

Attentional selection of multiple objects in the human visual system



Xilin Zhang^{*}, Nicole Mlynaryk, Shruti Japee, Leslie G. Ungerleider

Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

Keywords:

Object-based attention
Fusiform face area
Parahippocampal place area
Inferior frontal gyrus
Intraparietal sulcus
fMRI

ABSTRACT

Classic theories of object-based attention assume a single object of selection but real-world tasks, such as driving a car, often require attending to multiple objects simultaneously. However, whether object-based attention can operate on more than one object at a time remains unexplored. Here, we used functional magnetic resonance imaging (fMRI) to address this question as human participants performed object-based attention tasks that required simultaneous attention to two objects differing in either their features or locations. Simultaneous attention to two objects differing in features (face and house) did not show significantly different responses in the fusiform face area (FFA) or parahippocampal place area (PPA), respectively, compared to attending a single object (face or house), but did enhance the response in the inferior frontal gyrus (IFG). Simultaneous attention to two circular arcs differing in locations did not show significantly different responses in the primary visual cortex (V1) compared to attending a single circular arc, but did enhance the response in the intraparietal sulcus (IPS). These results suggest that object-based attention can simultaneously select at least two objects differing in their features or locations, processes mediated by the frontal and parietal cortex, respectively.

1. Introduction

Since neural resources are severely limited, the efficient processing of visual information requires selecting only a few behaviorally relevant items at any given moment in time. Attention is the main mechanism that controls this selection process. Numerous studies have demonstrated that attentional selection can be based on either a visual feature (Liu et al., 2007; Maunsell and Treue, 2006; Saenz et al., 2002; Serences and Boynton, 2007; Treue and Martinez-Trujillo, 1999; Zhang and Luck, 2009) or a spatial location (Brefczynski and DeYoe, 1999; Corbetta and Shulman, 2002; Kanwisher and Wojciulik, 2000; Kastner and Ungerleider, 2000; Martínez et al., 1999; Posner et al., 1980; Tootell et al., 1998; Zhang et al., 2012, 2016). Increasing evidence has indicated that attentional selection can also be deployed to entire objects: specific objects can be selected by directing attention to their features or to their locations (Scholl, 2001; Chen, 2012).

Directing attention to one feature of an object, such as its moving direction, can result in selection of the whole object as an integrated perceptual unit, including both its task-relevant and task-irrelevant features (Baldauf and Desimone, 2014; Cohen and Tong, 2013; Jiang et al., 2016; O'Craven et al., 1999; Schoenfeld et al., 2014; Serences et al., 2004). These findings are consistent with the predictions of the

integrated competition (Duncan et al., 1997) and incremental grouping (Roelfsema, 2006) models, both of which suggest that the perception of a unified object can occur through the selection and binding of its various features in the cortical areas that represent them. The feature integration theory (Treisman and Gelade, 1980) proposes that spatially directed attention also contributes to object-based attention: attending the location of an object leads to binding of the object's features. Several psychophysical (Chou and Yeh, 2012; Egly et al., 1994; Fiebelkorn et al., 2013; Zhang and Fang, 2012), physiological (Pooresmaeili and Roelfsema, 2014; Roelfsema et al., 1998), and brain imaging (Martínez et al., 2006; Müller and Kleinschmidt, 2003; Shomstein and Behrmann, 2006) studies further show that attention directed to one location in an object can automatically spread throughout the whole object. Together, it appears that object-based attention can be induced by both feature- and location-selection.

These models of object-based attention assume a single object of selection but many real-world tasks require attention to multiple objects. Numerous studies have shown that the human visual system is able to process task-relevant features of multiple objects, including multiple object recognition in cluttered scenes (Riesenhuber and Poggio, 1999), tracking of multiple objects (Cavanagh and Alvarez, 2005), as well as holding multiple objects in short-term (working) memory (Luck and

^{*} Corresponding author. Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Building 10, RM 4C205, 10 Center Drive, Bethesda, MD 20892, USA.

E-mail address: xilin.zhang@nih.gov (X. Zhang).

<https://doi.org/10.1016/j.neuroimage.2017.09.050>

Received 3 August 2017; Received in revised form 18 September 2017; Accepted 22 September 2017

Available online 23 September 2017

1053-8119/Published by Elsevier Inc.

Vogel, 1997; Wheeler and Treisman, 2002; Xu and Chun, 2006, 2009), feature binding (Treisman, 1996), and feature misbinding (Zhang et al., 2014). However, whether the human visual system is also able to process task-irrelevant features of multiple objects remains unclear. Previous studies regarding object-based attention have shown that a task-irrelevant object feature can be processed by both feature- and location-selection (Scholl, 2001; Chen, 2012). However, whether there are different neural mechanisms by which feature- and location-selection enhance object-based attention remains unclear. Moreover, previous studies have demonstrated that feature-selection can globally improve the processing of all stimuli with the same attended feature (Andersen et al., 2013; Maunsell and Treue, 2006; Saenz et al., 2002; Treue and Martinez-Trujillo, 1999; Zhang and Luck, 2009), and location-selection can be split between discrete regions of space without a cost to the spatial attention effect (McMains and Somers, 2004). We thus predicted that our visual system would be able to process task-irrelevant features of multiple objects simultaneously in an object-based attention task. In other words, we predicted that object-based attention can simultaneously select more than one object (for example, two objects) by directing attention to their particular features or their locations. If this is true, then directing attention to one feature (such as the motion direction) in two objects (face and house), the moving face and moving house would be constructed simultaneously as separate perceptual units through selection and binding of the task-relevant (the motion direction) and task-irrelevant features (the face and house shape) in FFA and PPA, respectively. Directing attention to two discrete locations of two objects (such as two circular arcs) can spread simultaneously throughout the entire object and facilitate the processing of task-irrelevant regions located within the boundaries of those two circular arcs.

To test these predictions, we used fMRI while human participants performed object-based attention tasks that required simultaneous attention to two objects differing in either their features (Experiment 1) or locations (Experiment 2). In Experiment 1, participants viewed stimuli consisting of a face transparently superimposed on a house (Fig. 1A). Attention to one feature (the motion direction or static position) in two objects (face and house) versus in a single object (face or house) did not

show significantly different evoked responses in their respective sensory modules: FFA and PPA, respectively, but enhanced the response in IFG. In Experiment 2, participants were randomly cued to attend to one end of a single circular arc (the single cue condition) or of two different circular arcs (the double cue condition) (Fig. 3A). A comparison of the single and double cue conditions did not show a significant difference in the object-based attentional effect in V1, but enhanced the response in IPS for the double cue condition. Together, our study provides strong evidence for multiple object-based attention in the human visual system and implies a crucial involvement of the IFG and IPS in this process when induced by feature- and location-selection, respectively.

2. Materials and methods

2.1. Participants

A total of 16 adults (11 male, 22–35 years old) participated in both Experiments 1 and 2. All were naïve to the purpose of the study. They reported normal or corrected-to-normal vision, and had no known neurological, psychiatric or visual disorders. They gave written informed consent in accordance with a protocol approved by the National Institute of Mental Health (NIMH) Institutional Review Board.

2.2. Experiment 1

Each stimulus was constructed by transparently superimposing one of sixteen grayscale front-view photographs of faces on one of sixteen grayscale photographs of houses. The stimulus subtended about 9.3° of visual angle (Fig. 1A).

Using an event-related design, the experiment consisted of eight functional runs, four for the attend moving direction (Attend Moving scan) and four for the attend static position (Attend Static scan). The order of these two different scans was counterbalanced across participants. These two different scans had three of the same trial types: a moving face with a static house (MFSH), a static face with a moving house (SFMH), and fixation. The stationary item was displaced about

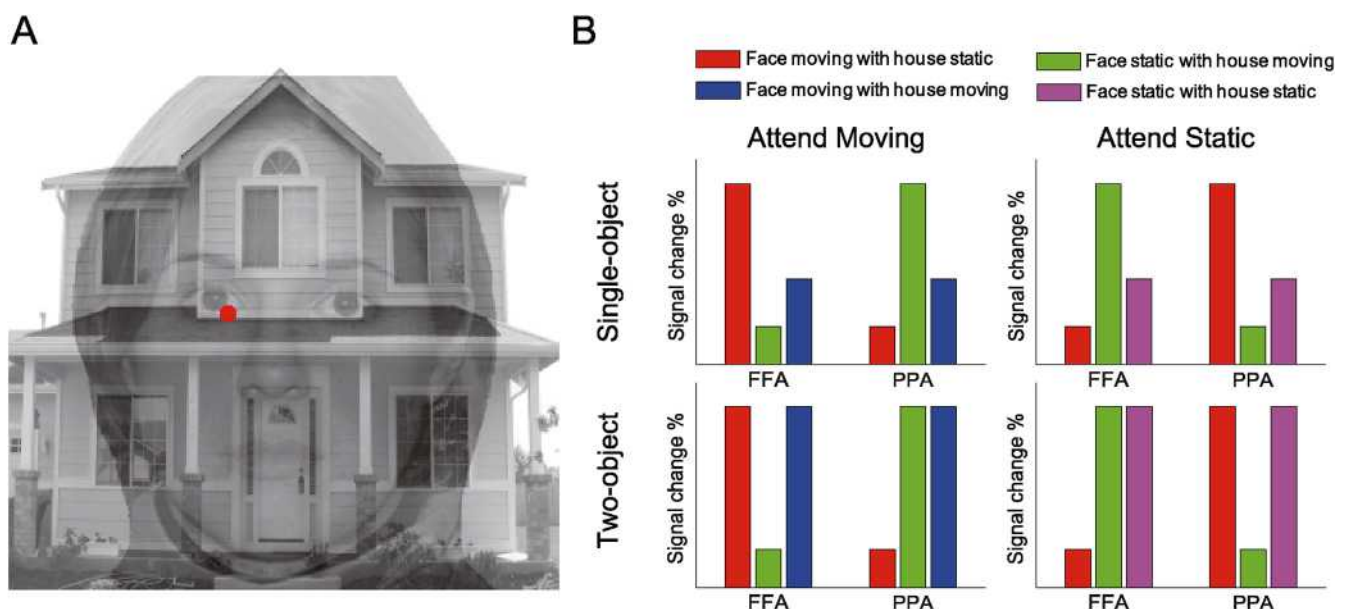


Fig. 1. A sample stimulus and two hypotheses of Experiment 1. **A** A sample stimulus of a superimposed face and a house, with the face and house displaced (i.e., a face static with a house static condition, SFSH) to the right of fixation (about 1.0° eccentricity) as required for the static position-repetition detection task. **B** Two hypotheses of Experiment 1. Single-object attention (top): The attended feature (the motion direction or static position) in two objects (face and house) versus in a single object (face or house) should show reduced responses in FFA and PPA, respectively. Two-object attention (bottom): The attended feature in two objects (face and house) versus in a single object (face or house) should not show significantly different responses in FFA and PPA, respectively.

1.0° in one of four cardinal directions from central fixation (i.e., above, below, left, or right). The moving item moved along a straight path on one of four non-cardinal axes, making one excursion out from and back to the center. The fourth trial type was a moving face with a moving house (same motion direction, MFMH) during the Attend Moving scans and a static face with a static house (same position, SFSH) during the Attend Static scans. In a given trial, each stimulus presentation lasted 400 ms, followed by a 1600-ms fixation period. In a fixation trial, only the fixation point was presented for 2 s. In each scan, there were 32 trials for each type of trial. The order of the trials was counterbalanced across four scans using M-sequences (Buracas and Boynton, 2002). These are pseudo-random sequences that have the advantage of being perfectly counterbalanced n trials back, so that each type of trial was preceded and followed equally often by all types of trials, including itself. Participants performed a one-back repetition detection task on the motion direction (one of four non-cardinal directions) and static position (above, below, left, or right relative to fixation) during the Attend Moving and Attend Static scans, respectively. Repetitions of the motion direction or static position occurred randomly with 12.5% probability and were independent of the M-sequences (Buracas and Boynton, 2002) of stimulus conditions (i.e., MFSH, SFMH, and MFMH in the Attend Moving scan; MFSH, SFMH, and SFSH in the Attend Static scan). They were asked to press a button to respond to the repetition. Methods of this experiment closely followed those used by Chen et al. (2016) and therefore, for consistency, we largely reproduce that description here, noting differences as necessary.

A block-design scan was used to localize the regions-of-interest (ROIs) in face-, house-, and motion-selective areas. Participants viewed images of moving faces, static faces, moving houses, and static houses (the stimuli subtended about 9.3° of visual angle), which were presented at the center of the screen. All images appeared at a rate of 2 Hz in blocks of 12 s, interleaved with 12-s blank blocks. Each image was presented for 300 ms, followed by a 200-ms blank interval. Each block type was repeated 4 times in the run, which lasted 384 s. Participants performed a one-back task during scanning. A general linear model (GLM) procedure was used for the ROI analysis. The fusiform face area (FFA) was defined as an area that responded more strongly to faces than houses ($p < 10^{-3}$, uncorrected). The parahippocampal place area (PPA) was defined as an area that responded more strongly to houses than faces ($p < 10^{-3}$, uncorrected). The middle temporal plus (MT+) area was defined as an area that responded more strongly to moving stimuli than static stimuli ($p < 10^{-3}$, uncorrected). The mean Talairach coordinates of ROIs for the left and right hemispheres in FFA were $[-40 \pm 0.68, -48 \pm 1.95, -19 \pm 0.93]$ and $[36 \pm 0.80, -52 \pm 1.83, -18 \pm 0.73]$, respectively, those in PPA were $[-26 \pm 0.60, -47 \pm 1.20, -11 \pm 1.00]$ and $[25 \pm 0.84, -43 \pm 1.05, -9 \pm 0.66]$, respectively, and those in MT+ were $[-44 \pm 1.17, -66 \pm 1.15, -4 \pm 1.08]$ and $[43 \pm 1.24, -65 \pm 1.37, -4 \pm 1.11]$, respectively.

Event-related blood oxygen level-dependent (BOLD) signals were calculated separately for each subject, following the method used by Kourtzi and Kanwisher (2000). For each event-related scan, the time course of the MR signal intensity was first extracted by averaging the data from all the voxels within the predefined ROI. The average event-related time course was then calculated for each type of trial, by selectively averaging to stimulus onset and using the average signal intensity during the fixation trials as a baseline to calculate percent signal change. Specifically, in each scan, we averaged the signal intensity across the trials for each type of trial at each of 7 corresponding time points starting from the stimulus onset. These event-related time courses of the signal intensities were then converted to time courses of percent signal change for each type of trial by subtracting the corresponding value for the fixation trials and then dividing by that value. Because M-sequences have the advantage that each type of trial was preceded and followed equally often by all types of trials, the overlapping BOLD responses due to the short interstimulus interval are removed by this averaging procedure (Buracas and Boynton, 2002).

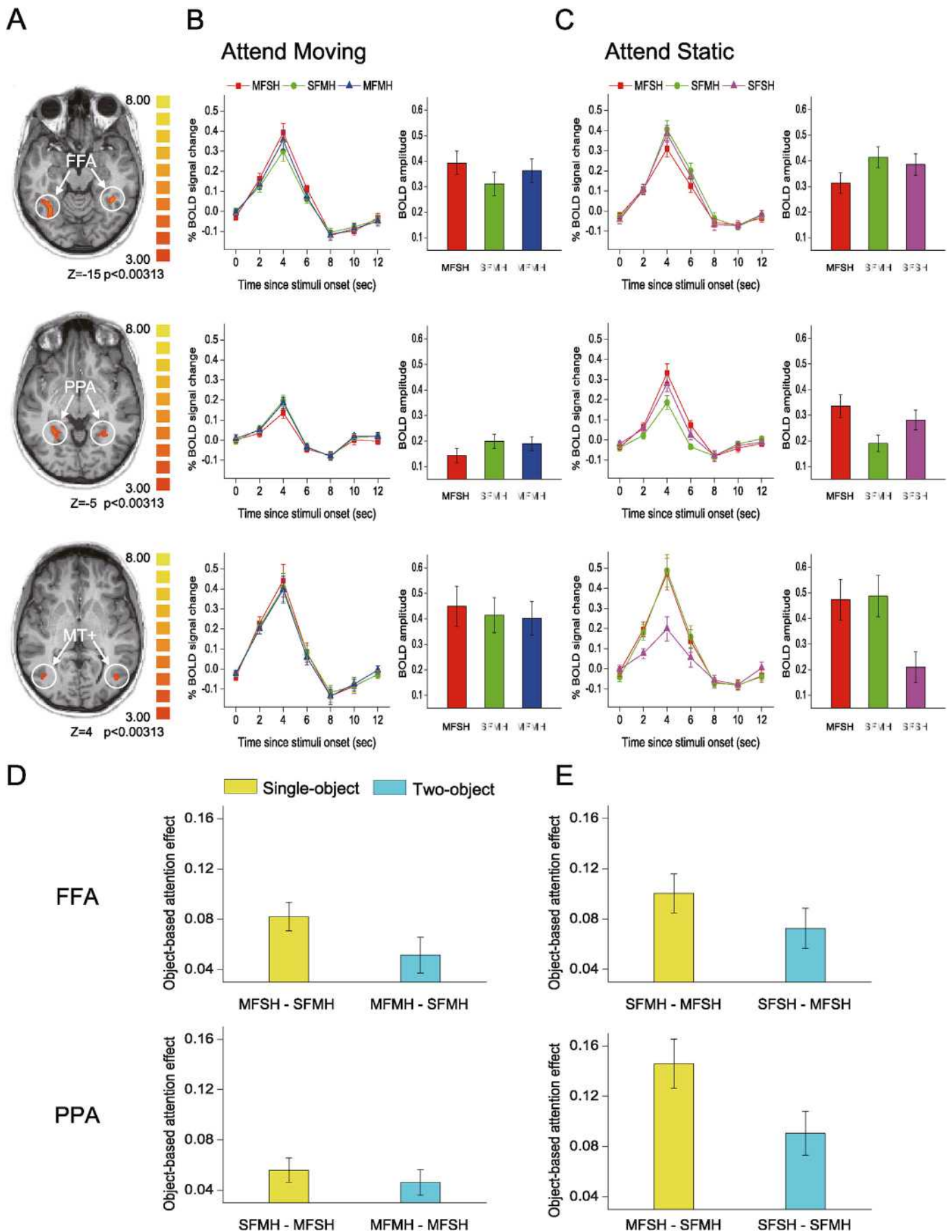
The resulting time course for each type of trial was then averaged across scans for each subject and then across participants.

2.3. Experiment 2

Experiment 2 was composed of two stimuli, each containing the same three circular arcs (luminance: 76.61 cd/m²; radian: 1/3 π ; length: 7.67°–8.89°; width: 1.16°) at 7.91° eccentricity. For one stimulus, two of three circular arcs were presented in the right visual field (Fig. 3A), and for the other stimulus, they were presented in the left visual field. The six ends of the three circular arcs were the possible locations of cues and targets. The single cue (luminance: 0.87 cd/m²; radian: 1/30 π ; length: 0.767°–0.889°; width: 1.16°) was an empty wedge that overlapped one end of a circular arc. For the double cue, two single cues were always diametrically opposite to each other, and were presented in either the same (double cue same hemifield, Fig. 3B) or opposite hemifields (double cue opposite hemifield, Fig. 3C). The target (radian: 1/30 π ; length: 0.767°–0.889°; width: 1.16°) was a solid red or green wedge that overlapped one end of a circular arc.

Using a slow event-related design, the experiment consisted of twelve functional runs, six for each stimulus (Fig. 3A). Each run consisted of 24 trials, which lasted 432 s. On each trial, three circular arcs were always present on the screen. Each trial began with the fixation cross presented for 8, 10, or 12 s. Then the single cue or the double cue was presented for 6, 8, or 10 s, indicating the location that would most likely contain the target stimulus to follow. The single cue and the double cue appeared with equal probability and randomly in the experiment. Following the cue display, the target was presented immediately (with zero delay) for 120 ms at an end of a circular arc. Participants were asked to use the same finger on the same hand to press one of two buttons as rapidly and correctly as possible to indicate the color of the target (red or green) (Fig. 3E). In 75% of the trials, the cue was valid and the target appeared at the pre-cued location (Valid Cue, VC). In 12.5% of the trials, the cue was invalid and the target appeared either at the uncued end of the same circular arc (Invalid Cue Same Object, ICSO) or, in 12.5% of the trials, at the equidistant end of the uncued circular arc (Invalid Cue Different Object, ICDO). All conditions (VC, ICSO, and ICDO) were randomized in a scan. Participants were told the percentage of correct responses after each scan.

Retinotopic visual areas, including V1, were defined by a standard phase-encoded method developed by Sereno et al. (1995) and Engel et al. (1997), in which participants viewed rotating wedge and expanding ring stimuli that created traveling waves of neural activity in visual cortex. A block-design scan was used to localize the ROIs in V1 corresponding to the six possible cue and target locations (Fig. 3D). The scan consisted of 4 12-s stimulus blocks, interleaved with 4 12-s blank intervals. In a stimulus block, participants passively viewed 8-Hz flickering patches. A GLM procedure was used for the ROI analysis. The ROIs in V1 were defined as areas that responded more strongly to the flickering patches than the blank screen ($p < 10^{-3}$, uncorrected). During the cue phase, the BOLD response to the cue was averaged across voxels within the predefined ROIs. The 2-s preceding the cue presentation served as a baseline, and data were collapsed over location and different durations of the cue (6, 8, and 10 s). During the target phase, the 2-s preceding the target presentation served as a baseline to avoid a confound from the cueing period, and data were collapsed over locations and different durations of the fixation (8, 10, and 12 s). BOLD signals evoked by the targets were extracted from the ROIs in V1 and then selectively averaged according to the different conditions (VC, ICSO, and ICDO). Moreover, we analyzed the valid target and invalid target trials separately, and to exclude stimulus-driven activation by the target, only trials without the target at the corresponding locations were analyzed for each predefined ROI during the target phase.



2.4. MRI data acquisition

MRI data were collected using a 3 T Siemens Trio scanner with a 32-channel phase-array coil. In the scanner, the stimuli were rear-projected via a video projector (refresh rate: 60 Hz; spatial resolution: 1280×800) onto a translucent screen placed inside the scanner bore. Participants viewed the stimuli through a mirror located above their eyes. The viewing distance was 115 cm. BOLD signals were measured with an echoplanar imaging sequence (TR: 2000 ms; TE: 30 ms; FOV: $192 \times 192 \text{ mm}^2$; matrix: 64×64 ; flip angle: 70; slice thickness: 3 mm; gap: 0 mm; number of slices: 34; slice orientation: axial). The bottom slice was positioned at the bottom of the temporal lobes. A 3D MPRAGE structural data set (resolution: $1 \times 1 \times 1 \text{ mm}^3$; TR: 2600 ms; TE: 30 ms; FOV: $256 \times 224 \text{ mm}^2$; flip angle: 7; number of slices: 176; slice orientation: sagittal) was collected in the same session before the functional scans. Participants underwent three sessions, one for retinotopic mapping and ROI localization and the other two for Experiments 1 and 2, respectively.

2.5. MRI data analysis

The anatomical volume for each subject in the retinotopic mapping session was transformed into a brain space that was common for all participants (Talairach and Tournoux, 1988) and then inflated using BrainVoyager QX. Functional volumes in all three sessions for each subject were preprocessed, including 3D motion correction, linear trend removal, and high-pass (0.015 Hz, Smith et al., 1999) filtering using BrainVoyager QX. Head motion within any fMRI session was $<3 \text{ mm}$ for all participants. The images were then aligned to the anatomical volume from the retinotopic mapping session and transformed into Talairach space. The first 8-s of BOLD signals was discarded to minimize transient magnetic saturation effects. A general linear model (GLM) procedure was used to determine ROIs in FFA, PPA, and MT+ (Experiment 1) and V1's boundaries (Experiment 2).

In the whole-brain group analysis, for both Experiments 1 and 2, a fixed-effects general linear model (FFX-GLM) was performed for each subject on the spatially non-smoothed functional data in Talairach space. The design matrix consisted of two predictors (the two-object and single-object conditions), which were modeled as epochs using the default BrainVoyager QX's two-gamma haemodynamic response function. Six additional parameters resulting from 3D motion correction (x , y , z rotation and translation) were included in the model. First, we calculated fixed effects analyses for each subject for the two predictors. Second, a second-level group analysis ($n = 16$) was performed with a random-effects general linear model (RFX-GLM) to calculate the contrast between the two predictors. Statistical maps were thresholded at $p < 0.01$, corrected by false discovery rate (FDR) correction (Genovese et al., 2002).

2.6. Eye movement recordings

Eye movements were recorded with an ASL EyeTrac 6000 (Applied Science Laboratories, Bedford, MA) in a psychophysics lab (outside the scanner). Its temporal resolution was 60 Hz and spatial resolution was 0.25° . Recording data were collected when participants performed the same tasks as in Experiments 1 and 2. Figs. S1 and S2 show that

participants' eye movements were small and statistically indistinguishable across all conditions in Experiments 1 and 2, respectively.

3. Results

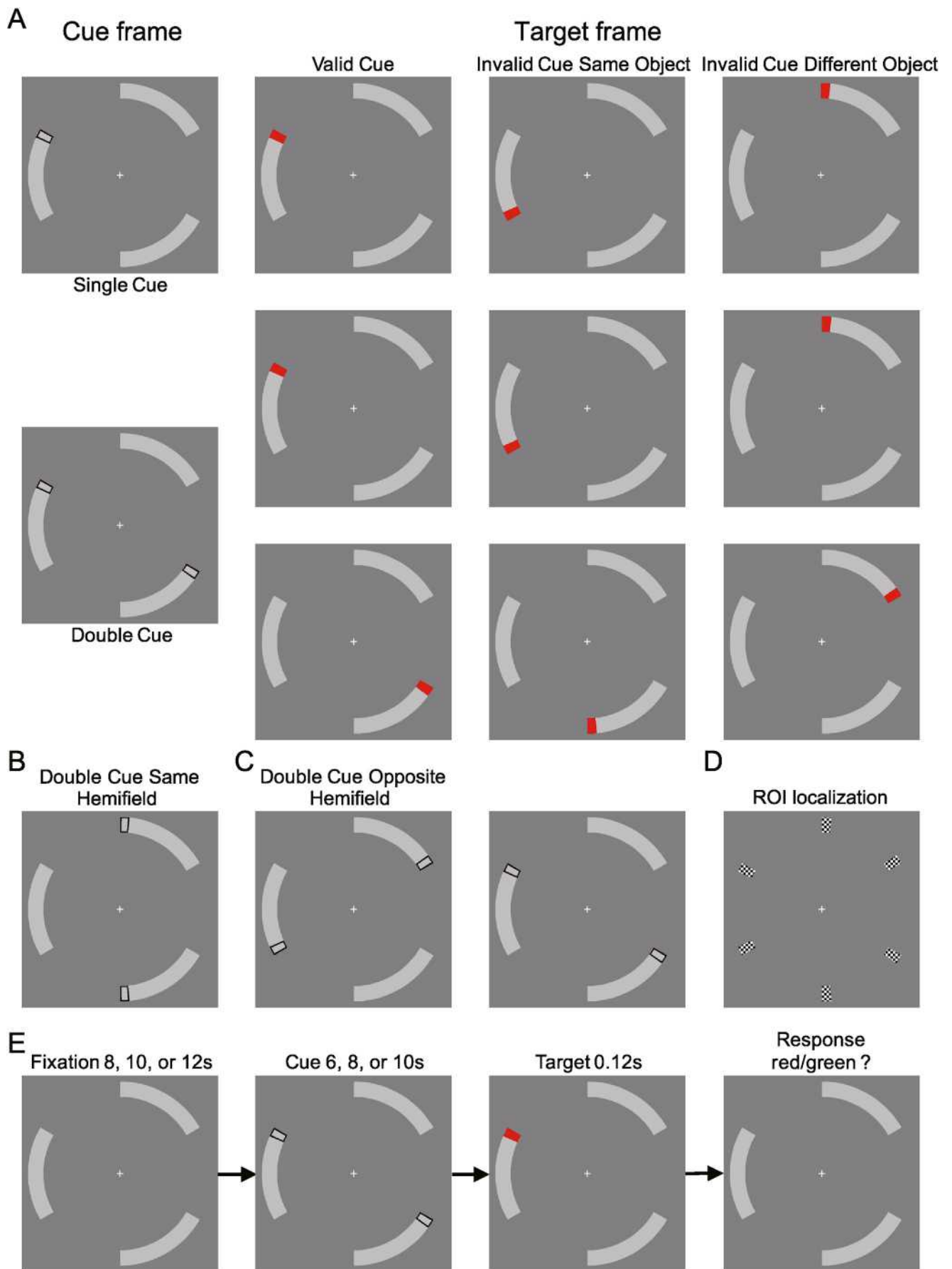
3.1. Experiment 1: objects differing in features

In Experiment 1, participants viewed stimuli consisting of a face transparently superimposed on a house (Fig. 1A). They performed a one-back repetition detection task on the motion direction (the Attend Moving scan, Fig. 1B, left) in conditions with a moving face with a static house (MFSH), a static face with a moving house (SFMH), and a moving face with a moving house (MFMH), or on the static position (the Attend Static scan, Fig. 1B, right) in the MFSH condition, SFMH condition, and in a condition with a static face with a static house (SFSH). During the Attend Moving scan, a repeated measures ANOVA revealed no significant differences in hit rates ($92.72 \pm 1.47\%$, $93.10 \pm 1.48\%$, and $93.64 \pm 1.61\%$, $F_{2, 30} = 0.25$, $p = 0.74$) or in false alarm rates ($6.95 \pm 0.95\%$, $7.04 \pm 0.68\%$, and $8.39 \pm 0.81\%$, $F_{2, 30} = 1.84$, $p = 0.18$) for the MFSH, SFMH, and MFMH conditions. Similarly, during the Attend Static scan, there was no significant difference in hit rate ($60.47 \pm 1.29\%$, $60.80 \pm 1.57\%$, and $62.59 \pm 1.60\%$, $F_{2, 30} = 1.55$, $p = 0.23$) or in false alarm rate ($15.69 \pm 0.78\%$, $15.35 \pm 0.59\%$, and $14.40 \pm 0.74\%$, $F_{2, 30} = 1.15$, $p = 0.33$) for the MFSH, SFMH, and SFSH conditions. For all these measurements, there was no significant difference among the three types of conditions in each scan.

BOLD signals were extracted from the ROIs in FFA, PPA, and MT+ and averaged according to the different conditions. The peak BOLD signal for the stimulus was used as a measure of response amplitude. We hypothesized that, if only a single object can be attended, then attention to one feature (the motion direction or static position) in two objects (face and house) versus in a single object (face or house) should show significantly reduced responses in FFA and PPA, respectively (Fig. 1B, left). However, if two objects can be attended simultaneously during object-based attention, then attention to one feature in two objects (face and house) versus in a single object (face or house) should not show significantly different responses in FFA and PPA, respectively (Fig. 1B, right).

During the Attend Moving scan, the BOLD amplitudes from the ROI in FFA, PPA, and MT+ for the three types of moving conditions (i.e., MFSH, SFMH, and MFMH) are shown in Fig. 2B, and were submitted to a repeated-measures ANOVA with moving condition as a within-subject factor. The main effect in MT+ was not significant ($F_{2, 30} = 2.29$, $p = 0.13$), but was significant in both FFA ($F_{2, 30} = 15.79$, $p < 0.001$) and PPA ($F_{2, 30} = 14.96$, $p < 0.001$). For FFA, compared to the SFMH condition, the enhanced response in the MFSH ($t_{15} = 7.29$, $p < 0.001$) and MFMH ($t_{15} = 3.59$, $p = 0.008$) conditions indicated an object-based attention effect in the single-object and two-object conditions, respectively, with no significant difference between these two conditions ($t_{15} = 1.71$, $p = 0.11$, Fig. 2D, top). For PPA, compared to the MFSH condition, the enhanced response in the SFMH ($t_{15} = 5.76$, $p < 0.001$) and MFMH ($t_{15} = 4.54$, $p = 0.001$) conditions indicated an object-based attention effect in the single-object and two-object conditions, respectively, with no significant difference between these two conditions ($t_{15} = 0.77$, $p = 0.46$, Fig. 2D, bottom).

Fig. 2. Results of Experiment 1. **A** Face-, house-, and motion-selective areas depicted on the brain of a single subject. The ROI in FFA was defined as an area that responded more strongly to faces than houses. The ROI in PPA was defined as an area that responded more strongly to houses than faces. The ROI in MT+ was defined as an area that responded more strongly to the moving stimuli than static stimuli. **B** Event-related BOLD signals and their peak amplitudes of ROIs in FFA (top), PPA (middle), and MT+ (bottom) during the Attend Moving scan. **C** Event-related BOLD signals and their peak amplitudes of ROIs in FFA (top), PPA (middle), and MT+ (bottom) during the Attend Static scan. **D** Object-based attention effect of the single-object and two-object conditions in FFA and PPA during the Attend Moving scans. Compared to the SFMH condition, enhanced responses in the MFSH and MFMH conditions indicated the object-based attention effect in FFA for the single-object and two-object conditions, respectively; compared to the MFSH condition, enhanced responses in the SFMH and MFMH conditions indicated an object-based attention effect in PPA for the single-object and two-object conditions, respectively. **E** Object-based attention effect of the single-object and two-object conditions in FFA and PPA during the Attend Static scans. Compared to the MFSH condition, the enhanced response in the SFMH and SFSH conditions indicated an object-based attention effect in FFA for the single-object and two-object conditions, respectively; compared to the SFMH condition, enhanced responses in the MFSH and SFSH conditions indicated an object-based attentional effect in PPA for the single-object and two-object conditions, respectively. Error bars denote 1 SEM calculated across sixteen participants.



During the Attend Static scan, the BOLD amplitudes from the ROI in FFA, PPA, and MT+ for the three types of static conditions (i.e., MFSH, SFMH, and SFSH) are shown in Fig. 2C, and were submitted to a repeated-measures ANOVA with static condition as a within-subject factor. The main effects in FFA ($F_{2, 30} = 28.48, p < 0.001$), PPA ($F_{2, 30} = 25.86, p < 0.001$), and MT+ ($F_{2, 30} = 55.93, p < 0.001$) were all significant. For FFA, compared to the MFSH condition, the enhanced response in the SFMH ($t_{15} = 6.50, p < 0.001$) and SFSH ($t_{15} = 5.36, p < 0.001$) conditions indicated an object-based attention effect in the single-object and two-object conditions, respectively, with no significant difference between these two conditions ($t_{15} = 1.78, p = 0.096$, Fig. 2E, top). For PPA, compared to the SFMH condition, the enhanced response in the MFSH ($t_{15} = 7.44, p < 0.001$) and SFSH ($t_{15} = 6.18, p < 0.001$) conditions indicated an object-based attention effect in the single-object and two-object conditions, respectively, with no significant difference between these two conditions ($t_{15} = 1.96, p = 0.068$, Fig. 2E, bottom). For MT+, there was no significant difference between MFSH and SFMH conditions ($t_{15} = 0.72, p = 1.00$), and both were significantly greater than the SFSH condition (MFSH versus SFSH: $t_{15} = 7.59, p < 0.001$; SFMH versus SFSH: $t_{15} = 8.71, p < 0.001$). Together, these results indicate that object-based attention can be simultaneously directed to at least two objects differing in their features.

3.2. Experiment 2: objects differing in locations

In Experiment 2, each stimulus contained the same three circular arcs, whose six ends were the possible locations of the cue and target. The single cue was an empty wedge that overlapped one end of a circular arc; the double cue contained two single cues on different circular arcs, presented either within the same (Fig. 3B) or opposite hemifields (Fig. 3C). The target was a solid red or green wedge that overlapped one end of a circular arc. Using a modified spatial cuing paradigm (Egly et al., 1994), each trial began with the fixation cross presented for 8, 10, or 12 s. Then the single cue or double cue was randomly presented for 6, 8, or 10 s with equal probability, followed with zero delay by the target presented for 120 ms at one end of a circular arc. Participants were asked to use the same finger on the same hand to press one of two buttons as rapidly and accurately as possible to indicate the color of the target (red or green, Fig. 3E). Miss rates were 1.56% and 1.65%, and false alarm rates were 1.70% and 2.17% for the single and double cue conditions, respectively. Correct reaction times (RTs) less than 150 ms and greater than three standard deviations from the mean RT in each condition were removed, resulting in 2.95% and 3.13% removal rates in the single and double cue conditions, respectively. There was no significant difference (all $p > 0.05$) for any of these measurements between the single and double cue conditions.

The main goal of Experiment 2 was to compare the object-based attention effect between the single and double cue conditions. The object-based attention effect was quantified as the RT difference between trials where the invalidly cued target appeared at the opposite end of the cued circular arc (the invalid cue same object, ICSO) and trials where the invalidly cued target appeared at the equidistant end of the uncued circular arc (the invalid cue different object, ICDO) (Fig. 3A). We also calculated the spatial attention effect, quantified as the RT difference between valid cue (VC) and invalid cue (i.e., the ICSO and ICDO conditions), during the single and double cue conditions (Fig. 3A), in order to compare effect strengths. Note that the spatial attention effect here was confounded with the physical stimulus difference—that is, the cue was

continuously present at the valid cue location but not present at the invalid cue locations. However, the object-based attention effect (i.e., ICSO - ICDO) was not affected by this confound.

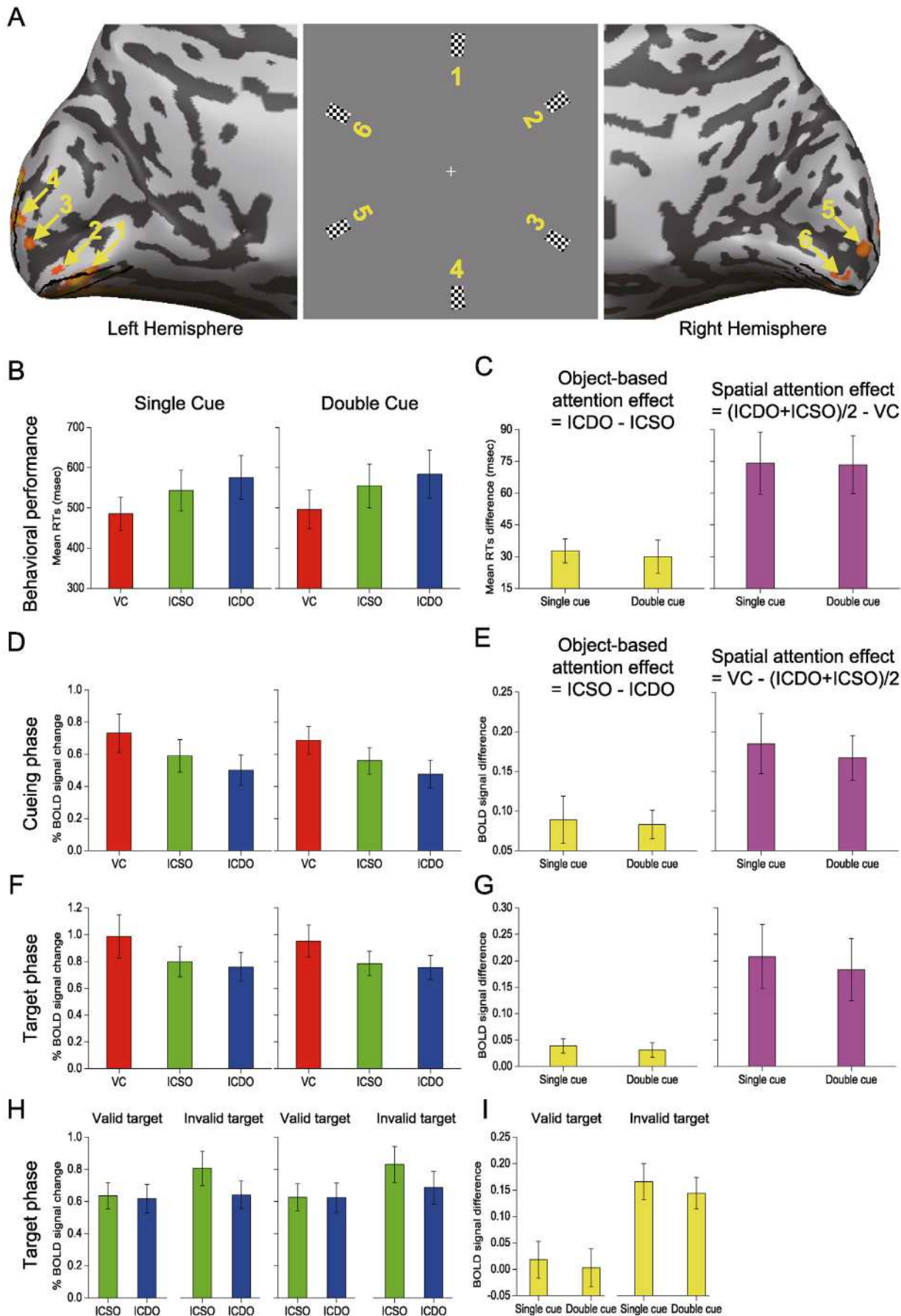
The object-based attention effect (RT: ICDO - ICSO) for the single cue (32.71 ± 5.71 ms, $t_{15} = 5.73, p < 0.001$) and double cue (29.99 ± 7.86 ms, $t_{15} = 3.84, p = 0.002$) conditions were significantly above zero (Fig. 4C, left), as was the spatial attention effect (RT: (ICDO + ICSO)/2 - VC: single cue: 74.08 ± 14.65 ms, $t_{15} = 5.06, p < 0.001$; double cue: 73.43 ± 13.66 ms, $t_{15} = 5.38, p < 0.001$, Fig. 4C, right). A repeated-measures ANOVA with cue type (single, double) and effect type (object, spatial) as within-subject factors showed that the main effect of cue type ($F_{1, 15} = 0.10, p = 0.75$) and the interaction between cue type and effect type ($F_{1, 15} = 0.07, p = 0.80$) were not significant. The main effect of effect type was significant ($F_{1, 15} = 24.76, p < 0.001$), and showed a greater spatial attention effect than object-based attention effect.

ROIs in V1 were defined as those V1 regions responding significantly to the six ends of the three circular arcs (Fig. 4A). We focused our analysis on V1 because activated areas in extrastriate regions corresponding to these six ends of the three circular arcs showed a great deal of overlap. BOLD signals were extracted from the ROIs in V1 and then selectively averaged according to the different conditions. During the cue phase, the 2 s preceding the cue presentation served as a baseline, and data were collapsed over locations and different durations of the cue (6, 8, and 10 s). The object-based attention effect (BOLD amplitude: ICSO - ICDO) of the single cue ($0.09 \pm 0.03, t_{15} = 2.96, p = 0.010$) and double cue ($0.08 \pm 0.02, t_{15} = 4.58, p < 0.001$) were significantly above zero (Fig. 4E, left), as was the spatial attention effect (BOLD amplitude: VC - (ICSO + ICDO)/2; single cue: $0.19 \pm 0.04, t_{15} = 4.89, p < 0.001$; double cue: $0.17 \pm 0.03, t_{15} = 5.91, p < 0.001$, Fig. 4E, right). A repeated-measures ANOVA showed that the main effect of cue type ($F_{1, 15} = 0.21, p = 0.65$) and the interaction between cue type and effect type ($F_{1, 15} = 0.30, p = 0.59$) were not significant. The main effect of effect type was significant ($F_{1, 15} = 13.27, p = 0.002$), showing a greater spatial attention effect than object-based attention effect in V1.

BOLD signals evoked by the targets were extracted from the ROIs in V1 and then selectively averaged according to the different conditions. During the target phase, the 2 s preceding the target presentation served as a baseline to avoid a confound from the cueing period, and data were collapsed over locations and different durations of the fixation (8, 10, and 12 s). The object-based attention effect (BOLD amplitude: ICSO - ICDO) of the single cue ($0.04 \pm 0.01, t_{15} = 2.90, p = 0.01$) and double cue ($0.03 \pm 0.01, t_{15} = 2.20, p = 0.04$) were significantly above zero (Fig. 4G, left), as was the spatial attention effect (BOLD amplitude: VC - (ICSO + ICDO)/2; single cue: $0.21 \pm 0.06, t_{15} = 3.41, p = 0.004$; double cue: $0.18 \pm 0.06, t_{15} = 3.11, p = 0.01$, Fig. 4G, right). A repeated-measures ANOVA showed that the main effect of cue type ($F_{1, 15} = 0.34, p = 0.57$) and the interaction between cue type and effect type ($F_{1, 15} = 0.07, p = 0.79$) were not significant. The main effect of effect type was significant ($F_{1, 15} = 9.16, p = 0.009$), showing a greater spatial attention effect than object-based attention effect in V1. These results support previous studies (e.g., Martínez et al., 2006) by showing a higher neural response to invalid targets belonging to the cued object versus the uncued object.

Moreover, previous studies (Avrahami, 1999; Moore et al., 1998; Müller and Kleinschmidt, 2003; Shomstein and Yantis, 2002) have also suggested that attention can spread from the cued spatial part of an object to other parts within the boundaries of that object when the target is not

Fig. 3. Stimuli and design of Experiment 2. **A** Stimuli and conditions. The stimulus display contained three circular arcs, two of which were presented in the same visual field (in this case, in the right visual field). Single cue: an empty wedge overlapped one end of a circular arc; double cue: two single cues, diametrically opposite to each other, overlapped the ends of two different circular arcs. The target appeared frequently (75% of the trials) at the pre-cued location (Valid Cue, VC). It could also appear either at the uncued end of the pre-cued circular arc (Invalid Cue Same Object, ICSO, 12.5% of the trials), or at the equidistant end of the uncued circular arc (Invalid Cue Different Object, ICDO, 12.5% of the trials). **B** Double cue same hemifield. Two cues were presented in the same visual field. **C** Double cue opposite hemifields. Two cues were presented in the opposite visual field. **D** ROI definition. Checkered wedges were used to define ROIs in human V1. **E** fMRI protocol. The fixation cross was presented for 8, 10, or 12 s, followed by the single cue or the double cue presented for 6, 8, or 10 s. Then a target was presented for 120 ms at one end of a circular arc. Participants pressed one of two buttons as rapidly and correctly as possible to indicate the color of the target (red or green).



detected at the cued location (the invalid target trials, i.e., the target needed to be searched at other locations). There was a strong and preferential activation for locations pertaining to the same object as the initially cued location even without the target. Thus, to examine this effect, we analyzed the valid target and invalid target trials separately, and to exclude stimulus-driven activation by the target, only trials without the target at the corresponding locations were analyzed for each predefined ROI. For example, when the top end of the left circular arc (Fig. 3A) was cued, for the valid target trials (i.e., the target appeared at the same location with the cue), we compared the BOLD signal from the bottom end of the left circular arc (the ROI for the ICSO condition) with that from the left end of the top circular arc (the ROI for the ICDO condition). For the invalid target trials, the BOLD signal from the bottom end of the left circular arc was analyzed (i.e., the response for the ICSO condition) when the target appeared at the left end of the top circular arc; the BOLD signal from the left end of the top circular arc was analyzed (i.e., the response for the ICDO condition) when the target appeared at bottom end of the left circular arc. In valid target trials (i.e., VC), the object-based attention effect (BOLD amplitude: ICSO - ICDO) of the single cue (0.02 ± 0.04 , $t_{15} = 0.52$, $p = 0.61$) and double cue (0.003 ± 0.04 , $t_{15} = 0.07$, $p = 0.95$) conditions were not significantly above zero; however, in invalid target trials, both object-based attention effects were significantly above zero (single cue: 0.17 ± 0.03 , $t_{15} = 4.90$, $p < 0.001$; double cue: 0.14 ± 0.03 , $t_{15} = 4.75$, $p < 0.001$) (Fig. 4I). A repeated-measures ANOVA with cue type (single cue and double cue) and target type (valid and invalid cue) as within-subject factors showed that the main effect of cue type ($F_{1, 15} = 0.28$, $p = 0.60$) and the interaction between cue type and target type ($F_{1, 15} = 0.01$, $p = 0.92$) were not significant. The main effect of target type was significant ($F_{1, 15} = 22.26$, $p < 0.001$), showing a greater object-based attention effect in invalid trials than valid trials. These results are consistent with previous studies by showing that when the target is not detected at the cued location (i.e., the invalid target trials), attention can spread from the cued part of an object to other parts within the boundaries of that object (Moore et al., 1998; Müller and Kleinschmidt, 2003; Shomstein and Yantis, 2002). Together, these results suggest that object-based attention can be simultaneously directed to at least two objects differing in their locations.

3.3. Experiments 1 and 2: attending to two objects versus a single object

To examine potential cortical or subcortical area(s) that showed a significant difference between attending to two objects versus a single object, we performed a group analysis and a whole-brain search with a general linear model (GLM) procedure (Friston et al., 1995) for both Experiments 1 and 2. In Experiment 1, the two-object condition was MFMH and SFSH trials in the Attend Moving and Attend Static scans, respectively; the single-object condition contained both MFSH and SFMH trials in these two different scans. The results showed that only the inferior frontal gyrus bilaterally (IFG, left: -45 ± 1.01 , 27 ± 1.78 , 13 ± 1.52 ; right: 41 ± 1.34 , 19 ± 1.94 , 14 ± 1.27 , $p < 0.01$ with FDR correction) demonstrated a greater response in the two-object condition than the single-object condition in both the Attend Moving ($t_{15} = 3.34$, $p = 0.004$, Fig. 5A, top) and Attend Static scans ($t_{15} = 3.29$, $p = 0.005$, Fig. 5A, bottom). Neither FFA nor PPA demonstrated this effect (Fig. S3). Moreover, both MT+ and V1 demonstrated a greater response in the single-object condition than the two-object condition in the Attend Static scan (Fig. S3B), but not in the Attend Moving scan (Fig. S3A).

In Experiment 2, the two-object and the single-object conditions were

the double and single cue conditions, respectively. The results showed that only early visual cortex (Fig. S3C) and the intraparietal sulcus bilaterally (IPS, left: -20 ± 1.22 , -88 ± 1.56 , 25 ± 1.48 ; right: 27 ± 1.16 , -83 ± 1.49 , 27 ± 1.39 , $p < 0.01$ with FDR correction) demonstrated a greater response in the two-object condition than the single-object condition ($t_{15} = 3.63$, $p = 0.002$, Fig. 5C). Note that early visual cortical areas showed a greater response here since the double cue condition contained more physical stimuli than the single cue condition. No area showed a greater response in the single-object condition than the two-object condition.

In addition, we computed a cost for the object-based attention effect (C_{OBAE}) to quantify how much the object-based attention effect showed in the single-object condition relative to the two-object condition in both Experiments 1 and 2. The cost was calculated as follows: $C_{OBAE} = OBAE_{\text{Single-object}} - OBAE_{\text{Two-object}}$, where $OBAE_{\text{Single-object}}$ and $OBAE_{\text{Two-object}}$ are the object-based attention effect in the single-object and two-object conditions, respectively. Similarly, in Experiment 2, the cost for the spatial attention effect was calculated as follows: $C_{SAE} = SAE_{\text{Single-object}} - SAE_{\text{Two-object}}$, where $SAE_{\text{Single-object}}$ and $SAE_{\text{Two-object}}$ are the spatial attention effect in the single-object and two-object conditions, respectively. In Experiment 1, across individual participants, the enhanced IFG response in the two-object condition relative to the single-object condition significantly predicted the C_{OBAE} in both FFA (Attend Moving scan: $r = -0.69$, $p = 0.003$; Attend Static scan: $r = -0.62$, $p = 0.01$) and PPA (Attend Moving scan: $r = -0.71$, $p = 0.002$; Attend Static scan: $r = -0.61$, $p = 0.01$) (Fig. 5B). In Experiment 2, across individual participants, the enhanced IPS response in the two-object condition relative to the single-object condition significantly predicted both the C_{SAE} ($r = -0.57$, $p = 0.02$) and C_{OBAE} ($r = -0.53$, $p = 0.03$) (Fig. 5D). These results suggest that object-based attention can simultaneously select two objects differing in their features and locations, which may be mediated by the enhanced response in IFG and IPS, respectively.

4. Discussion

We used fMRI to examine whether object-based attention can operate simultaneously on more than one object, i.e., whether task-irrelevant features of multiple objects can be processed simultaneously in object-based attention. In Experiment 1, object-based attention was enhanced by directing attention to a specific feature (the motion direction or static position) of either a single object or two objects simultaneously. When the attended feature was in a single object only (face or house), we replicated prior brain imaging results (O'Craven et al., 1999) by showing a classic object-based attention effect: enhanced activation in face-selective (FFA) and house-selective (PPA) cortical regions depending on which feature was attended. More importantly, we found that simultaneously attending to two objects differing in features (face and house) did not show significantly different responses in FFA or PPA, respectively, compared to attending to a single object (face or house). For example, when participants attended motion direction, no significant difference of the enhanced activation was found between MFSH and MFMH conditions in FFA, or between SFMH and MFMH conditions in PPA. However, a whole-brain analysis revealed that attending to one feature (the motion direction or static position) in the two-object condition relative to the single-object condition showed a greater response in IFG. In Experiment 2, object-based attention was engaged by directing attention to one end of a single circular arc (the single cue condition) or of two different circular arcs (the double cue condition). During the

Fig. 4. Results of Experiment 2. A ROIs in V1 were defined as retinotopic loci corresponding to the six ends of the three circular arcs. B Mean RTs are shown for VC, ICSO, and ICDO in the single (left) and double (right) cue conditions. C Object-based attention effect (left) and spatial attention effect (right) of the single and double cue conditions for RTs. D BOLD amplitude of ROIs in V1 evoked by VC, ICSO, and ICDO in the single (left) and double (right) cue conditions during the cueing phase. E Object-based attention effect (left) and spatial attention effect (right) of the single and double cue conditions during the cueing phase. F BOLD amplitude of ROIs in V1 evoked by the target for VC, ICSO, and ICDO in the single (left) and double (right) cue conditions. G Object-based attention effect (left) and spatial attention effect (right) of the single and double cue conditions during the target phase. H BOLD amplitude of ROIs in V1 evoked by ICSO and ICDO in the single (left) and double (right) cue conditions during the target phase. I Object-based attention effect of the single and double cue conditions for the valid and invalid target conditions. Error bars denote 1 SEM calculated across sixteen participants.

single cue condition, we confirmed both prior psychophysical (Egly et al., 1994; Zhang and Fang, 2012) and brain imaging (Martínez et al., 2006; Müller and Kleinschmidt, 2003; Shomstein and Behrmann, 2006) results by showing a classic object-based attention effect. Participants performed better (Fig. 4B) and showed greater BOLD responses (Fig. 4D) when the invalidly cued target appeared at uncued locations on the cued object

than at equidistant locations on the uncued object. Notably, we did not find a significant difference in object-based attention effects between the single and double cue conditions in human V1. However, a whole-brain analysis revealed that the double cue condition evoked a greater response in IPS than the single cue condition. These results thus provide strong evidence that object-based attention can operate simultaneously on

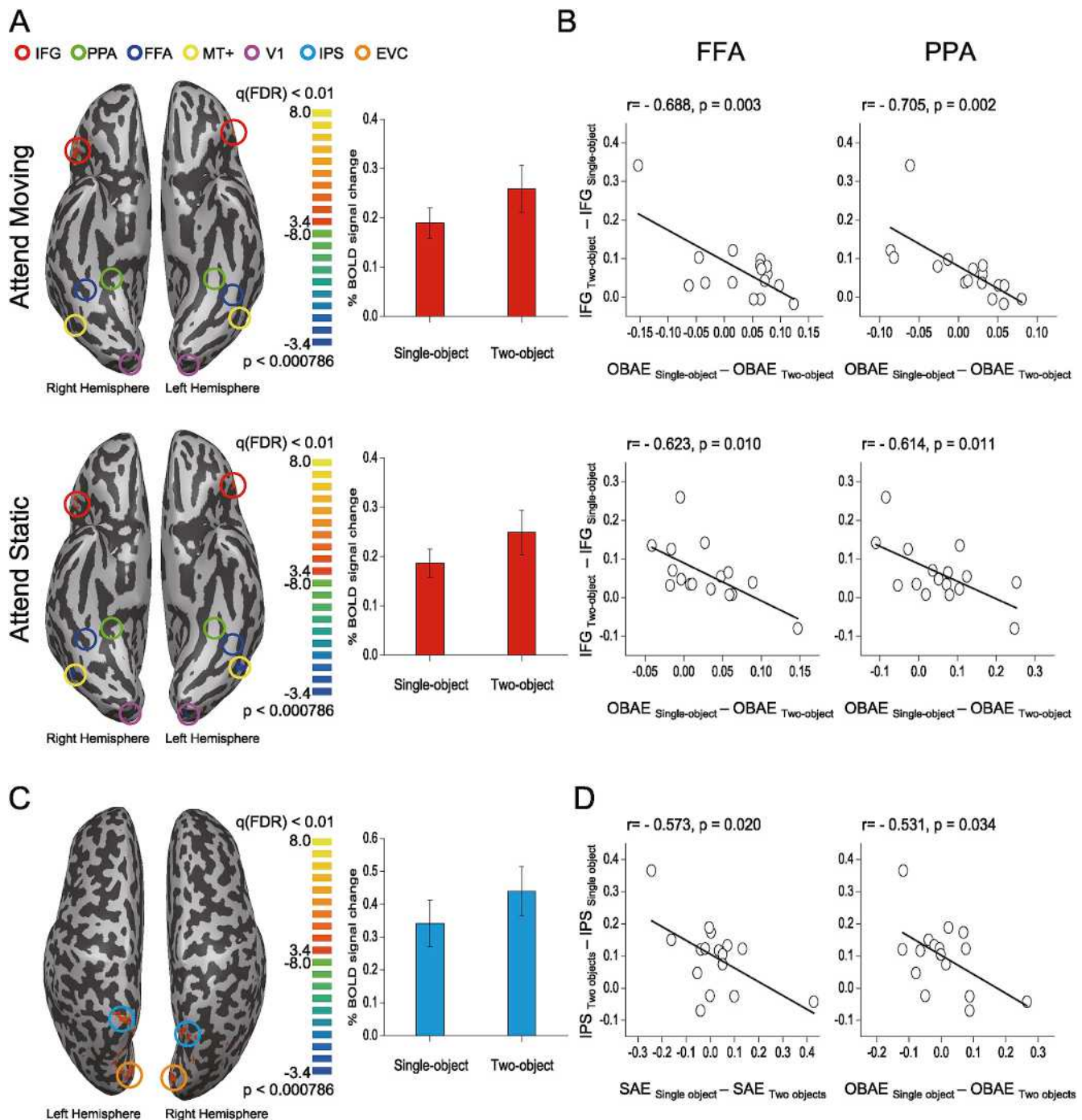


Fig. 5. Results of a whole brain analysis. **A** Experiment 1: The inferior frontal gyrus (IFG, talairach coordinates: Left: $-45 \pm 1.01, 27 \pm 1.78, 13 \pm 1.52$; Right: $41 \pm 1.34, 19 \pm 1.94, 14 \pm 1.27$, $p < 0.01$ with FDR correction) showed a greater response in the two-object condition than the single-object condition in both the Attend Moving (top) and Attend Static (bottom) scans, and their BOLD signal amplitudes. **B** Correlations between the enhanced IFG response in the two-object condition relative to the single-object condition and the C_{OBAE} in FFA (left) and PPA (right). $C_{\text{OBAE}} = \text{OBAE}_{\text{Single-object}} - \text{OBAE}_{\text{Two-object}}$, where $\text{OBAE}_{\text{Single-object}}$ and $\text{OBAE}_{\text{Two-object}}$ are the object-based attention effect in the single-object and two-object conditions, respectively. **C** Experiment 2: The intraparietal sulcus (IPS, talairach coordinates: Left: $-20 \pm 1.22, -88 \pm 1.56, 25 \pm 1.48$; Right: $27 \pm 1.16, -83 \pm 1.49, 27 \pm 1.39$, $p < 0.01$ with FDR correction) showed a greater response in the two-object condition than the single-object condition, and its BOLD signal amplitudes. EVC: early visual cortex. **D** Correlations between the enhanced IPS response in the two-object condition relative to the single-object condition and C_{SAE} (left), and C_{OBAE} (right). $C_{\text{SAE}} = \text{SAE}_{\text{Single-object}} - \text{SAE}_{\text{Two-object}}$, where $\text{SAE}_{\text{Single-object}}$ and $\text{SAE}_{\text{Two-object}}$ are the spatial attention effect in the single-object and two-object conditions, respectively. Error bars denote 1 SEM calculated across sixteen participants.

multiple objects in the human visual system, i.e., task-irrelevant features of multiple objects can be processed simultaneously in object-based attention. Our results also imply a crucial involvement of IFG and IPS in this process induced by feature- and location-selection, respectively. We acknowledge that our conclusion relies on a null difference between the single-object and two-object conditions in Experiments 1 and 2. However, both conditions showed significant object-based attention effects and, more importantly, the fronto-parietal attention network showed a significantly greater activation in the two-object condition than the single-object condition. Hence, our results reflect positive findings.

In addition, our results cannot be explained by the rapid switching of attention between the features or locations of the two objects (Wager et al., 2004). In Experiment 1, participants were asked to detect the consecutive repetition of either the motion direction or the static position of the test objects. Behavioral data showed no significant difference in subject performance between the single-object and two-object conditions, which would have occurred in the presence of rapid switching. Moreover, this task required participants to attend a given visual feature (motion direction or static position); the other, unattended features (e.g., face and house shape) were never task-relevant. Thus, participants did not need to switch their attention between these unattended features. In Experiment 2, the double cue contained two single cues on different circular arcs, presented either within the same (Fig. 3B) or opposite hemifields (Fig. 3C). Previous studies have proposed that shifts in attention are much easier within the same hemifield than across hemifields (Ibos et al., 2009). Accordingly, for both RT and V1 BOLD signal changes, we compared the object-based attention effect and the spatial attention effect on the same and opposite hemifield trials and found no significant difference for either effect (Fig. S2). If participants rapidly switched their attention between two cued circular arcs, then we should have observed a stronger object-based attention effect and spatial attention effect in the same hemifield trials than the opposite hemifield trials.

Our finding of multiple object-based attention is consistent with previous studies suggesting that the human visual system is able to process multiple objects, including multiple object recognition in cluttered scenes (Riesenhuber and Poggio, 1999), tracking of multiple objects (Cavanagh and Alvarez, 2005), as well as holding multiple objects in short-term (working) memory (Luck and Vogel, 1997; Wheeler and Treisman, 2002; Xu and Chun, 2006, 2009), feature binding (Treisman, 1996), and feature misbinding (Zhang et al., 2014). It is important to note that, in those studies, all the features of the objects were task-relevant (attended). Attention to these task-relevant features may reflect a feature-similarity gain model (Treue and Martinez-Trujillo, 1999), whereby feature-based attention enhances the gain of cortical neurons tuned to the attended feature anywhere in the visual field, and this similarity might be based on the spatial location or any other feature (Maunsell and Treue, 2006). However, in our study, the task required participants to attend a given visual feature, while we measured the representation of the other unattended features (namely, the face and house shape in Experiment 1, and the uncued location in Experiment 2) of the same objects. Thus, our findings cannot be explained by feature-based attention.

In Experiment 1, our results supported the predictions of the integrated competition (Duncan et al., 1997) and incremental grouping (Roelfsema, 2006) models. The integrated competition model proposes that attending to one feature of an object produces top-down biasing signals that result in an enhanced response in the cortical area that represents the attended feature, which is then conveyed to the cortical areas that represent other features of the same object. The network of these cortical areas can then bind both attended and unattended features into a unified percept of the object. This model has been extended by the incremental grouping model, which similarly proposes that the enhanced neural responses spread across the network via recurrent connections across cortical areas to bind the features of the same object (Roelfsema, 2006). Our results extend these models by showing that at least two

objects can be constructed simultaneously as separate perceptual units through selection and binding of the attended and unattended features in their respective cortical areas. Moreover, we found that IFG is likely to be responsible for this top-down feature biasing that allows two objects to be constructed simultaneously as separate perceptual units. First, IFG showed a greater response in the two-object condition than the single-object condition (Fig. 5A). Second, the enhanced response in IFG significantly predicted the cost ($C_{OBAE} = OBAE_{\text{Single-object}} - OBAE_{\text{Two-object}}$, where $OBAE_{\text{Single-object}}$ and $OBAE_{\text{Two-object}}$ are the object-based attention effect in the single-object and two-object conditions, respectively) of the object-based attention effect (Fig. 5B) in the two-object condition relative to the single-object condition. Combined with a recent study (Baldauf and Desimone, 2014) reporting that human inferior frontal junction (IFJ) could play a key role in attention-biased perception through neural synchrony with FFA and PPA, we thus proposed that object-based attention induced by feature-selection was mediated by the inferior frontal cortex.

In Experiment 2, our results confirmed previous studies (McMains and Somers, 2004; Müller et al., 2003) by showing that spatial attention could be split between at least two discrete regions of space without a cost to the spatial attention effect. Remarkably, we found that attention directed to two discrete locations of two objects can spread simultaneously throughout the entire object and facilitate the processing of unattended regions located within the boundaries of those two objects. Furthermore, we found that this spread of attention throughout two objects was mediated by the IPS. First, IPS showed a greater response in the two-object condition than the single-object condition (Fig. 5C). Second, the enhanced response in IPS significantly predicted the cost of the object-based attention effect (C_{OBAE} , Fig. 5D) in the two-object condition relative to the single-object condition. In other words, object-based attention can simultaneously select two objects differing in their locations, which may be mediated by the enhanced response in IPS. Our results are consistent with previous studies (Xu and Chun, 2006, 2009) showing an involvement of the IPS in object individuation via their locations.

In addition, numerous previous studies have suggested that the frontal/prefrontal and parietal cortical areas show increased activity with short-term (working) memory load (e.g., Cohen et al., 1997; Courtney et al., 1997; Curtis and D'Esposito, 2003; Pessoa et al., 2002; Todd and Marois, 2004; Vogel and Machizawa, 2004; Xu and Chun, 2006, 2009). In Experiment 2, we used a modified version of the classical cuing paradigm developed by Egly et al. (1994), and participants likely maintained the cued (attended) objects in short-term (working) memory to guide or facilitate their response to the targets. Compared to the single cue condition, two circular arcs were cued (attended) in the double cue condition (Fig. 3A), which would indeed increase the memory load. Thus, the enhanced response in IPS during the double cue compared to the single cue condition might be caused by increased short-term (working) memory load. However, the memory load cannot explain the enhanced response in IFG in Experiment 1. In this experiment, participants attended a given visual feature (motion direction or static position), while we measured activation evoked by other, task-irrelevant features (face and house shape) in both the single-object and two-object conditions. Moreover, on each trial, both the single-object (face or house) and the two-object (face and house) conditions had only one motion direction or static position. Note that the moving face and moving house (MFMH, the two-object condition in the Attend Moving scan) moved in the same direction; the static face and static house (SFSH, the two-object condition in the Attend Static scan) were located at the same position. Thus, the attended feature (motion direction or static position) was exactly the same the single-object and two-object conditions. Hence, there was no difference in memory load between these two conditions. Therefore, the enhanced response in IFG in Experiment 1 cannot be explained by increased memory load.

Based on our findings, three intriguing questions need to be addressed in future research: 1) What are the limits on the number of distinct

objects that can be attended simultaneously? 2) What is the cost of attending to multiple objects? 3) In what ways does object-based attention due to feature- and location-selection differ? Investigations of these questions may consider Xu and Chun (2006, 2009) proposal, suggesting that our visual system first selects a fixed number of about four objects from a crowded scene based on their locations and then encodes a variable subset of the attended objects, depending on their feature complexity and the encoding demands of the task-relevant feature. Xu (2010) showed that the encoding of a task-irrelevant shape feature in object-based encoding could be seen with no more than two objects when there is a high encoding demand of the task-relevant feature.

In sum, our study has three novel aspects. First, our study provides strong evidence that object-based attention can operate simultaneously on task-irrelevant features of at least two objects. This finding adds a new dimension to the literature on object-based attention. Second, our study extends Duncan et al.'s integrated competition model (1997) and Roelfsema's incremental grouping model (2006) by showing that at least two objects can be constructed simultaneously as separate perceptual units through selection and binding of the attended and unattended features in their respective cortical areas. Finally, our study implicates, to the best of our knowledge, for the first time, distinct roles for frontal and parietal cortices in object-based attention induced by feature- and location-selection, respectively. This aspect of our study is consistent with recent neurophysiological findings that have begun to address the relative contributions of frontal and parietal cortex in visual attention (Suzuki and Gottlieb, 2013). Our findings open up a new avenue of research regarding how many distinct objects can be constructed as separate perceptual units simultaneously and how the fronto-parietal attention network may be involved in this process.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

We wish to thank John Ingeholm for eye tracking support. This work was supported by the National Institute of Mental Health Intramural Research Program (NIH Clinical Study Protocol 93-M-0170, NCT00001360, ZIAMH002918-09).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.neuroimage.2017.09.050>.

References

- Andersen, S.K., Hillyard, S.A., Müller, M.M., 2013. Global facilitation of attended features is obligatory and restricts divided attention. *J. Neurosci.* 33, 18200–18207.
- Avrami, J., 1999. Objects of attention, objects of perception. *Percept. Psychophys.* 61, 1604–1612.
- Baldauf, D., Desimone, R., 2014. Neural mechanisms of object-based attention. *Science* 344, 424–427.
- Brefczynski, J.A., DeYoe, E.A., 1999. A physiological correlate of the “spotlight” of visual attention. *Nat. Neurosci.* 2, 370–374.
- Buracas, G.T., Boynton, G.M., 2002. Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage* 16, 801–813.
- Cavanagh, P., Alvarez, G.A., 2005. Tracking multiple targets with multifocal attention. *Trends Cogn. Sci.* 9, 349–354.
- Chen, C., Zhang, X., Wang, Y., Zhou, T., Fang, F., 2016. Neural activities in V1 create the bottom-up saliency map of natural scenes. *Exp. Brain Res.* 234, 1769–1780.
- Chen, Z., 2012. Object-based attention: a tutorial review. *Atten. Percept. Psychophys.* 74, 784–802.
- Chou, W.L., Yeh, S.L., 2012. Object-based attention occurs regardless of object awareness. *Psychon. Bull. Rev.* 19, 225–231.
- Cohen, E.H., Tong, F., 2013. Neural mechanisms of object-based attention. *Cereb. Cortex* 25, 1080–1092.
- Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E., 1997. Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.
- Courtney, S.M., Ungerleider, L.G., Keil, K., Haxby, J.V., 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386, 608.
- Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423.
- Duncan, J., Humphreys, G., Ward, R., 1997. Competitive brain activity in visual attention. *Curr. Opin. Neurobiol.* 7, 255–261.
- Egley, R., Driver, J., Rafal, R.D., 1994. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.* 123, 161.
- Engel, S.A., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7, 81–192.
- Fiebelkorn, I.C., Saalman, Y.B., Kastner, S., 2013. Rhythmic sampling within and between objects despite sustained attention at a cued location. *Curr. Biol.* 23, 2553–2558.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 211–224.
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878.
- Ibos, G., Duhamel, J.R., Hamed, S.B., 2009. The spatial and temporal deployment of voluntary attention across the visual field. *PLoS One* 4, e6716.
- Jiang, J., Summerfield, C., Eger, T., 2016. Visual prediction error spreads across object features in human visual cortex. *J. Neurosci.* 36, 12746–12763.
- Kanwisher, N., Wojciulik, E., 2000. Visual attention: insights from brain imaging. *Nat. Rev. Neurosci.* 1, 91–100.
- Kastner, S., Ungerleider, L.G., 2000. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341.
- Kourtzi, Z., Kanwisher, N., 2000. Cortical regions involved in perceiving object shape. *J. Neurosci.* 20, 3310–3318.
- Liu, T., Larsson, J., Carrasco, M., 2007. Feature-based attention modulates orientation selective responses in human visual cortex. *Neuron* 55, 313–323.
- Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.
- Martínez, A., Anillo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., Hillyard, S.A., 1999. Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369.
- Martínez, A., Teder-Sälejärvi, W., Vazquez, M., Molholm, S., Foxe, J.J., Javitt, D.C., Di Russo, F., Worden, M.S., Hillyard, S.A., 2006. Objects are highlighted by spatial attention. *J. Cogn. Neurosci.* 18, 298–310.
- Maunsell, J.H., Treue, S., 2006. Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322.
- McMains, S.A., Somers, D.C., 2004. Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42, 677–686.
- Moore, C., Yantis, S., Vaughan, B., 1998. Object-based visual selection. Evidence from perceptual completion. *Psychol. Sci.* 9, 104–110.
- Müller, N.G., Kleinschmidt, A., 2003. Dynamic interaction of object-and space-based attention in retinotopic visual areas. *J. Neurosci.* 23, 9812–9816.
- Müller, M.M., Malinowski, P., Gruber, T., Hillyard, S.A., 2003. Sustained division of the attentional spotlight. *Nature* 424, 309–312.
- O'Craven, K.M., Downing, P.E., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. *Nature* 401, 584–587.
- Pessoa, L., Gutierrez, E., Bandettini, P.A., Ungerleider, L.G., 2002. Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron* 35, 975–987.
- Pooresmaeili, A., Roelfsema, P.R., 2014. A growth-cone model for the spread of object-based attention during contour grouping. *Curr. Biol.* 24, 2869–2877.
- Posner, M.I., Snyder, C.R.R., Davidson, B.J., 1980. Attention and the detection of signals. *J. Exp. Psychol.* 109, 160–174.
- Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Roelfsema, P.R., 2006. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227.
- Roelfsema, P.R., Lamme, V.A., Spekreijse, H., 1998. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381.
- Saenz, M., Buracas, G.T., Boynton, G.M., 2002. Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* 5, 631–632.
- Schoenfeld, M.A., Hopf, J.M., Merkel, C., Heinze, H.J., Hillyard, S.A., 2014. Object-based attention involves the sequential activation of feature-specific cortical modules. *Nat. Neurosci.* 17, 619–624.
- Scholl, B.J., 2001. Objects and attention: the state of the art. *Cognition* 80, 1–46.
- Serences, J.T., Boynton, G.M., 2007. Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55, 301–312.
- Serences, J.T., Schwarzbach, J., Courtney, S.M., Golay, X., Yantis, S., 2004. Control of object-based attention in human cortex. *Cereb. Cortex* 14, 1346–1357.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B.H., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Shomstein, S., Behrmann, M., 2006. Cortical systems mediating visual attention to both objects and spatial locations. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11387–11392.
- Shomstein, S., Yantis, S., 2002. Object-based attention: sensory modulation or priority setting? *Percept. Psychophys.* 64, 41–51.
- Smith, A.M., Lewis, B.K., Ruttimann, U.E., Ye, F.Q., Sinnwell, T.M., Yang, Y., Duyn, J., Frank, J.A., 1999. Investigation of low frequency drift in fMRI signal. *Neuroimage* 9, 526–533.

- Suzuki, M., Gottlieb, J., 2013. Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat. Neurosci.* 16, 98–104.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: an Approach to Cerebral Imaging* (New York: Thieme).
- Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751.
- Tootell, R.B., Hadjikhani, N., Hall, E.K., Marrett, S., Vanduffel, W., Vaughan, J.T., Dale, A.M., 1998. The retinotopy of visual spatial attention. *Neuron* 21, 1409–1422.
- Treisman, A., 1996. The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178.
- Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.
- Treue, S., Martinez-Trujillo, J.C., 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579.
- Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748.
- Wager, T.D., Jonides, J., Reading, S., 2004. Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* 22, 1679–1693.
- Wheeler, M.E., Treisman, A.M., 2002. Binding in short-term visual memory. *J. Exp. Psychol. Gen.* 131, 48.
- Xu, Y., 2010. The neural fate of task-irrelevant features in object-based processing. *J. Neurosci.* 30, 14020–14028.
- Xu, Y., Chun, M.M., 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95.
- Xu, Y., Chun, M.M., 2009. Selecting and perceiving multiple visual objects. *Trends Cogn. Sci.* 13, 167–174.
- Zhang, W., Luck, S.J., 2009. Feature-based attention modulates feedforward visual processing. *Nat. Neurosci.* 12, 24–25.
- Zhang, X., Fang, F., 2012. Object-based attention guided by an invisible object. *Exp. Brain Res.* 223, 397–404.
- Zhang, X., Japee, S., Safiullah, Z., Mlynaryk, N., Ungerleider, L.G., 2016. A normalization framework for emotional attention. *PLoS Biol.* 14, e1002578.
- Zhang, X., Qiu, J., Zhang, Y., Han, S., Fang, F., 2014. Misbinding of color and motion in human visual cortex. *Curr. Biol.* 24, 1354–1360.
- Zhang, X., Zhaoping, L., Zhou, T., Fang, F., 2012. Neural activities in V1 create a bottom-up saliency map. *Neuron* 73, 183–192.