Object-based attention to one of two superimposed surfaces alters responses in human early visual cortex

Vivian M. Ciaramitaro, Jude F. Mitchell, Gene R. Stoner, John H. Reynolds and Geoffrey M. Boynton

Object-based attention to one of two superimposed surfaces alters responses in human early visual cortex

Vivian M. Ciaramitaro,[1,2] **Jude F. Mitchell,**[3] **Gene R. Stoner,**[4] **John H. Reynolds,**[3]
and Geoffrey M. Boynton[5]

[1]*Department of Psychology, University of Massachusetts, Boston, Massachusetts;* [2]*Department of Psychology, University of California, San Diego, La Jolla;* [3]*Systems Neurobiology Laboratory and* [4]*Vision Center Laboratory, The Salk Institute of Biological Studies, La Jolla, California; and* [5]*Department of Psychology, University of Washington, Seattle, Washington*

Ciaramitaro VM, Mitchell JF, Stoner GR, Reynolds JH, Boynton GM. Object-based attention to one of two superimposed surfaces alters responses in human early visual cortex. *J Neurophysiol* 105: 1258–1265, 2011. First published January 12, 2011; doi:10.1152/jn.00680.2010.—Faced with an overwhelming amount of sensory information, we are able to prioritize the processing of select spatial locations and visual features. The neuronal mechanisms underlying such spatial and feature-based selection have been studied in considerable detail. More recent work shows that attention can also be allocated to objects, even spatially superimposed objects composed of dynamically changing features that must be integrated to create a coherent object representation. Much less is known about the mechanisms underlying such object-based selection. Our goal was to investigate behavioral and neuronal responses when attention was directed to one of two objects, specifically one of two superimposed transparent surfaces, in a task designed to preclude space-based and feature-based selection. We used functional magnetic resonance imaging (fMRI) to measure changes in blood oxygen level-dependent (BOLD) signals when attention was deployed to one or the other surface. We found that visual areas V1, V2, V3, V3A, and MT+ showed enhanced BOLD responses to translations of an attended relative to an unattended surface. These results reveal that visual areas as early as V1 can be modulated by attending to objects, even objects defined by dynamically changing elements. This provides definitive evidence in humans that early visual areas are involved in a seemingly high-order process. Furthermore, our results suggest that these early visual areas may participate in object-specific feature "binding," a process that seemingly must occur for an object or a surface to be the unit of attentional selection.

surface-based selection; spatial attention; feature-based attention; human visual cortex

VISUAL ATTENTION allows prioritized processing of behaviorally relevant stimuli. Most studies of attention have focused on spatial or feature-based attention. However, over the past two decades it has become increasingly clear that attention can select not only locations and/or features but also entire objects and/or surfaces. In such object-based selection, if one feature of an object is selected, other features of that object are also selected. This selection has been shown to persist even if the selected object's features change over time.

Studies of object-based attention show that attention can be directed to objects defined by the conjunction of low-level, superimposed features, such as color and motion, which are constantly changing (Blaser et al. 2000), or by the integration of locally coplanar elements to create a surface (He and Nakayama 1995). Attention can also select one of two super-imposed surfaces (surface-based selection), even if the properties (e.g., motion direction) of those surfaces change unpredictably over time, thereby ruling out feature-based selection (Valdes-Sosa et al. 1998a). Although the neural mechanisms mediating spatial (reviewed in Reynolds and Chelazzi 2004) and feature-based selection (reviewed in Maunsell and Treue 2006) have been studied extensively, less is known about the neural mechanisms mediating object or surface-based selection.

Functional magnetic resonance imaging (fMRI) studies of object-based selection have reported that the activity in face- and house-selective areas of inferotemporal cortex is modulated when human observers shift attention between superimposed images of faces and houses (O'Craven et al. 1999; Serences et al. 2004). In such experiments, observers are selecting between different classes of objects, with unique visual areas responsive to each class. Object-based selection has also been investigated in event-related potential (ERP) studies using superimposed moving random-dot patterns. These studies have used variations of the psychophysical paradigm developed by Valdes-Sosa et al. (1998a) and compared ERPs to translations of attended surfaces to those of unattended surfaces. Unlike objects such as faces and houses, these surfaces require the binding together of dynamically changing elements into a single cohesive object and do not allow for object selection based on object class; both objects are composed of the same types of elements, and both are members of the same class (surfaces).

These ERP studies of surface-based attention have yielded somewhat conflicting results regarding the earliest processing stage at which surface-based attention effects are found. In particular, in the original ERP study of Valdes-Sosa et al. (1998b), endogenous cueing was found to modulate the P1 component, which is associated with early extrastriate visual areas. A follow-up study by Pinilla et al. (2001) did not, however, find modulation of that early component. More recently, Khoe et al. (2005), using only an exogenous cue, found modulation of the even earlier C1 component, which is usually associated with striate cortex. That study built on the results of Reynolds et al. (2003), who had earlier found that an exogenous cue was present in the original Valdes-Sosa paradigm and that this exogenous cue was sufficient to elicit a surface-based psychophysical advantage. The timing and scalp distribution of Khoe et al.'s (2005) C1 component may, however, have reflected an extrastriate origin.

Address for reprint requests and other correspondence: V. M. Ciaramitaro, Univ. of Massachusetts, Boston, Dept. of Psychology, 100 Morrissey Boulevard, Boston, MA 02125 (e-mail: vivian.ciaramitaro@umb.edu).

Consistent with the above ERP studies, single-unit physiological studies using transparent-motion stimuli have found object-based attention effects in extrastriate visual areas: Fallah et al. (2007) found enhanced responses to the color of exogenously cued versus uncued surfaces in area V4, and Wannig et al. (2007) found enhanced responses to translations of endogenously cued versus uncued surfaces in MT. To date, however, there has been no single-unit examination of object-based effects in earlier areas (i.e., V1 or V2) using stimuli that rule out spatial selection. Given that ERP studies have yielded somewhat conflicting data and, more generally, have difficulty localizing brain responses precisely, it is unknown which early visual cortical areas are modulated by surface-based attention.

The question of whether early visual areas are influenced by object-based attention is of particular interest, because one might reasonably suppose that object-based attention is mediated by the anterior visual areas that are thought to process complex objects (Gross et al. 1972; Desimone et al. 1984; Damasio et al. 1982). Consistent with this, neuroimaging studies have indeed found object-based attention effects in higher order visual areas (Kanwisher et al. 1997). We asked whether object-based attention might also modulate responses of neurons in early visual areas, including primary visual cortex.

In the present study, we modified the exogenous cueing paradigm used by Reynolds et al. (2003) to measure surface-based selection in a paradigm that precludes space-based and feature-based selection. We used fMRI to determine which early visual areas were modulated by exogenously cued surface-based selection, our measure of object-based attention.

MATERIALS AND METHODS

Subject selection. Subjects were undergraduate, graduate, or postdoctoral students recruited from the Salk Institute or the University of California, San Diego (UCSD) community (4 females, 3 males, 23–32 yr old). Experiments were approved by the Institutional Review Boards of the Salk Institute and UCSD. Written informed consent was obtained from all subjects. Subjects had normal or corrected-to-normal vision and no psychiatric or neurological history.

Stimulus generation and presentation. Stimuli were generated using an Apple Macintosh computer (Powerbook G3 processor, 300 MHz) with the Psychophysics Toolbox (Brainard 1997; Pelli 1997), the Video Toolbox, and the MATLAB (version 5.2) programming language. In the laboratory, visual stimuli were projected via a liquid crystal display video projector (NEC model LT157, 640 × 480 pixels, 60 Hz; maximum luminance 400 cd/m^2) onto a backprojection screen, producing a 52 × 40-cm image 1 m away from the observer. Subjects sat 57 cm distant from the screen while seated upright, with a chin rest and forehead rest stabilizing head position. In the MRI scanner, subjects viewed the screen 57 cm away through an angled front-surfaced mirror mounted on the head coil while lying supine with a bite-bar stabilizing head position. Viewing conditions were the same in the scanner and laboratory, except for the above-mentioned differences.

Experimental design. Subjects maintained gaze on a central fixation point while two overlapping fields of randomly positioned red dots were presented; one red dot field rotated clockwise, whereas the other rotated counterclockwise (see Fig. 1). The dots defining the two surfaces were of maximum contrast with the gun values set to [0, 0, 0] for the background and [255, 0, 0] for the dots. These moving dot fields yielded a percept of two superimposed surfaces. After a fixed duration of rotation (*rotation interval 1*), one of the dot fields underwent a brief (100 ms) translation, in which 60% of the dots moved

coherently in one of four possible directions along the noncardinal axes while the remaining 40% of dots on this dot field moved incoherently. During this period, the second dot field continued rotating. From trial to trial, the surface that translated and the direction of translation were unpredictable. After the first translation, both dot fields resumed rotation (*rotation interval 2*), after which a second translation occurred (100 or 50 ms), with 60% of the dots on a given dot field moving coherently in one of the four possible directions. The second translation was of either surface, selected at random, and the direction of translation was randomized. After this second translation, both dot fields resumed rotation (*rotation interval 3*). The period of rotational motion that followed each period of translational motion served to mask each translational motion stimulus. Subjects could respond at any time after the start of the second translation, within 500 ms after the onset of *rotation interval 3*. They reported the direction of the second translation by pressing one of four keys. We refer to the condition where the first and second translations are of the same surface as our cued condition and that where they are of different surfaces as our uncued condition.

All observers were extensively trained on the task so that overall performance was well above the chance level of 25% correct. We required observers to achieve a minimum of 60% correct before testing on the same task in the scanner. This practice ensured that behavior during scanning would be well controlled. All behavioral analyses were conducted on data collected during scanning.

The task was difficult and used a very rapid sequence of events. This experimental design, with the duration of the second translation accounting for ~5–10% of total trial duration, helped maximize fMRI responses to the second translation. Many subjects had difficulty learning the task, making extensive training necessary. Eight additional observers were excluded due to their failure to learn the task, obtaining less than 60% correct after 4–5 days of training on the experiment. Previous studies using a closely related paradigm have imposed a minimum level of performance of 70–75% correct (Mitchell et al. 2003; Lopez et al. 2004) to exclude subjects who either could not understand the task or could not perceive the surfaces clearly. Our task was more challenging, so we imposed a somewhat more lenient threshold of 60% correct, but well above the performance expected by chance (25% correct). This criterion provided a disincentive for subjects to track individual dots, since only 60% of the dots translated coherently.

During training, feedback was provided at the end of each trial. The color at central fixation indicated whether responses were correct, incorrect due to the wrong direction of motion being selected, or incorrect due to the response being too slow (green, red, and purple, respectively). Trials for which subjects were too slow to respond accounted for only 3.8% of all trials (±1.4% SE). No feedback was provided for the data collected for analysis during scanning, since this could have influenced the blood oxygen level-dependent (BOLD) signal.

For both training and testing, experiments were carried out in a darkened room. Each dot field had an average dot density of five dots per square degree of visual arc. Each dot subtended 0.03°. Dot fields rotated at 50°/s. To discourage the tracking of individual dots, a red circle surrounded fixation, and only 60% of the dots underwent translational motion. Thus dot tracking was difficult since no dots were present in the vicinity of fixation, and dot tracking was only minimally beneficial since only a subset of dots moved in the correct direction. Subjects were instructed to discriminate the translation.

fMRI acquisition. MR images for retinotopic mapping, high-resolution anatomical scans, and functional scans were acquired on a Signa EXCITE 3 Tesla GE "short-bore" scanner at the Center for Functional Magnetic Resonance Imaging at the UCSD medical school campus. Data were collected using an eight-channel head coil (parallel imaging was not used). Functional images were acquired using a ±62.5 kHz acquisition bandwidth T2*-weighted echo planar imaging (EPI) pulse sequence in 23 axial slices (TR = 1,250 or 1,050 ms,

TE = 30 ms, flip angle = 90°, FOV = 256 mm, slice thickness = 4 mm, in-plane resolution = 4 × 4 mm). The first 20 or 21 s of data were discarded to avoid magnetic saturation affects and to allow subjects to reach a stable behavioral state.

Each scanning session started with a localizer scan (described below) using the same pulse sequences described above. The localizer scan was used to constrain the activation considered for analysis within each region of interest (ROI). Each scanning session ended with a low-resolution anatomical scan, a T1-weighted gradient echo pulse sequence (SPGR; TE = 5 ms, flip angle = 50°, FOV = 256 mm, resolution = 1 × 1 × 4 mm). This low-resolution anatomical scan was used to coregister functional data gathered across days to each observer's high-resolution anatomical scan acquired once on the first day of scanning (SPGR; TE = 5 ms, flip angle = 50°, FOV = 256 mm, resolution = 1 × 1 × 1 mm).

Retinotopic mapping and ROI-based analysis. We used standard retinotopic mapping, cortical segmentation, and flattening techniques to define early visual areas (see Ciaramitaro et al. 2007 for details). The boundaries of each visual area were delineated on a flat map based on borders between phase transitions. Regions of activation within each subject's uniquely defined ROI were constrained based on a localizer scan. Voxels within each ROI were considered for analysis if their activity correlated with the localizer scan at a correlation threshold of 0.35. Data were collapsed across corresponding regions of the left and right cortex.

fMRI experimental design. We used a block design for localizer scans and a rapid event-related design for attention scans. Each localizer scan contained 6 cycles of 40- or 42-s blocks alternating between 2 conditions every 20 or 21 s (96 or 120 trials/condition, respectively). A flickering checkerboard (same visual angle and eccentricity as stimuli in the attention experiment) alternated with a uniform gray field of the same mean luminance. For the attention scans, three trial types were presented in each scan with order counterbalanced using an m-sequence (Buracas and Boynton 2002): *1*) first and second translation of the same surface ("cued"), *2*) first and second translation of different surfaces ("uncued"), or *3*) fixation alone. Motion direction for translational and rotational motion was randomized from trial to trial, outside of the m-sequence used to randomize across trial types. Each scan lasted 626 or 526 s, 250 trials total, and 2,500 or 2,100 ms/trial, respectively. Subjects completed one to two localizer scans and four to six attention scans per scanning session.

Measuring fMRI response amplitudes. For each scan, we estimated the peak amplitude of the fMRI response for each of the two attention conditions as follows. *1*) fMRI time series were averaged across the subset of voxels within each ROI that showed modulation with the localizer scan with a correlation of 0.35 or higher. *2*) Linear deconvolution (Dale 1999) was then applied to each averaged time series to provide an estimated hemodynamic response function (HDR) for the two attentional conditions. *3*) The average of this HDR across the two attentional conditions was fit with a parametric HDR function (difference of gamma) to minimize least squared error. *4*) The HDRs for the two attentional conditions were then refit with this parametric HDR, allowing only the amplitude parameters to vary. *5*) The peaks of these best-fitting parametric fits were calculated for each of the two attention conditions and served as our dependent measure. This within-scan measure provides a more reliable estimate of peak amplitude than measures of HDRs averaged across scans, because differences in the time of the peak across scans and subjects result in underestimation of effect size.

Statistical analysis. We used both ratio and nonratio measures in our behavioral and BOLD analyses (see RESULTS). Ratio measures were artificially bounded, and a Shapiro-Wilk *W* test confirmed that these data were not normally distributed. Accordingly, we used nonparametric statistics (2-tailed Wilcoxon signed rank test). Although application of the Shapiro-Wilk *W* test to our nonratio data (percent correct, reaction time, etc.) did not demonstrate deviation

from a normal distribution, for consistency we again quantified significance using the two-tailed paired Wilcoxon signed rank test. The use of nonparametric statistics to analyze these data also ensured a conservative estimate of significance.

RESULTS

Subjects performed a task that has been found to engage surface-based attention. They maintained fixation while viewing two overlapping dot fields. One dot field rotated clockwise while the other rotated counterclockwise, yielding a percept of two superimposed surfaces (Fig. 1). Subsequent to this period of rotation, one of the dot fields underwent a brief translation (60% of the dots of 1 of the dot fields moved coherently in 1 of 4 possible directions). Subjects could not predict from trial to trial which surface would translate or the direction of the translation. Previous studies have found that this brief motion transient acts as a potent exogenous cue, which automatically draws attention to the translating surface (first demonstrated in Reynolds et al. 2003). After this first translation, both dot fields resumed rotation, after which there was an unpredictable second translation of one of the surfaces, followed by a rotation to mask the translational motion. Subjects had to report the direction of the second translation. The condition where the first and second translations are of the same surface is our cued condition, and where they are of different surfaces is our uncued condition.

Behavioral results: surface-based attention effects on behavior. Observers practiced the task in the laboratory before scanning took place. Seven subjects completed a total of 22 ~1.5-h practice sessions (28,250 practice trials) over 2–5 days (mean = 3.14 days; 8–30 repeats/subject; 250 trials/repeat). All behavioral data presented below were acquired during scanning, while we simultaneously measured fMRI responses. The 7 observers completed a total of 97 scans, with 24,250 trials (1,750–4,000 trials/observer) over 1–5 days (mean = 3.29 days).

Consistent with previous reports using similar paradigms, subjects discriminated the direction of the second translation more accurately for the cued vs. the uncued surface. Figure 2*A* plots mean percent correct averaged across all scans for a single subject (S3) when the second translation occurred on the cued vs. the uncued surface. This subject, like our other subjects, was more accurate at discriminating the direction of the second translation when it occurred on the cued vs. the uncued surface. The effect of cuing on performance for each of the individual subjects tested is shown in Fig. 2*B*. Each of the individual subjects showed a significant improvement in percent correct performance for the cued condition relative to the uncued condition (2-tailed paired Wilcoxon signed rank test: S1, $P = 0.001$; S2, $P = 0.013$; S3, $P < 0.0001$; S4, $P = 0.001$; S5, $P < 0.0001$; S6, $P = 0.001$; S7, $P = 0.016$).

We further quantified our behavioral effects using a metric that normalizes for differences in the absolute magnitude of effects across subjects and that provides a single composite measure of the performance difference between cued and uncued conditions. Specifically, we computed a percent correct attention index: for each subject and each scan, the behavioral measure for the uncued surface (U) was subtracted from the measure for the cued surface (C) and divided by the sum of these measures [$(C - U)/(C + U)$]. Figure 2*C* plots the mean attention index for each subject for percent correct. Subjects showed improved accuracy for the cued relative to the uncued
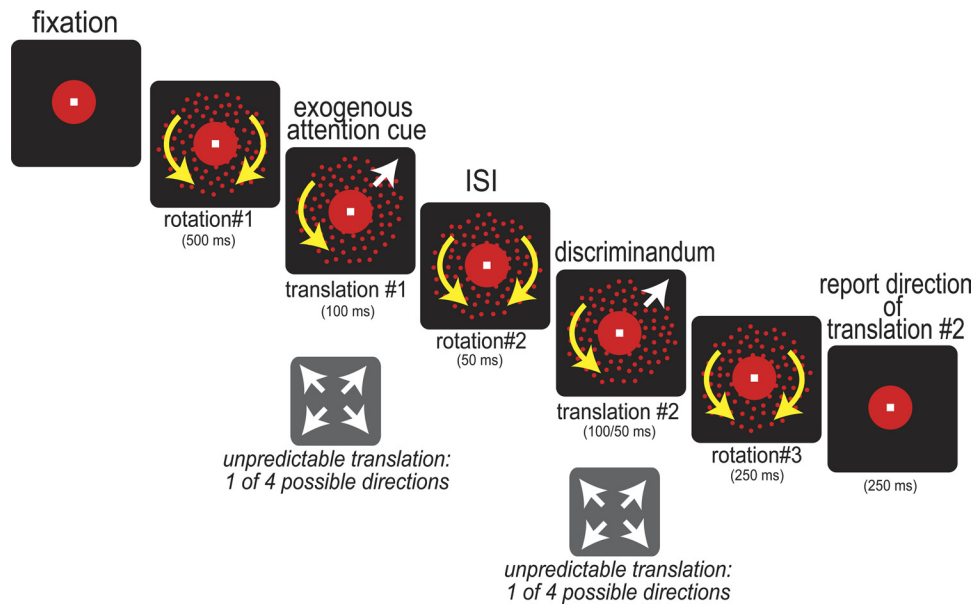
Fig. 1. Stimuli and task. Sequence of events in a trial: trials began with a central white fixation point (0.3 × 0.3° visual angle) presented simultaneously with a surrounding red circle (radius, 1° visual angle; duration, 800/850 or 1,200/1,250 ms) against a black background. Two superimposed circular, high-contrast red dot patterns then appeared, one rotating clockwise and the other counterclockwise (radius, 3.5° visual angle; *rotation 1*, 500 ms). Sixty percent of the dots in 1 of the dot fields then translated in 1 of 4 possible directions, selected at random (*translation 1*, 100 ms) while the remaining dots in this dot field, and dots in the other dot field, continued rotating in their original directions. This first translation served to exogenously direct attention to the translating dot field. The dot pattern undergoing this first translation is therefore referred to as the "cued surface." After this first translation, both surfaces rotated as before (*rotation 2*, 50 ms), and then a second brief translation occurred; 60% of dots in 1 of the dot fields translated in 1 of 4 possible directions, selected at random (*translation 2*, 100 or 50 ms). After this second translation, both dot fields again rotated as before (*rotation 3*, 250 ms). Subjects could respond during *translation 2*, up to 500 ms after the onset of *rotation 3*. They reported the direction of the second translation by pressing 1 of 4 keys.

surface, corresponding to a positive attention index for percent correct. All subjects showed a significantly positive attention index for percent correct (2-tailed Wilcoxon signed rank test: S1, $P = 0.0006$; S2, $P = 0.013$; S3, $P < 0.0001$; S4, $P = 0.0012$; S5, $P < 0.0001$; S6, $P = 0.001$; S7, $P = 0.016$). Figure 2D shows the group average for the percent correct attention index across subjects, with mean $\pm$ SE across all subjects, all repeats. Across observers, the percent correct attention index was significantly different from zero (2-tailed Wilcoxon signed rank test, $P < 0.0001$).

In our paradigm, subjects were not instructed to respond as quickly as possible but did have to respond within a limited window of time. We found that subjects tended to be faster at discriminating the direction of the second translation when it occurred on the cued surface. Figure 3A plots mean reaction time across scans for the subject (S3) whose percent correct data are shown in Fig. 2A. This subject was faster at discriminating the direction of the second translation for the cued vs. the uncued surface. Figure 3B plots mean reaction time performance for the cued and uncued surface for each of the individual subjects. Subjects S3, S4, and S5 showed a statistically significant reduction in reaction time performance for the cued relative to the uncued condition (2-tailed paired Wilcoxon signed rank test: S3, $P = 0.001$; S4, $P = 0.003$; S5, $P < 0.004$). Thus subjects were not more accurate at discriminating translations of the cued surface because they were taking longer to respond; percent correct performance was not due to a speed-accuracy tradeoff. Rather, subjects tended to respond both more quickly and more accurately on cued vs. uncued trials.

We also computed an attention index for reaction time for each subject and each scan $[(C - U)/(C + U)]$. Figure 3C plots

the mean reaction time attention index for each subject. A negative attention index for reaction time indicates faster reaction times for the cued relative to the uncued surface. Three of the seven subjects showed individually significant negative attention index for reaction time (2-tailed Wilcoxon signed rank test: S3, $P = 0.0008$; S4, $P = 0.003$; S5, $P < 0.008$). The average reaction time attention index across subjects is plotted in Fig. 3D, with mean $\pm$ SE across all subjects, all repeats. The reaction time attention index was significantly different from zero across observers (2-tailed Wilcoxon signed rank test, $P < 0.0001$).

Thus, consistent with earlier studies, we found that accuracy was better for cued than uncued trials. We also found that response latencies were generally shorter for cued than uncued trials. Although reaction time was not our main dependent measure, since subjects were not instructed to respond as quickly as possible, our task was very demanding, and subjects had a limited time to make their decision (750 ms). Critically, we found that the effect of cueing on accuracy did not reflect a speed-accuracy trade-off.

fMRI results: surface-based attention effects on fMRI responses. We examined fMRI responses in visual areas V1, V2, V3, V3A, V4V, and MT+ to cued and uncued translations. We constrained responses based on a localizer scan (see MATERIALS AND METHODS) such that only the subregion of each visual area that was responsive to the parameters of our visual stimulus was analyzed.

We found that BOLD responses tended to be larger when the cued rather than the uncued surface translated. This is shown in Fig. 4. Figure 4A plots the hemodynamic response function (HDR) across visual areas when the second translation was of the cued vs. the uncued surface for subject S3, based on a
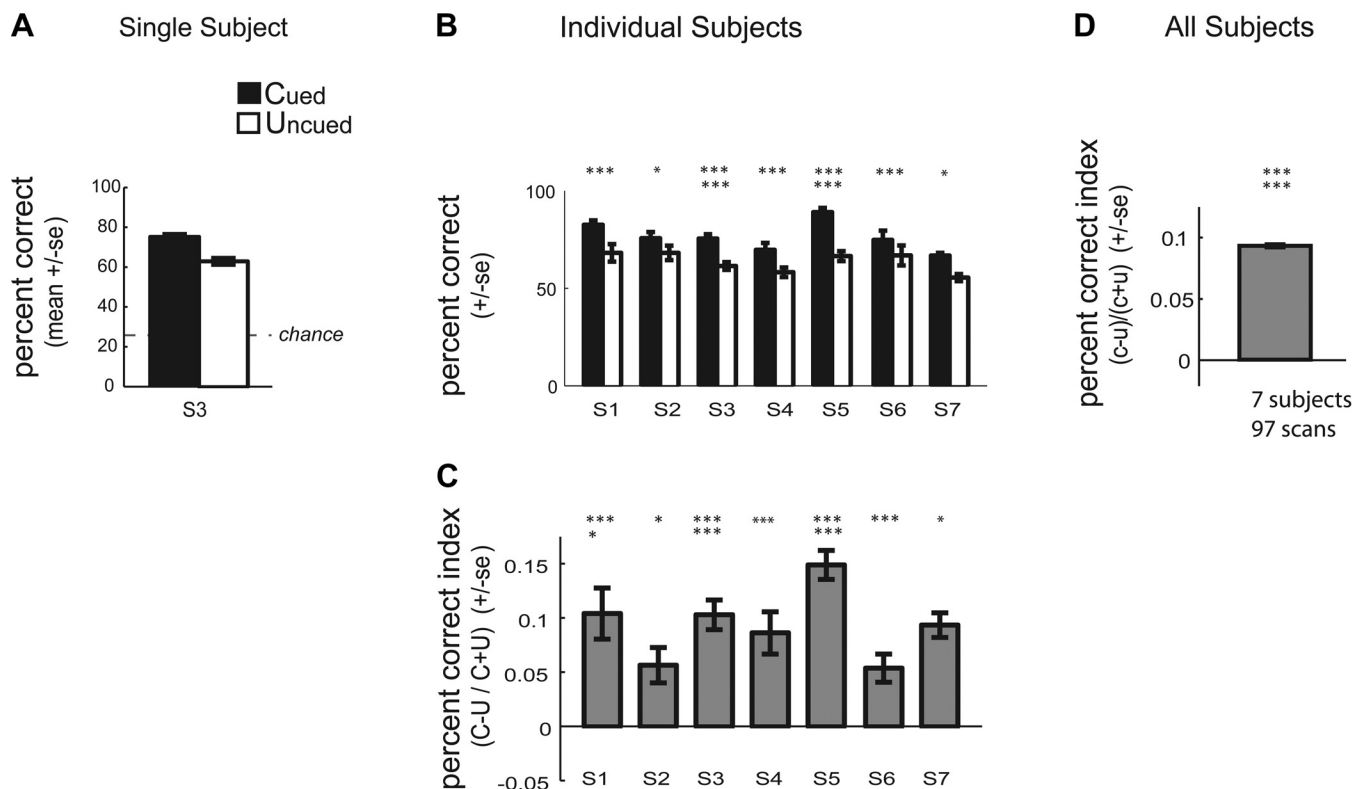
Fig. 2. Psychophysical performance during scanning: accuracy. *A*: plot of the mean percent correct (±SE) for a single subject (S3) for both attention conditions: cued (C; first and second translation of the same dot field) and uncued (U; first and second translation of different dot fields). *B*: plot of the mean percent correct for the cued and uncued condition for each of the 7 subjects (S1–S7). To normalize for differences in the absolute magnitude of the attention effect across subjects, we also computed an attention index, defined as $(C + U)/(C - U)$. *C*: plot of the average percent correct attention index for each subject (±SE across scans). *D*: plots the average percent correct attention index across subjects (±SE across all subjects, all repeats). *$P < 0.05$; **$P < 0.01$; ***$P < 0.005$; ****$P < 0.001$; *****$P < 0.0005$; ******$P < 0.0001$ (2-tailed paired Wilcoxon signed rank test for *B* and 2-tailed paired Wilcoxon signed rank test for *C* and *D*).

single scan. Figure 4*B* shows the computed peak amplitude of the HDRs on cued and uncued trials for the same subject, S3, over a single scan. Figure 4*C* shows the mean across all scans for this same subject. Across scans, this subject showed significant cueing effects in several of the early visual areas we studied (2-tailed paired Wilcoxon signed rank test: V1, $P = 0.034$; V2, $P = 0.021$; V3, $P = 0.044$; V3A, $P < 0.0001$; V4V, $P = 0.025$; MT+, $P = 0.009$).

To quantify cueing effects across subjects, we first computed a BOLD attention index. For each scan, the estimated peak amplitude of the uncued HDR (U) was subtracted from the peak of the cued HDR (C) and divided by the sum, yielding the index $(C - U)/(C + U)$. A positive BOLD index indicates larger BOLD responses for the cued relative to the uncued translation. Figure 4*D* plots the BOLD attention index for the single subject shown in Fig. 4, *A–C*, subject S3. Subject S3 showed significant cueing effects in several visual areas with a positive attention index, indicating stronger BOLD responses for the cued vs. the uncued surface (2-tailed Wilcoxon signed rank test: V3A, $P = 0.001$; V4V, $P = 0.029$; MT+, $P = 0.021$). Other visual areas showed a trend in the same direction but did not reach significance (V1, $P = 0.058$; V2, $P = 0.051$; V3, $P = 0.051$).

Figure 4*E* shows the average BOLD index for each visual area (±SE) across all subjects, all scans. The BOLD index tended to be positive, indicating stronger BOLD responses for attended vs. unattended translations. Cueing effects were significantly greater than zero in visual areas V1, V2, V3, V3A,

and MT+ (2-tailed Wilcoxon signed rank test: V1, $P = 0.025$; V2, $P = 0.003$; V3, $P = 0.013$; V3A, $P = 0.006$; MT+, $P = 0.007$). Although area V3 tended to show a larger BOLD attention index, there were no significant differences in the magnitude of attentional modulation across pairs of visual areas, as revealed by a Tukey-Kramer test ($P > 0.05$). Thus, across subjects, we found that BOLD responses to translations of the cued surface were significantly elevated throughout the visual system, including primary visual cortex.

DISCUSSION

Many studies of attention have addressed the mechanisms involved in the prioritized processing of attended spatial locations or features. Yet, attention can also be allocated to objects, and much less is known about the mechanisms mediating such object-based selection.

Several object-based attention studies have used paradigms that manipulate within-object vs. between-object selection by changing the spatial locus of attention (Pei et al. 2002; Müller and Kleinschmidt 2003; Shomstein and Behrmann 2006). These studies reveal important interactions between spatial and object-based selection but cannot isolate the unique contributions of object-based selection in the absence of spatial selection because different objects occupy different positions in space. Studies using superimposed faces and houses to study object-based attention (O'Craven et al. 1999; Serences et al. 2004) may also allow for space-based selection, since the
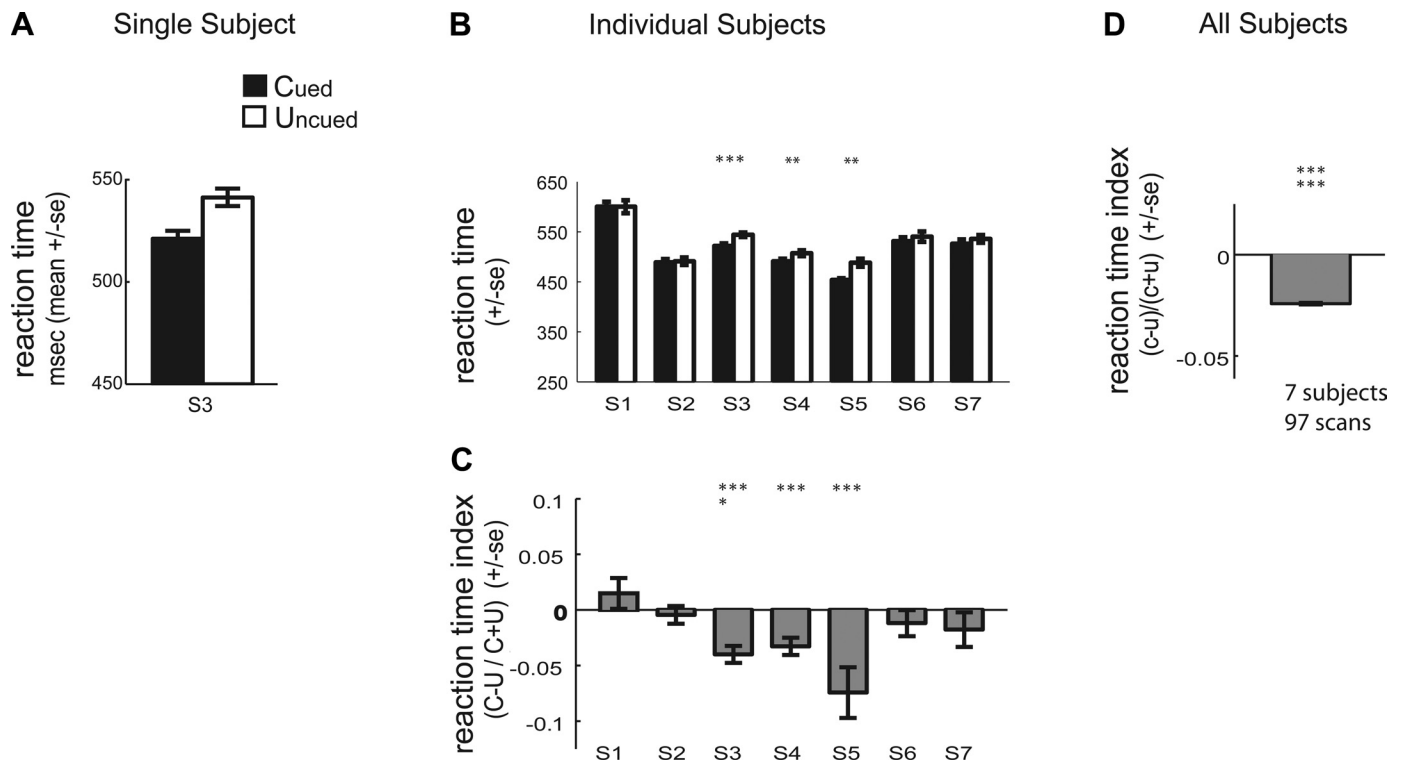
Fig. 3. Psychophysical performance during scanning: reaction time. *A*: plot of the mean reaction time (±SE) for subject S3 for the cued attention condition (first and second translation of the same dot field) and the uncued attention condition (first and second translation of different dot fields). *B*: plot of the mean reaction time for the cued and uncued condition for each of the 7 subjects (S1–S7). We computed an attention index, $(C + U)/(C - U)$, to normalize for differences in the absolute magnitude of the attention effect across subjects. *C*: plot of the average reaction time attention index for each subject (±SE across scans). *D*: plot of the average reaction time attention index across subjects (±SE across all subjects, all repeats). $*P < 0.05$; $**P < 0.01$; $***P < 0.005$ (2-tailed paired Wilcoxon signed rank test for *B* and 2-tailed Wilcoxon signed rank test for *C* and *D*).

features that distinguish spatially overlapping objects are not spatially coincident and allow for selection based on the location where covert attention may be allocated to a unique feature. Furthermore, studies presenting overlapping faces and houses are presenting objects whose low-level features differ in spatial frequency and could potentially result in feature-based selection of a particular frequency (highlighted in Watt 1998).

In the present study we adapted an object-based attention paradigm developed by Valdes-Sosa and colleagues (1998a,b) that was designed to preclude spatial and feature-based selection. We studied the behavioral correlates of exogenous object-based attention in combination with fMRI to localize the underlying neuronal mechanisms of object-based attention to select early human visual areas.

Unit of selection: spatial location, features, or surfaces. The improved discrimination we found for translations of the cued surface was not due to space-based selection. Subjects could not solve our task by attending a particular spatial location, because the two virtual surfaces were spatially superimposed. Moreover, the locations of the individual dots that defined the two surfaces were constantly changing. Furthermore, the stimulus change was global and distributed randomly over the entire surface, with only 60% of the dots translating coherently. Thus subjects could not solve our task by attending a particular spatial subregion.

Improved discrimination also was not due to feature-based selection. Subjects could not solve this task by attending a particular color, because both surfaces were the same color (as in Mitchell et al. 2003). The cueing effect also could not be

attributed to attending to a particular direction of motion, because any given direction was equally likely to occur on either surface and the direction of translation was unpredictable. Furthermore, there was no systematic relationship between the direction of translation and rotation; either surface was equally likely to undergo translation, and the direction of translation was uncorrelated with the direction of rotation. Although rotational motion initially segregated the two surfaces, and the first translation distinguished the cued from the uncued surface, directing feature-based attention to a particular feature, the direction of rotational motion or of the first translation could not account for differences in perceiving a subsequent translation direction, since this would not favor any particular direction for the second translation.

The observed cueing effects imply that attention was directed to the motions of the cued surface even as those motions changed unpredictably. This object- or surface-based selection appears to require that those successive motions, rotation and translation, were linked or bound together in an object-specific manner. The mechanisms underlying this type of selective processing have been the topic of current debate and pose an important topic for future research (see Treue and Katzner 2007; Freiwald 2007).

Our finding of surface-based attention effects in human area MT+ (which includes visual areas MT and MST) corroborates single-unit physiological recording studies in area MT in non-human primates using a similar paradigm (Wannig et al. 2007). This result extends earlier studies in monkey showing that MT plays a role in the organization of motion fields into surfaces,

A Single Subject
*Hemodynamic Responses
Single Scan*

■ Cued Trials
┈ Uncued Trials

B Single Subject
*Peak Amplitudes
Single Scan*

C Single Subject
*Peak Amplitudes
All Scans*

D Single Subject
Attention Index

■ (Cued-Uncued)/(Cued+Uncued)
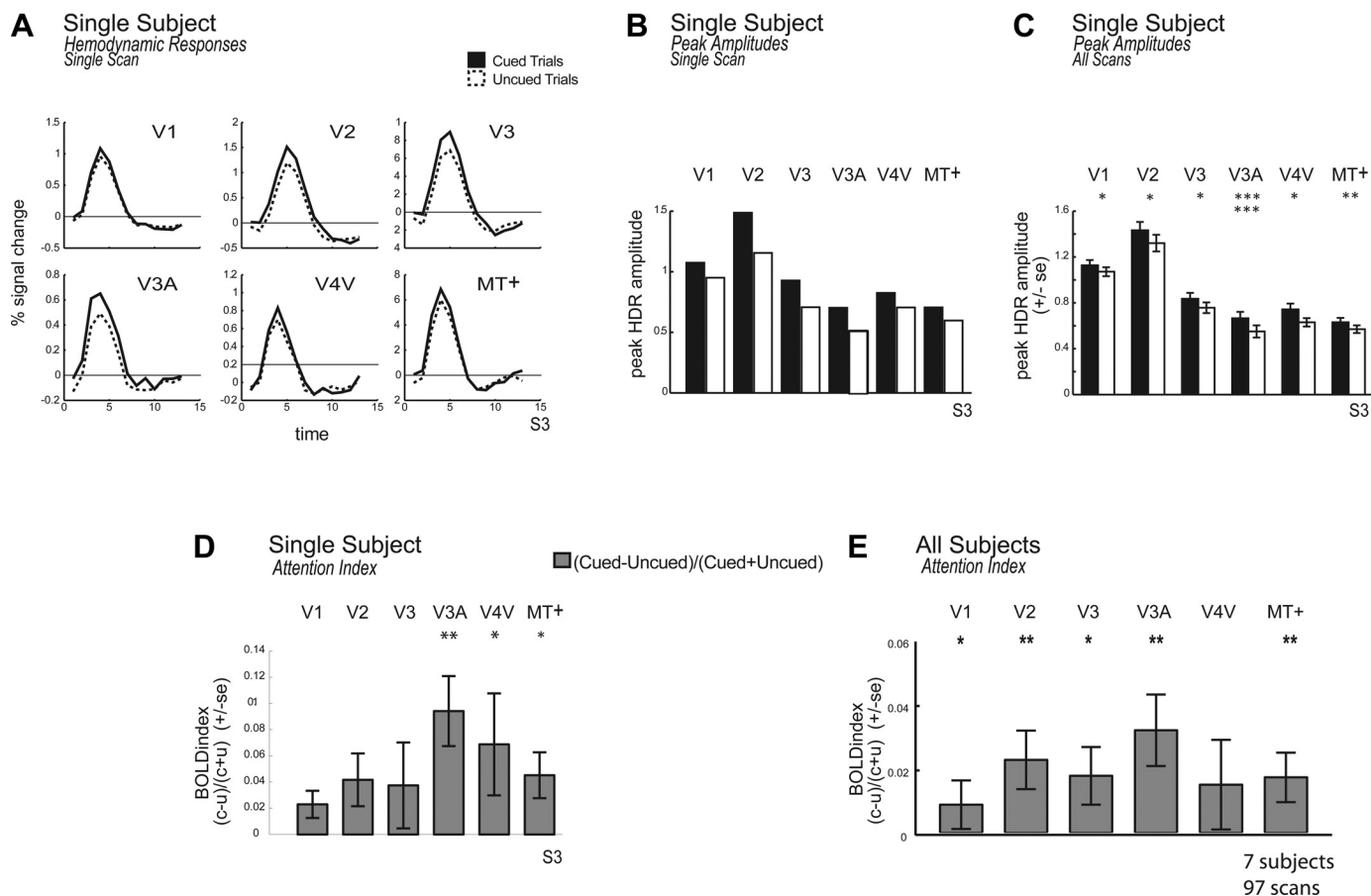
E All Subjects
Attention Index

7 subjects
97 scans

Fig. 4. Functional magnetic resonance imaging (fMRI) responses in early visual areas: cued vs. uncued surfaces. *A*: plot of the hemodynamic response functions (HDRs) for observer S3 for a single scan in areas V1, V2, V3, V4V, V3A, and MT+. Data were averaged across the left and right hemisphere. HDRs when the first and second translation were of the same dot field (cued condition) are shown as a solid line. HDRs when the first and second translation were of different dot fields (uncued condition) are shown as a dotted line. *B*: plot of the peak amplitude of the best-fitting difference of gamma functions for the HDR for the same single scan for subject S3 for the cued and uncued conditions. *C*: plot of the peak amplitude of the blood oxygen level-dependent (BOLD) for subject S3 across all scans. *D*: plot of the BOLD attention index for S3 across all scans, defined as $(C - U)/(C + U)$. *E*: plot of the average BOLD attention index across subjects ($\pm$SE across all subjects, all scans). $*P < 0.05$; $**P < 0.01$; $***P < 0.005$ (2-tailed paired Wilcoxon signed rank test for *B* and *C* and 2-tailed Wilcoxon signed rank test for *D* and *E*).

such as in structure from motion displays (Bradley et al. 1998), and that MST is responsive to rotational motion, the feature that originally defines the two surfaces in our experiment. These previous studies highlight the role of MT/MST in delineating object boundaries. Our results show that once objects have been delineated, these same visual areas are also modulated by attending to objects. Furthermore, our results suggest that the modulation of the N200 ERP component observed by Rodriguez and Valdes-Sosa (2006) in their human study may have arisen from area MT. Unlike Fallah et al. (2007), who found object-based color modulation of single units in area V4, we did not use differently colored dot fields. It is therefore not surprising that we failed to find surface-based attention effects in human area V4. Overall, this result is thus generally consistent with previous single-unit and ERP studies.

What is more surprising is our finding that earlier visual areas also respond more strongly to the attended surface. Previous studies have highlighted the role of early visual areas in the organization and segregation of visual elements into coherent units, or surfaces. V2 has been implicated in surface-based segmentation (Qui and von der Heydt 2005). Also, V3A has been found to be differentially responsive to different types of global motion (Koyama et al. 2005). The present study is,

however, the first to find unambiguous evidence of object-based attentional selection in human area V1 in a task that precludes spatial and feature-based selection. This study corroborates previous work finding evidence for object-based attention in macaque area V1 in a very different paradigm (Roelfsema et al. 1998). However, in the study by Roelfsema et al., the attended objects were spatially distinct, except for one point of overlap, and thus spatial and object-based selection were not dissociable. Our results in human V1 also bear on the finding of Khoe et al. (2005) of modulation of the C1 ERP component by surface-based attention. Although their results were equivocal with regard to the origin of that component, our findings support a striate origin. We also found significant cueing effects in areas V2, V3, and V3A.

The discovery of object-based modulation within area V1 may seem somewhat surprising given that the objects in question, surfaces defined by random dots, extended far beyond the receptive fields of individual neurons in area V1. Moreover, an individual V1 neuron would have been stimulated by one object and then the other in fairly rapid succession as the moving elements of each object, the dots, passed through a given region of visual space. A recent psychophysical and modeling study by Stoner and Blanc (2010) potentially sheds

light on this result. Stoner and Blanc devised a variation of Reynolds et al.'s paradigm (2003) in which they unexpectedly switched the properties, motion directions, and/or colors of the two superimposed dot fields on some trials. This manipulation allowed them to determine whether object-based cueing was specific to those properties. They found that the performance advantage did not depend on the color or direction of the two surfaces at the moment of translation but was instead specific to the dots of the surface that had been exogenously cued (in their experiment by an abrupt onset). Stoner and Blanc (2010) concluded that neurons with receptive fields small enough to distinguish the dots of the two superimposed fields must be involved. They offered a mechanistic account that incorporated feature-specific feedback from extrastriate cortex (i.e., areas V4 and MT) onto V1 and non-feature-specific local connections within V1. By enhancing the processing of all V1 neurons with receptive fields that currently contain dots of the cued surface, this hypothetical mechanism dynamically links the attributes of the cued surface and achieves object-based selection within area V1. Testing the validity of this model is an exciting goal for future neurophysiological experiments.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

Blaser E, Pylysh ZW, Holcombe AO. Tracking an object through feature space. *Nature* 408: 196–199, 2000.

Bradley DC, Chang GC, Andersen RA. Encoding of three-dimensional structure-from-motion by primate MT neurons. *Nature* 392: 714–717, 1998.

Brainard DH. The Psychophysics Toolbox. *Spat Vis* 10: 433–436, 1997.

Brefczynski JA, DeYoe EA. A physiological correlate of the "spotlight" of visual attention. *Nat Neurosci* 2: 370–374, 1999.

Buracas GT, Boynton GM. Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage* 16: 801–813, 2002.

Ciaramitaro VM, Buracas GT, Boynton GM. Spatial, and cross-modal attention alter responses to unattended sensory information in early visual and auditory human cortex. *J Neurophysiol* 98: 2399–2413, 2007.

Dale A. Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8: 109–114, 1999.

Damasio AR, Damasio H, van Hoesen GW. Prosopagnosia: anatomic basis, and behavioral mechanisms. *Neurology* 32: 331–341, 1982.

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.

Fallah M, Stoner GR, Reynolds JH. Stimulus-specific competitive selection in macaque extrastriate visual area V4. *Proc Natl Acad Sci USA* 104: 4165–4169, 2007.

Freiwald WA. Attention to objects made of features. *Trends Cogn Sci* 11: 453–454, 2007.

Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of Macaque. *J Neurophysiol* 35: 96–111, 1972.

He ZJ, Nakayama K. Visual attention to surfaces in three-dimensional space. *Proc Natl Acad Sci USA* 92: 11155–11159, 1995.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302–4311, 1997.

Khoe W, Mitchell JF, Reynolds JH, Hillyard SA. Exogenous attentional selection of transparent superimposed surfaces modulates early event-related potentials. *Vision Res* 45: 3004–3014, 2005.

Koyama S, Sasaki Y, Andersen GJ, Tootell RB, Matsuura M, Watanabe T. Separate processing of different global-motion structures in visual cortex is revealed by FMRI. *Curr Biol* 15: 2027–2032, 2005.

Lopez M, Rodriquez V, Valdes-Sosa M. Two-object attentional interference depends on attentional set. *Int J Psychophysiol* 53: 127–134, 2004.

Maunsell JHR, Treue S. Feature-based attention in visual cortex. *Trends Neurosci* 29: 317–322, 2006.

Mitchell JF, Stoner GR, Fallah M, Reynolds JR. Attentional selection of superimposed surfaces cannot be explained by modulation of the gain of color channels. *Vision Res* 43: 1323–1328, 2003.

Müller NG, Kleinschmidt A. Dynamic interaction of object-, and space-based attention in retinotopic visual areas. *J Neurosci* 23: 9812–9816, 2003.

O'Craven KM, Downing PE, Kanwisher N. fMRI evidence for objects as the units of attentional selection. *Nature* 401: 584–587, 1999.

Pei F, Pettet MW, Norcia AM. Neural correlates of object-based attention. *J Vis* 2: 588–596, 2002.

Pelli DG. The Video Toolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10: 437–442, 1997.

Pinilla T, Cobo A, Torres K, Valdes-Sosa M. Attentional shifts between surfaces: effects on detection, and early brain potentials. *Vision Res* 41: 1619–1630, 2001.

Qui FT, von der Heydt R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron* 47: 155–166, 2005.

Reynolds JH, Alborzian S, Stoner GR. Exogenously cued attention triggers competitive selection of surfaces. *Vision Res* 43: 59–66, 2003.

Reynolds JH, Chelazzi L. Attentional modulation of visual processing. *Annu Rev Neurosci* 27: 611–647, 2004.

Rodriguez V, Valdes-Sosa M. Sensory suppression during shifts of attention between surfaces in transparent motion. *Brain Res* 1072: 110–118, 2006.

Roelfsema PR, Lamme VA, Spekreijse H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395: 376–381, 1998.

Saenz M, Buracas GT, Boynton GM. Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5: 631–632, 2002.

Serences JT, Schwarzbach J, Courtney SM, Golay X, Yantis S. Control of object-based attention in human cortex. *Cereb Cortex* 14: 1346–1357, 2004.

Shomstein S, Behrmann M. Cortical systems mediating visual attention to both objects, and spatial locations. *Proc Natl Acad Sci USA* 103: 11387–11392, 2006.

Stoner GR, Blanc G. Exploring the mechanisms underlying surface-based stimulus selection. *Vision Res* 50: 229–241, 2010.

Treue S, Katzner S. Visual attention: of features and transparent surfaces. *Trends Cogn Sci* 11: 451–452, 2007.

Valdes-Sosa M, Cobo A, Pinilla T. Transparent motion, and object-based attention. *Cognition* 66: B13–B23, 1998a.

Valdes-Sosa M, Bobes MA, Rodriguez V, Pinilla T. Switching attention without shifting the spotlight object-based attentional modulation of brain potentials. *J Cogn Neurosci* 10: 137–151, 1998b.

Wannig A, Rodriguez V, Freiwald W. Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54: 639–651, 2007.

Watt RJ. *Visual Processing: Computation, Psychophysical, and Cognitive Research.* Hillsdale, NJ: Erlbaum, 1998.