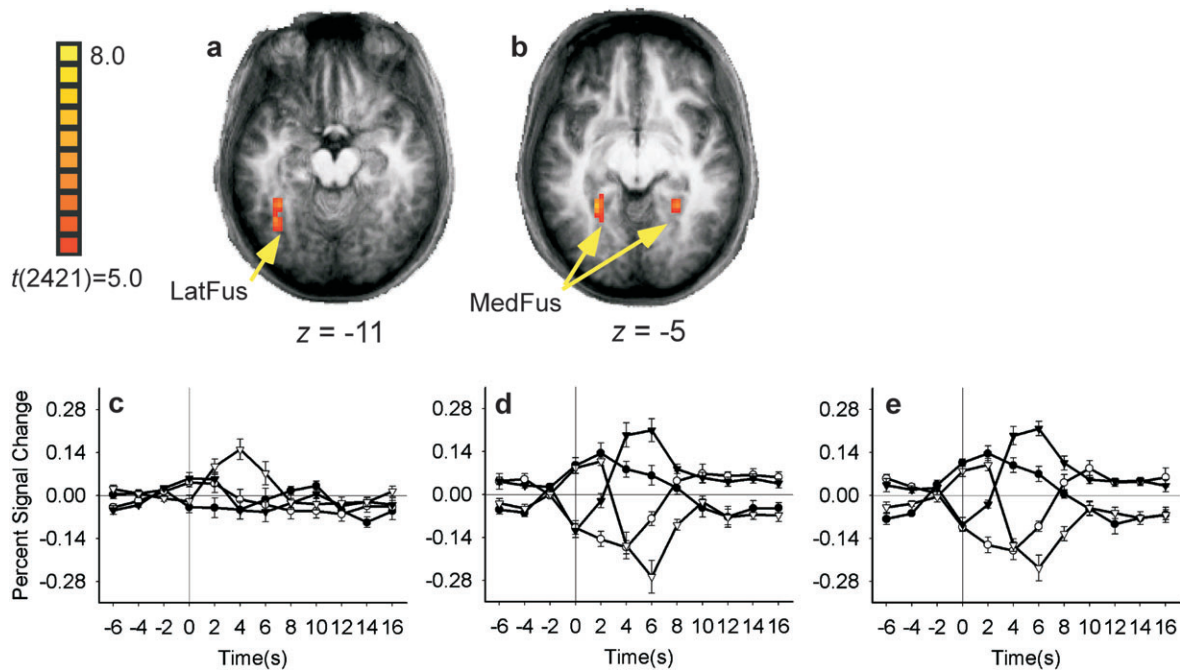
$86 \pm 3.9$ ,  $92 \pm 1.8$  and  $92 \pm 1.6$  for hold house, hold face, shift face-to-house and shift house-to-face events, respectively,  $F(1,7) = 14.1$ ,  $P < 0.01$ ]. No other significant effects were observed for detection rates or for the response times (in ms  $\pm$  SE:  $791 \pm 36$ ,  $818 \pm 36$ ,  $834 \pm 35$  and  $801 \pm 40$  for hold house, hold face, shift face-to-house and shift house-to-face events, respectively,  $F(1,7) = 2.6$ ,  $P > 0.14$ ).

### *Attention Effects in Localizer ROIs*

Figures 6c–e show the group mean event-related timecourses of the BOLD signal in the functional localizer defined right LatFus and the bilateral MedFus for the four event types in the attention task (hold house, hold face, shift face-to-house and shift house-to-face; see Table 3). Figure 6c shows the time course of activity in right LatFus during the attention task. The main effect of attending to faces versus attending to houses failed to reach significance in this region when considering timepoints 4–10 s post-stimulus (open circles and triangles, respectively). However, a paired one-tailed *t*-test revealed a significantly greater response to hold face compared to hold house stimuli at time 0, confirming the early onset of stimulus-specific attention effects in this region [ $t(7) = 2.6$ ,  $P < 0.05$ ]. In addition, the right LatFus region exhibited an interaction between stimulus identity and time, revealing that face selectivity evolved after attention was shifted from a house to a face stimulus [ $F(3,21) = 10.5$ ,  $P < 0.001$ ].

Figure 6d,e shows that the right and the left MedFus regions exhibited greater responses when attention was directed to house compared to face stimuli [right MedFus,  $F(1,7) = 37.4$ ,  $P < 0.001$ ; left MedFus,  $F(1,7) = 77.4$ ,  $P < 0.001$ ]. Paired one-tailed *t*-tests revealed significantly higher activation levels at timepoint 0 in both of these regions [right MedFus,  $t(7) = -4.5$ ,  $P < 0.005$ ; left MedFus,  $t(7) = -9.9$ ,  $P < 0.001$ ]. In addition,





**Figure 6.** Object-based attentional modulation in ventral extrastriate cortex in Experiment 2. (a) Region of right LatFus gyrus that exhibited an increased BOLD response at time 0 when attention is directed to faces versus when attention is directed to houses. In addition, this region showed a strong response when attention was shifted to faces compared to when attention was shifted to houses. (b) Bilateral regions of MedFus that exhibited an increased BOLD response when subjects attended to houses during the attention task. (c–e) Group event related averages from the attention task computed from the right LatFus, right MedFus and left MedFus, respectively (see Methods). Open circle, hold face; filled circle, hold house; filled triangle, shift face-to-house; open triangle, shift house-to-face. Error bars are  $\pm$ SEM.

**Table 3**

Brain regions from Experiment 2 identified in the independent functional localizers as responding more strongly to pictures of faces or places, respectively.

Localizer ROIs	(x, y, z)	Volume (ml)	Diff. at time 0, $t(7)$ , one-tailed	Stim. identity, $F(1,7)$	Time $\times$ identity, $F(3,21)$
Right LatFus	37, -48, -11	0.648	2.56, $P < 0.05^*$	2.1, $P > 0.1$	10.5, $P < 0.001^*$
Right MedFus	25, -45, -5	1.18	-4.5, $P < 0.005^*$	37.4, $P < 0.001^*$	14.3, $P < 0.001^*$
Left MedFus	-22, -44, -6	0.783	-9.9, $P < 0.001^*$	77.4, $P < 0.001^*$	24.8, $P < 0.001^*$

Coordinates from Talairach and Tournoux (1988). Statistical tests were performed on the event-related averages from each of these regions during the attention switching task (see Methods).

significant interactions were observed between stimulus identity and time, such that house stimuli came to dominate the BOLD response after attention was shifted from faces to houses [right MedFus,  $F(3,21) = 14.3$ ,  $P < 0.001$ ; left MedFus,  $F(3,21) = 24.8$ ,  $P < 0.001$ ].

#### Shift-related Activity

Figure 7a,b shows regions of medial SPL–precuneus and right SFS–PreCeG that showed greater BOLD responses following shift compared to hold targets (Table 4). Importantly, these regions exhibit a high degree of spatial correspondence with the SPL and SFS–PreCeG activations observed in Experiment 1. In contrast to Experiment 1, no significant clusters were observed outside of SPL and right SFS–PreCeG. While this null result does not rule out the involvement of other areas in object-based attentional control (such as the lingual gyrus activation depicted in Fig. 3f), it suggests that the occipital and ventral temporal shift-related activations seen in Experiment 1 are not as robust as the frontoparietal activations.

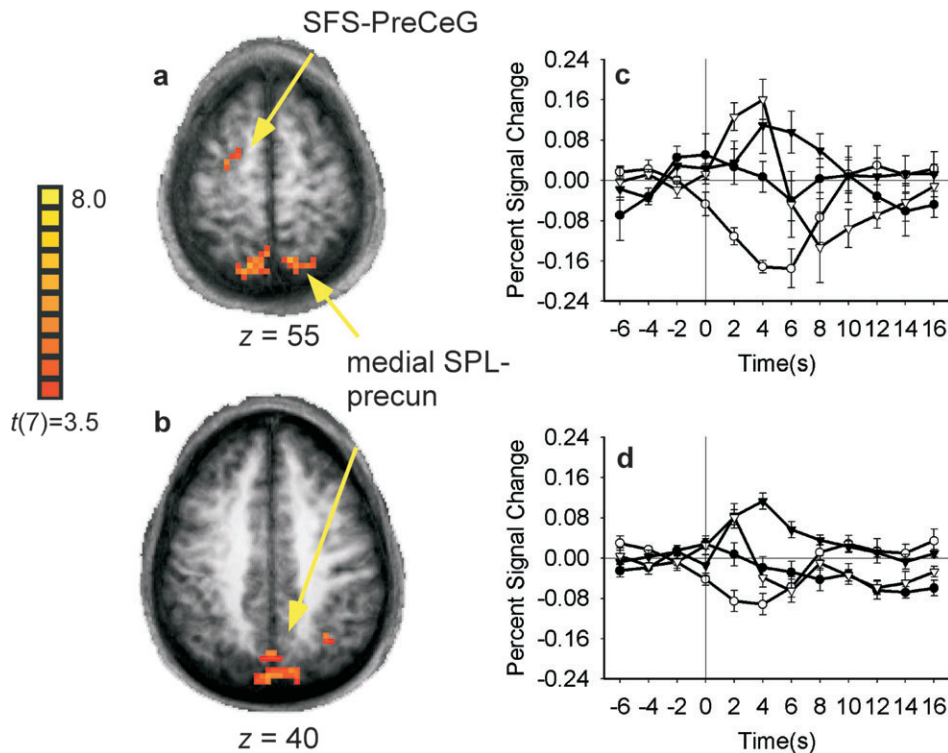
Figure 7c,d shows the group mean event-related BOLD time-course from the medial SPL (Fig. 7c) and the right SFS–PreCeG (Fig. 7d). In addition to the main effect of shifting versus

holding attention, these regions also showed a main effect of attending to houses compared to faces [medial SPL,  $F(1,7) = 11.1$ ,  $P < 0.05$ ; right SFS–PreCeG,  $F(1,7) = 11.2$ ,  $P < 0.05$ ].

We also observed a region in the inferior medial-frontal cortex, extending up the longitudinal fissure to the medial bank of superior frontal cortex that responded more strongly to hold targets than to shift targets. However, because the inferior medial-frontal cortex is highly susceptible to imaging artifacts (e.g. EPI image distortion) and because no such region was observed in Experiment 1, we will not further discuss this activation.

#### Eye Movements

A two-way ANOVA (stimulus identity  $\times$  target type) revealed no differences in the mean number of eye movements occurring in a 1.5 s window following the onset of shift and hold targets [mean number of eye-movements  $\pm$  SE:  $2.8 \pm 0.34$ ,  $2.8 \pm 0.43$ ,  $3.0 \pm 0.45$  and  $2.9 \pm 0.33$  for hold house, hold face, shift face-to-house and shift house-to-face events, respectively,  $F(1,5) = 0.36$ ,  $P > 0.57$ ]. The main effect of stimulus identity and the interaction term were also non-significant ( $P$ s  $> 0.72$ ). Note



**Figure 7.** Regions exhibiting a greater BOLD response to shift as compared to hold targets in Experiment 2. (a, b) Axial slices showing right SFS–PreCeG and medial SPL–precuneus regions. The slices are taken at the same vertical coordinates as Figure 3a,b and show the high degree of correspondence with the activated regions in Experiment 1. (c, d) Group event related averages from the medial SPL–precuneus activation, and right SFS–PreCeG, respectively. Open circle, hold face; filled circle, hold house; filled triangle, shift face-to-house; open triangle, shift house-to-face. Error bars are  $\pm$ SEM.

**Table 4**  
Brain areas exhibiting a main effect of switching attention versus holding attention

Shift > Hold	(x, y, z)	Volume (ml)	t(7)	Stimulus type $\times$ stimulus ID, F(1,7)
Medial SPL	-3, -66, 48	5.751	4.6	0.305, $P > 0.5$
Right SFS–PreCeG	21, -5, 51	0.513	4.3	4.9, $P = 0.06$

Abbreviations: SPL, superior parietal lobule; PreCeG, pre-central gyrus; SFG, superior frontal gyrus. Coordinates from Talairach and Tournoux (1988). Average  $t$ - and  $F$ -values computed across all voxels in activated clusters.

that the mean number of eye movements registered in Experiment 2 is higher than in Experiment 1; however, since different eye trackers and motion criteria were used, the results can not be compared directly. As in Experiment 1, we also examined the number of eye-movements made in a 1 s temporal window corresponding to the time when the target stimuli were physically present on the display: no significant differences were observed [main effect of shift versus hold:  $F(1,5) = 0.003, P > 0.95$ ]. Paired  $t$ -tests revealed no differences between the mean Cartesian coordinates of eye gaze during epochs of attention to faces and houses [ $t(5) = 0.47, P > 0.65$ ,  $t(5) = -0.6, P > 0.56$  for the comparisons of  $X$  and  $Y$  mean eye gaze position, respectively].

### Discussion

The results of Experiment 2 replicate the key findings reported in the first experiment: activation levels in regions of object selective ventral visual cortex are modulated by attention to faces and houses, respectively, and these stimulus-specific

modulations are accompanied by transient activation increases in regions of dorsal parietal and frontal cortex that are commonly thought to mediate voluntary attentional control in other domains (Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002; Yantis *et al.*, 2002).

In addition to the main effect of shifting attention observed in SPL and SFS–PreCeG regions, there was a main effect of attending to houses. This heightened response to house stimuli was also observed in Experiment 1 and most likely reflects our choice of stimuli. For instance, regions of dorsal parietal and frontal cortex have been shown to be more active in working memory tasks for houses compared to faces (Sala *et al.*, 2003). Evidently these areas have a modest selective preference for houses over faces.

While a spatial ‘zoom lens’ model might account for the data in Experiment 1, the stimuli used in the second experiment were designed to rule out this explanation. The house stimuli were reduced in size so that they overlapped with the eyes–nose–mouth region of the face stimuli, which corresponds to the region of space most relevant for discriminating faces (Yarbus, 1967; Levy *et al.*, 2001; Malach *et al.*, 2002). Thus, zooming out when attending to houses and zooming in when attending to faces would be ineffective in Experiment 2.

Eye movement data collected in the scanner revealed no difference either in the number of eye-movements made following shift and hold events, or in the spatial distribution of gaze during epochs of attention to faces or houses. Therefore, we conclude that the shift-related parietal and dorsal frontal activations observed in the present study can be attributed to the control of object-based, and not space-based, attentional control.

## General Discussion

Frontal cortical areas are intimately involved in the maintenance of behavioral goals and are a likely origin of attentional biasing signals (Mesulam, 1981; Desimone and Duncan, 1995; Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002). The posterior parietal cortex has a well documented role in mediating shifts of spatial attention and it has been proposed that dorsal frontal and posterior parietal areas constitute a neural network for voluntarily controlling shifts of spatial attention (Corbetta and Shulman, 2002). The present results suggest that similar areas also participate in controlling non-spatial object-based attention. Although the cortical locus of the activations seen here are similar to those observed in similar spatial attention paradigms (e.g. Yantis *et al.*, 2002), further detailed within-subject comparisons will be necessary to determine precisely the extent of anatomical overlap between space- and object-based attentional control systems.

Because fMRI is a correlational method, the present data do not rule out the possibility that parietal and frontal regions are the recipient of attention-related changes in extrastriate visual areas, rather than the source of the control signals as we have suggested. However, convergent evidence from the human neuropsychology literature supports the role of parietal cortex in controlling object-based attention. Patients with Balint's syndrome, a neurological deficit typically observed following bilateral lesions in the parieto-occipital area, can perceive only one object in the visual field at a time, even if the objects are spatially superimposed (Luria, 1959). Furthermore, unilateral damage to parietal cortex can cause visual neglect in an object-based frame of reference and not just in the contralesional hemifield of space (Driver and Halligan, 1991; Behrmann and Tipper, 1999). The present results, together with these neuropsychological findings, strongly suggest that parietal cortex is causally involved in the control of object-based attention.

### Mechanisms of Attentional Control

Previous reports have shown that regions of extrastriate, parietal, and frontal cortex exhibit a sustained increase in neural activity, the so-called 'baseline shift', when attention is directed to a location in the periphery or a to-be-fixated object, even before the stimulus itself appears (Luck *et al.*, 1997; Chelazzi *et al.*, 1998; Kastner *et al.*, 1999). The baseline shift is greater in parietal and frontal regions than in extrastriate visual cortex (Kastner *et al.*, 1999).

The task employed in the present report could not reveal the presence of a nonlateralized baseline shift because subjects were always attending to an object; there were no periods of rest or passive fixation. Consequently, any baseline shift would have been continuously present throughout the task. Instead, the current paradigm isolates signals associated with attention shifts from other cortical activity by equating perceptual, memory load, and motor requirements following shift and hold targets. If an area only participates in the static maintenance of attention, then no additional increase in the BOLD signal would be observed following shifts of attention. Therefore, we conclude that the brain regions exhibiting more activity following shift targets than following hold targets were transiently active (in addition to any sustained baseline shift they may have exhibited).

Corbetta and colleagues have reported sustained BOLD responses in parietal and frontal compared to occipital regions during a delay period following a spatial attention cue; they concluded that parietal and frontal areas influence attention by tonically maintaining the current locus of attention via a sustained baseline shift (Corbetta *et al.*, 2000; Corbetta and Shulman, 2002). We have found that parietal and frontal regions are transiently active during shifts of attention between objects; functionally similar parietal areas were also found in previous studies of spatial attention shifts (Vandenberghe *et al.*, 2001; Yantis *et al.*, 2002).

Together with data from other laboratories, the present results suggest a dual role for parietal and frontal regions in attentional control involving the initiation (via the transient signal) and the maintenance (via the baseline shift) of a desired attentive state. This functional distinction may involve superior and more inferior regions of parietal cortex, respectively (Corbetta *et al.*, 2000; Hopfinger *et al.*, 2000; Vandenberghe *et al.*, 2001; Yantis *et al.*, 2002).

### Conclusions

The present results reveal that regions in dorsal frontal and parietal cortex mediate nonspatial shifts of object-based attention. Similar areas have previously been implicated in the control of spatial attention (Corbetta *et al.*, 2000; Hopfinger *et al.*, 2000; Beauchamp *et al.*, 2001; Vandenberghe *et al.*, 2001; Yantis *et al.*, 2002). The transient timecourse of parietal and frontal activity suggests that these regions may serve to abruptly change the attentive state of the brain; this new state may then be maintained via sustained increases in activity elsewhere in the brain (Luck *et al.*, 1997; Kastner *et al.*, 1999; Corbetta *et al.*, 2000; Vandenberghe *et al.*, 2001). The functional similarity of the neural systems associated with shifts of spatial and object-based visual attention suggests a domain-general mechanism of attentional control.

### Notes

We thank N. Kanwisher for providing functional localizer stimuli and advice, J.B. Sala and C.E. Stark for helpful suggestions and T. Brawner, K. Kahl and J.J. Pekar for technical assistance with image acquisition. Supported by National Eye Institute Training grant (T32-EY07143) and NSF Graduate Research Fellowship to J.T.S. and NIH grant R01-DA13165 to S.Y.

Address correspondence to John T. Serences, Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 N. Charles St, Baltimore, MD, 21218-2686, e-mail: serences@jhu.edu.

### References

- Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. *Neuron* 21:373-383.
- Arrington CM, Carr TH, Mayer AR, Rao SM (2000) Neural mechanisms of visual attention: object-based selection of a region in space. *J Cogn Neurosci* 12:106-117.
- Beauchamp MS, Petit L, Ellmore TM, Ingelholm J, Haxby JV (2001) A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage* 14:310-321.
- Behrmann M, Tipper SP (1999) Attention accesses multiple reference frames: evidence from visual neglect. *J Exp Psychol Hum Percept Perform* 25:83-101.
- Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299:81-86.
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207-4221.

- Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport* 9:3735-3739.
- Chawla D, Rees G, Friston KJ (1999) The physiological basis of attentional modulation in extrastriate visual areas. *Nat Neurosci* 2:671-676.
- Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 80:2918-2940.
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201-215.
- Corbetta M, Kincade JM, Ollinger JM, McAvooy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci* 3:292-297.
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193-222.
- Driver J, Halligan PW (1991) Can visual neglect operate in object-centered co-ordinates? An affirmative single-case study. *Cogn Neuropsychol* 8:475-496.
- Duncan J (1984) Selective attention and the organization of visual information. *J Exp Psychol Gen* 113:501-517.
- Egeth HE, Virzi RA, Garbart H (1984) Searching for conjunctively defined targets. *J Exp Psychol Hum Percept Perform* 10:32-39.
- Epstein R, Harris A, Stanley D, Kanwisher N (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23:115-125.
- Fink GR, Dolan RJ, Halligan PW, Marshall JC, Frith CD (1997) Space-based and object-based visual attention: shared and specific neural domains. *Brain* 120:2013-2028.
- Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci* 2:568-573.
- Giesbrecht B, Woldorff MG, Song AW, Mangun GR (2003) Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage* 19:496-512.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425-2430.
- Hoffman JE, Subramaniam B (1995) The role of visual attention in saccadic eye movements. *Percept Psychophys* 57:787-795.
- Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. *Nat Neurosci* 3:284-291.
- Ishai A, Ungerleider LG, Haxby JV (2000) Distributed neural systems for the generation of visual images. *Neuron* 28:979-990.
- Kahneman D, Henik A (1981) Perceptual organization. In: *Perceptual organization* (Kubovy M, Pomerantz JR, eds), pp. 181-211. Hillsdale, NJ: Lawrence Erlbaum.
- Kanwisher N, Wojciulik E (2000) Visual attention: insights from brain imaging. *Nat Rev Neurosci* 1:91-100.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302-4311.
- Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci* 23:315-341.
- Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22:751-761.
- Le TH, Pardo JV, Hu X (1998) 4 T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. *J Neurophysiol* 79:1535-1548.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533-539.
- Liu T, Slotnick SD, Serences JT, Yantis S (2003) Cortical mechanisms of feature-based attentional control. *Cereb Cortex* 13:1334-1343.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24-42.
- Luria AR (1959) Disorders of 'simultaneous perception' in a case of bilateral occipitoparietal brain injury. *Brain* 82:437-449.
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176-184.
- Mesulam MM (1981) A cortical network for directed attention and unilateral neglect. *Ann Neurol* 10:309-325.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782-784.
- O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL (1997) Voluntary attention modulates fMRI activity in human MT-MST. *Neuron* 18:591-598.
- O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. *Nature* 401:584-587.
- Posner MI (1980) Orienting of attention. *Q J Exp Psychol* 32:3-25.
- Posner MI, Walker JA, Friedrich FJ, Rafal RD (1984) Effects of parietal injury on covert orienting of attention. *J Neurosci* 4:1863-1874.
- Rock I, Gutman D (1981) The effect of inattention on form perception. *J Exp Psychol Hum Percept Perform* 7:275-285.
- Roelfsema PR, Lamme VA, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395:376-381.
- Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5:631-632.
- Sala JB, Rama P, Courtney SM (2003) Functional topography of a distributed neural system for spatial and nonspatial information maintenance in working memory. *Neuropsychologia* 41:341-356.
- Shulman GL, d'Avossa G, Tansy AP, Corbetta M (2002) Two attentional processes in the parietal lobe. *Cereb Cortex* 12:1124-1131.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575-579.
- Vandenberghe R, Gitelman DR, Parrish TB, Mesulam MM (2001) Functional specificity of superior parietal mediation of spatial shifting. *Neuroimage* 14:661-673.
- Wojciulik E, Kanwisher N (1999) The generality of parietal involvement in visual attention. *Neuron* 23:747-764.
- Yantis S (1992) Multielement visual tracking: attention and perceptual organization. *Cogn Psychol* 24:295-340.
- Yantis S (1998) Control of visual attention. In: *Attention* (Pashler H, ed.), pp. 223-256. Hove, UK: Psychology Press.
- Yantis S, Gibson BS (1994) Object continuity in apparent motion and attention. *Can J Exp Psychol* 48:182-204.
- Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM (2002) Transient neural activity in human parietal cortex during spatial attention shifts. *Nat Neurosci* 5:995-1002.
- Yarbus AL (1967) Eye movements during perception of complex objects. In: *Eye movements and vision* (Riggs LA, ed.), pp. 171-196. New York: Plenum Press.