Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Neural representation of object-specific attentional priority

Taosheng Liu*

Department of Psychology, Michigan State University, East Lansing, MI 48824, USA Neuroscience Program, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history: Received 5 November 2015 Accepted 16 January 2016 Available online 26 January 2016

ABSTRACT

Humans can flexibly select locations, features, or objects in a visual scene for prioritized processing. Although it is relatively straightforward to manipulate location- and feature-based attention, it is difficult to isolate object-based selection. Because objects are always composed of features, studies of object-based selection can often be interpreted as the selection of a combination of locations and features. Here we examined the neural representation of attentional priority in a paradigm that isolated object-based selection. Participants viewed two superimposed gratings that continuously changed their color, orientation, and spatial frequency, such that the gratings traversed the same exact feature values within a trial. Participants were cued at the beginning of each trial to attend to one or the other grating to detect a brief luminance increment, while their brain activity was measured with fMRI. Using multi-voxel pattern analysis, we were able to decode the attended grating in a set of frontoparietal areas, including anterior intraparietal sulcus (IPS), frontal eye field (FEF), and inferior frontal junction (IFJ). Thus, a perceptually varying object can be represented by patterned neural activity in these frontoparietal areas. We suggest that these areas can encode attentional priority for abstract, high-level objects independent of their locations and features.

© 2016 Elsevier Inc. All rights reserved.

Selection of task-relevant information is necessary to guide efficient and adaptive behavior in a complex environment. Attention is the mechanism that can select different aspects of a scene, such as locations, features and objects (Carrasco, 2011; Scolari et al., 2014). Although the neural basis of attention has been extensively studied (Kastner and Ungerleider, 2000; Reynolds and Chelazzi, 2004), a central question remains: how is top-down selection implemented in the brain?

A key assumption of attention theories is that higher-order brain areas maintain attentional priority, akin to a template, that exerts topdown control to guide selection (e.g., Deco and Rolls, 2004; Desimone and Duncan, 1995; Wolfe, 1994). For the control of spatial attention, the neural representation of spatial priority has been strongly linked to spatiotopic neural responses in dorsal frontoparietal areas (Bisley and Goldberg, 2010). Neurophysiological evidence from microstimulation studies suggest that these higher-level topographic representations send top-down control signals to earlier visual areas to implement spatial selection (Ekstrom et al., 2008; Moore and Armstrong, 2003; Moore and Fallah, 2004). For the control of featurebased attention, evidence from human fMRI and monkey neurophysiology has suggested that the dorsal frontoparietal areas can also represent the attended visual feature such as specific color and motion direction (Liu et al., 2011; Liu and Hou, 2013; Mendoza-Halliday et al., 2014). However, real scenes typically contain many objects, and observers

E-mail address: tsliu@msu.edu.

often select whole perceptual objects (Scholl, 2001). This raises the question of how attentional priority for perceptual objects is represented in the brain.

One key challenge in studying object-based attention is that objects are always composed of features so it can be difficult to ascertain that selection occurred on the level of whole objects instead of elemental features. For example, in a popular paradigm where participants were instructed to attend to either a face or a house in a superimposed face/ house image, the face and house stimuli differ in terms of low level features such as curvature and spatial frequency (Watt, 1998). Thus behavior in these studies can be potentially facilitated by feature-level selection, making it difficult to attribute results to object-based attention.

The goal of the present study is to investigate the neural representation of attentional priority for perceptual objects. Based on previous work showing that the dorsal frontal and parietal areas represent attentional priority for non-spatial features, we hypothesized that these areas can also represent priority for whole perceptual objects. To isolate object-level selection, we employed a compound stimulus composed of two objects that continuously evolved in multiple feature dimensions (Blaser et al., 2000). We then applied both fMRI univariate analysis and multivariate pattern analysis to investigate neural signals that can represent specific attended objects. Because we employed a cueing approach to direct attention, static featural differences associated with the cue could potentially account for classification results. Thus we also ran a control experiment to rule out the contribution of featurebased attention to our results.







^{*} Department of Psychology, 316 Physics Rd., Michigan State University, East Lansing, MI 48824, USA.

Materials and methods

Participants

Twelve individuals (six females, mean age: 25.5), including the author, participated in the experiment. Participants were recruited from the Michigan State University community (graduate and undergraduate students and the author) and all had normal or corrected-to-normal visual acuity and reported to have normal color vision. Participants were paid at the rate of \$20/h for their time. They gave informed consent under the study protocol approved by the Institutional Review Board at Michigan State University. Sample size was determined prior to data collection and was based on comparable studies in the literature on fMRI studies of visual attention. We also conducted a power analysis, using effect size estimated from a previously published study in our lab that used a similar paradigm to decode attentional state (Hou and Liu, 2012). We pooled decoding accuracies from two frontoparietal sites (IPS and FEF) across participants to estimate the effect size. We then used G*Power 3.1.9 (Faul et al., 2007) to estimate the power in detecting a true effect using a two-tailed t-test for significant above-chance classification. This analysis showed that a sample of 12 participants would give a power of 0.82.

Stimulus and display

The visual stimuli consisted of two superimposed Gabor patches $(\sigma = 1.1^{\circ})$ that varied in their orientation, color, and spatial frequency simultaneously (Fig. 1). The evolution of the features followed fixed, cyclic trajectories in their respective dimensions. On each trial, the Gabors rotated counterclockwise through all possible orientations at a speed of 59°/s; the colors of the Gabors traversed through all hues on a color circle in the CIE L*a*b space (L = 30, center: a = b = 0, radius = 80) at a speed of 59°/s; the spatial frequency of the Gabors varied smoothly in a sinusoidal fashion from 0.5 cycles/deg. to 3 cycles/deg. at a speed of 0.41 cycles/deg./s. Thus, in 6.1 s (the duration of the stimulus movie), the Gabors rotated two full cycles in orientation, traversed one cycle in the color space, and one full period in the sinusoidal modulation of spatial frequency. All features evolved continuously and simultaneously with maximal offset between the two Gabors (opposite angles in color space, orthogonal orientations, opposite phases in the modulation of spatial frequency).

All stimuli were generated using MGL (http://justingardner.net/mgl), a set of custom OpenGL libraries running in Matlab (Mathworks, Natick, MA). Images were projected on a rear-projection screen located in the scanner bore by a Toshiba TDP-TW100U projector outfitted with a custom zoom-lens (Navitar, Rochester, NY). The screen resolution was set to 1024×768 and the display was updated at 60 Hz. Participants viewed the screen via an angled mirror attached to the head coil at a viewing distance of 60 cm. Color calibration was performed with a MonacoOPTIX colorimeter (X-rite, Grand Rapids, MI), which generated an ICC profile for a display. We then used routines in Matlab's Image Processing Toolbox to read the ICC profile and calculate a transformation from the CIE L*a*b space to the screen RGB values.

Task and design

Participants tracked one of the Gabor patches on each trial and performed a change detection task. At the beginning of each trial, a number ("1" or "2", 0.4° , white) appeared in the center of the display for 0.5 s. In prior practice sessions (see below), participants had learned to associate "1" with the Gabor that was initially red, horizontal, and high spatial frequency, and to associate "2" with the Gabor that was initially cyan, vertical, and low spatial frequency. The initial image of the two Gabors appeared together with the number cue. We referred to these two Gabors as "Object 1" and "Object 2" in the instruction, and we adopt the same terminology for the rest of this report. During the subsequent 6.1 s, the two objects continuously evolved through the features space as described above, and participants were instructed to track the cued object and monitor for a brief brightening event (0.2 s). On each trial, there was either a brightening of the cued object (target), a brightening of the uncued object (distracter), or no brightening of either object (null). The three trial types (target, distracter, null) were interleaved and equally probable (proportion 1/3 each). The timing of targets and distracters conformed to a uniform distribution in two possible time windows: 1.5-2.5 s or 4.5-5.5 s after trial onset. These time windows were chosen such that the two objects had similar spatial frequency, which made the task challenging. The magnitude of the brightening (luminance increment) was determined for each participant at the beginning of the scanning session with a thresholding procedure (see Practice Sessions below). Participants were instructed to press a button with their right index finger if they detected the target (a brief brightening on the cued object), and withheld response otherwise. They were



Fig. 1. Schematic of a trial in the main experiment. Two superimposed Gabors continuously changed color, orientation, and spatial frequency. Sample images are shown on the left; the trajectories in feature space are shown on the right (solid and dashed curves represent two objects). Note color and orientation can be conceived as circular dimensions, whereas spatial frequency is a linear dimension. Here a "2" instructs participants to attend to the second object (the initially low frequency, cyan, vertical Gabor).

told to respond after the stimulus disappeared, without any emphasis on response speed. The instruction emphasized that they should track the cued object throughout the duration of the movie on all trials. Target probability and timing were not conveyed to participants. We counted button responses on target trials as hits, and button responses on either distracter or null trials as false alarms.

There were two conditions in the experiment: attend to Object 1 and attend to Object 2. Each run in the scanner started with an 8.8 s fixation period, followed by a 316.8 s period in which trials of 6.6 s long were separated by ITIs jittered in the range of 4.4–8.8 s. Each participant completed 9 runs in the scanner, which yielded ~100 trials per condition.

Practice sessions

Each participant underwent 2–4 h of practice in the psychophysics laboratory before the MRI session. The first part of the practice involved a thresholding task which was identical to the main task as described above, except that the magnitude of the luminance increment was varied across trials using the method of constant stimuli. Six levels of luminance increments were used and the resulting psychometric function was fit with a Weibull function to derive a threshold corresponding to ~75% correct. Participants continued to practice the thresholding task until their thresholds stabilized.

In the second part of the practice, participants performed the change detection task with individually determined luminance increment threshold. At least four runs were completed, during which we also monitored participants' eye position with an Eyelink 1000 eye tracker (SR Research, Ontario, Canada). Eye position data were analyzed offline to assess the stability of fixation. For each participant, we submitted their trial-level horizontal and vertical eye position to a two-way ANOVA (factor 1: attend to Object 1 vs. Object 2; factor 2: time points). We did not find any significant main effect or interaction for any subject in either the vertical or horizontal eye position. Although we were not able to monitor eye position in the scanner, it seemed unlikely participants would change their fixation behavior from training to scanning sessions. We would also like to note that had subjects tracked the cued object, their eye movement pattern would be similar across the two objects because they rotated in the same direction. Thus it is unlikely that eye movement played a significant role in our results.

Retinotopic mapping

For each participant, we also defined early visual cortex containing topographic maps in a separate scanning session. We used rotating wedge and expanding/contracting rings to map the polar angle and radial component, respectively (DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995). Borders between visual areas were defined as phase reversals in a polar angle map of the visual field. Phase maps were visualized on computationally flattened representations of the cortical surface, which were generated from the high resolution anatomical image using FreeSurfer and custom Matlab code. In a separate run, we also presented moving vs. stationary dots in alternating blocks and localized the motion-sensitive area, MT +, as an area near the junction of the occipital and temporal cortex that responded more to moving than stationary dots (Watson et al., 1993). We were able to reliably indentify the following areas in all participants: V1, V2, V3, V3AB, V4, and MT+. There is controversy regarding the definition of visual area V4 (for a review, see Wandell et al., 2007). Our definition of V4 followed that of Brewer et al. (2005), which defines V4 as a hemifield representation directly anterior to V3v.

MRI data acquisition

All functional and structural brain images were acquired using a GE Healthcare (Waukesha, WI) 3 T Signa HDx MRI scanner with an 8channel head coil, in the Department of Radiology at Michigan State University. For each participant, high-resolution anatomical images were acquired using a T1-weighted MP-RAGE sequence (FOV = 256 mm × 256 mm, 180 sagittal slices, 1 mm isotropic voxels) for surface reconstruction and alignment purposes. Functional images were acquired using a T2*-weighted echo planar imaging sequence consisted of 30 slices (TR = 2.2 s, TE = 30 ms, flip angle = 80°, matrix size = 64×64 , in-plane resolution = 3 mm × 3 mm, slice thickness = 4 mm, interleaved, no gap). In each scanning session, a 2D T1-weighted anatomical image was also acquired that had the same slice prescription as the functional scans, for the purpose of aligning functional data to high resolution structural data.

fMRI data analysis

Data were processed and analyzed using mrTools (http://www.cns. nyu.edu/heegerlab/wiki/doku.php?id=mrtools:top) and custom code in Matlab. Preprocessing of function data included head motion correction, linear detrending and temporal high pass filtering at 0.01 Hz. The 2D T1-weighted image was used to compute the alignment between the functional images and the high-resolution T1-weighted image, using an automated robust image registration algorithm (Nestares and Heeger, 2000). Functional data were converted to percent signal change by dividing the time course of each voxel by its mean signal over a run, and data from the nine scanning runs were concatenated. All region-ofinterest (ROI) analyses were performed on individual participant's native anatomical space. For group-level analysis, we used surface-based spherical registration as implemented in Caret to co-register the individual participant's functional data to the Population-Average, Landmark- and Surface-based (PALS) atlas (Van Essen, 2005). Grouplevel statistics (random effects) were computed in the atlas space and the statistical parameter maps were visualized on a standard atlas surface (the "very inflated" surface). To correct for multiple comparisons, we set the threshold of the maps based on individual voxel level pvalue in combination with a cluster constraint, using the 3dFWHMx program to estimate the smoothing parameter and the 3dClustSim program to estimate the cluster threshold; both programs are distributed as part of AFNI.

Univariate analysis: deconvolution

Each voxel's time series was fitted with a general linear model whose regressors contained two attention conditions (Object 1, Object 2). Each regressor modeled the fMRI response in a 25 s window after trial onset. The design matrix was pseudo-inversed and multiplied by the time series to obtain an estimate of the hemodynamic response for each attention condition. To measure the response magnitude of a region, we averaged the deconvolved response across all the voxels in a ROI. For each voxel, we also computed a goodness of fit measure (r^2 value), which is the amount of variance in the fMRI time series explained by the deconvolution model. The r^2 value is analogous to an omnibus F statistic in ANOVA, in that it indicates the degree to which a voxel's time course is modulated by the task events (Gardner et al., 2005).

For the visualization of the univariate result at the group level, we computed a group-averaged r^2 map. The statistical significance of the r^2 values was assessed via a permutation test (for details see Gardner et al., 2005; Liu et al., 2011). At the group level, we transformed the individually obtained r^2 maps to the PALS atlas space and averaged their values (Fig. 3). We used a p of 0.005 and a cluster extent of 18 to threshold the group r^2 map, which corresponded to a whole-brain false positive rate of 0.01 according to AFNI 3dClustSim.

Multivariate analysis: multi-voxel pattern analyses

Our goal in the multivariate analysis was to identify patterns of neural activity that can represent attended object in our task. To this end, we performed multi-voxel pattern analysis (MVPA) across the whole brain using a "searchlight" procedure (Kriegeskorte et al., 2006) to identify voxels that can be used to decode the attended object. We restricted our search to the cortical surface based on each individual participants' surface reconstruction. For each voxel in the gray matter (center voxel), we defined a small neighborhood containing all gray matter voxels within a 12 mm radius (measured as distance on the folded cortical surface). This radius produced neighborhoods containing ~100 voxels on average and MVPA was performed on each of such neighborhoods.

For each neighborhood, we extracted single-trial fMRI response for each voxel with the following procedure. We first averaged the deconvolved fMRI response obtained from univariate analysis (see above) across all voxels and conditions in the neighborhood, which served as an estimate of the hemodynamic impulse response function (HIRF) in that neighborhood. We then constructed four boxcar functions, with a boxcar corresponding to each individual trial. The first two functions coded trials on which participants made a button-press response, one function for Object 1 and another function for Object 2. Similarly, the second two functions coded trials on which participants did not make a response (one for each object). The boxcar functions were then convolved with the estimated HIRF to produce a design matrix coding for each individual trial in each condition. The design matrix was then pseudo-inversed and multiplied by the time series to obtain an estimate of the response amplitude (beta weight) for each individual trial in each voxel. MVPA was performed using beta weights on noresponse trials, which were composed of most of the null and distracter trials and a small portion of target trials. This precluded neural signal related to target detection and motor response from contributing to our results. The no-response trials accounted for the majority of trials due to our design (1/3 target-present trials) and we obtained ~70 trials per condition for each participant. We used linear support vector machines (SVM) as implemented in LIBSVM (Chang and Lin, 2011) in a leave-one-run-out cross-validation procedure to perform the multivoxel pattern analysis. The SVM was trained to discriminate between multivoxel patterns associated with attending to Object 1 vs. Object 2. We trained the SVM using trial data from 8 runs and tested its accuracy on the trial data from the left-out run. This was repeated 9 times until all trials were classified by the SVM, and the resulting classification accuracy was assigned to the center voxel of the neighborhood. The above procedure was repeated for all neighborhoods to obtain a map of classification accuracy across the cortical surface.

For the group analysis, we averaged individual classification maps in the PALS atlas space. Voxel-wise p-value was derived by a t-test (two-tailed) comparing the classification accuracies from all participants (N = 12) against the chance-level (0.5). We used 3dFWHMx to estimate the smoothness of individual classification maps and used the average value to determine the threshold of the group analysis via 3dClustSim. The group-average map was thresholded at a voxel p value of 0.002 and a cluster extent of 18 voxels, which corresponded to a whole-brain false positive rate of 0.01. The threshold was slightly different from the univariate analysis above because the classification map had a larger smoothing parameter than the fMRI time series.

Methods for the control experiment

We also ran a separate control experiment to assess the contribution of static featural differences in the main experiment (see Results for the full rationale). The control experiment was very similar to the main experiment, so only a brief description of the methods is provided here, with differences between experiments emphasized.

Another group of 12 individuals participated in the control experiment; all were graduate and undergraduate students at Michigan State University (mean age: 23.8), with one individual already participated in the main experiment. Stimuli were identical to those of the main experiment, except we introduced two additional Gabor objects for the tracking task (Fig. 5). Object 3 was identical to Object 1 at the beginning of the trial, but its trajectory in the color and orientation dimension was reversed with respect to Object 1; similarly, Object 4 started in an identical form as Object 2 but evolved along a reversed trajectory in the color and orientation dimension. On each trial, two objects appeared together and an auditory cue ("one", "two", "three", "four") instructed participants which object to attend in order to detect a brightening event on that object. There were two possible pairings of objects: either Object 1 and 2 were shown, or Object 3 and 4 were shown. All trial types were interleaved within a run. The timing, target prevalence, and training procedure were all identical to the main experiment. Participants completed 10 runs in the scanner, which yielded ~56 trials per condition.

Imaging data were acquired and analyzed similarly as the main experiment. We performed searchlight MVPA across the whole brain using trials in which participants did not make a button response (on average ~ 38 trials per condition). Four sets of classification analyses were conducted, two within-pairing and two cross-pairing. In the within-pairing classification, a SVM was trained and tested to classify Object 1 vs. Object 2 in a leave-one-run-out cross-validation procedure; similarly, another SVM was trained and tested to classify Object 3 vs. Object 4. In the cross-pairing classification, a SVM was trained to classify Object 1 vs. Object 2 and tested to classify Object 3 vs. Object 4, or vice versa. Whole-brain classification accuracy maps were then averaged across participants, thresholded, and displayed on the atlas surface as the main analysis. Due to fewer trials in the control experiment than the main experiment, the classifier performance is expected to be lower. We thus thresholded the group classification map at a lower threshold, using a voxel p value of 0.01 and a cluster extent of 29 voxels, which corresponded to a whole-brain false positive rate of 0.05.

Results

Behavior: main experiment

Participants reported small luminance increments in the cued object via button presses. They detected around 70% of luminance changes and made less than 5% of false alarms (Fig. 2). We compared performance between attending to Object 1 vs. Object 2 using t-tests. There was no difference in Hit, FA, or Hit-FA scores (all p > 0.6). Thus, behavioral results suggest that our task was attentionally demanding and task difficulty was similar for detecting the changes in the two objects. Although average performance did not differ for the two objects, idio-syncratic variations in performance at the individual level could still drive multivariate pattern classification results (Todd et al., 2013). Thus we also examined individual participant's behavioral data. For



Fig. 2. Behavioral results in the scanner for the main experiment. Average hit and false alarm (FA) rates for each object. The difference score, hit-fa, was also shown. Error bars indicate s.e.m. across participants.

each participant, we obtained two 2 × 2 contingency table, one for the number of hits, one for the number of false alarms, across the two attention conditions. We then performed a χ^2 test for independence and found 1 out of the 24 tests showed a significant effect (p < 0.05). This significant effect would well be a false positive given that no correction for multiple tests was applied. Thus, the behavioral data did not exhibit discernible idiosyncratic variation between conditions at the individual level.

Univariate fMRI: main experiment

We used the goodness-of-fit of the deconvolution model (r^2) to localize brain areas whose activity was modulated by our task (see Methods). At the group level, a network of frontoparietal areas, as well as the occipital visual areas, showed significant modulation by our task (Fig. 3A). In the frontal cortex, two activation loci were found along the precentral sulcus: a dorsal locus at the junction of precentral and superior frontal sulcus, the putative human frontal eye field (FEF) and a ventral locus near the inferior frontal sulcus, which we referred to as inferior frontal junction (IFJ). In the parietal cortex, activity ran along the intraparietal sulcus (IPS).

Fig. 3B plots mean fMRI time courses from a few visual areas, as defined by the retinotopic mapping procedure, and FEF, which was defined on the group-averaged r² map (Fig. 3A). These areas all showed robust task-related activity with nearly identical responses for the two attention conditions. For this ROI-based analysis, we did not find any significant difference in response amplitude between the two conditions in all brain areas we defined, even without correcting for multiple comparisons (other retinotopic areas include V2, V3, V3AB and the taskdefined area IFJ, data not shown). We also conducted whole-brain contrast analysis, comparing the two attention conditions. This analysis did not yield any significant difference in activation at the same statistical threshold as in Fig. 3A. Even at more lenient statistical thresholds, the contrast did not reveal any significant activation. Thus, the univariate analysis showed equivalent overall neural responses in the two attention conditions.

Multivariate fMRI: main experiment

We used whole-brain "searchlight" MVPA to determine whether patterns of neural activity can be used to predict the attended object. For this analysis, we only used fMRI data from trials on which participants did not make a response, to eliminate the contribution from processes associated with detecting the target and making a motor response (manual button press). This analysis revealed significant above-chance classification in FEF, IFJ, and anterior IPS regions at the group level (Fig. 4). These three areas were present in both hemispheres, showing bilateral symmetry. Significant classification was also observed in a lateral occipital region in the left hemisphere. Thus, multivoxel patterns in these areas can be used to decode the attended object, even though average univariate response was not informative about which object was attended.

For this multivariate analysis, we used trials where participants did not make a response to limit contributions from neural processes associated with target detection. However, the ignored distracter trials contained brightening events that were not completely equated between the two attention conditions. To explore whether such subtle perceptual difference contributed to our results, we performed an additional analysis using only the null trials for the searchlight classification. The two attention conditions were perceptually identical in these null trials. Here we obtained similar results as in the main analysis (see Supplementary Material). Furthermore, a perceptual difference would



Fig. 3. Univariate results: main experiment. A. Group averaged r² map visualized on the PALS atlas surface. Scale bar indicates r² value. B. Mean time course in select brain areas, V1, V4, MT+, and FEF. Error bars indicate s.e.m. across participants.



Fig. 4. Multivariate pattern analysis: main experiment. Group-averaged whole-brain classification map visualized on the PALS atlas surface. Scale bar indicates classification accuracy.

predict strong decoding performance in early visual cortex, which we did not observe. Thus we conclude that subtle perceptual difference due to distracter brightening could not account for our decoding results.

Behavior: control experiment

Because both objects started with a fixed set of features and remained still during the cue period, it is possible that the classifier was picking up pattern differences caused by attention to two sets of static features during the cue period. In principle, we can test the impact of the cue period on classification if we could separate fMRI response for the cue period vs. the rest of the trial. However, such an analysis is not practical due to the low temporal resolution of fMRI data and our event-related design. Hence we performed a control experiment to assess any potential contribution from feature-based attention.

In this experiment, a separate group of participants tracked one of four possible objects (Fig. 5). Objects 1 and 2 were the same as the main experiment, whereas Objects 3 and 4 started identically as Objects 1 and 2, respectively, but traversed the color and orientation dimension in an opposite direction. Two objects were presented on each trial, with two possible pairings: Objects 1 and 2, or Objects 3 and 4. An auditory cue instructed participants to track a particular object (see Methods).

The behavioral results paralleled those of the main experiment (Fig. 6). Participants detected the luminance increment equally well for the four objects (one-way repeated measures ANOVA for Hit-FA, p > 0.3). For each participant, we constructed contingency tables for both hits and false alarms for both pairing conditions (Object 1/2 and Object 3/4), yielding a total of 48 tables. We then performed a χ^2 test for independence and found 2 out of the 48 tests showed a significant effect (p < 0.05). Again, this analysis showed that there was no systematic performance difference at the individual participant level that can account for the multivariate classification results below.

Multivariate fMRI: control experiment

We performed both within-pairing and cross-pairing MVPA. We expect the within-pairing classification to yield similar results as the main experiment. Critically, if decoding in the main experiment was supported by featural differences during the cue period, we would expect significant decoding in the cross-pairing classification, because the visual stimuli were identical during the cue period for Object 1/2 trials and Object 3/4 trials. However, if decoding was supported by object-level differences, we would expect reduced or chance-level cross-pairing classification, because each object had a unique trajectory in the feature



Fig. 5. Schematic of a trial in the control experiment. Two superimposed Gabors continuously changed color, orientation, and spatial frequency. Objects 1 and 2 were defined as in the main experiment; Objects 3 and 4 traversed the color and orientation dimension in an opposite direction. On each trial, either Objects 1 and 2 or Objects 3 and 4 were shown (i.e., two possible pairings). An auditory cue was played at the beginning of the trial instructing participants which object to track on a particular trial.



Fig. 6. Behavioral results in the control experiment (same format as in Fig. 2).

space, and a classifier trained to discriminate one pair of objects should not generalize well to a different pair of objects.

For the within-pairing classification discriminating Object 1 vs. Object 2, we found significant classification in bilateral FEF, right IFI, and bilateral anterior IPS regions (Fig. 7A). In addition, bilateral superior temporal regions also exhibited significant classification. Similarly, the classifier discriminating Object 3 vs. Object 4 revealed significant classification in bilateral FEF, bilateral anterior IPS, and a left superior temporal region (Fig. 7B). We note that, however, the classifier for Object 3 vs. Object 4 showed a reduced spatial extent of above-threshold voxels than the classifier for Object 1 vs. Object 2. Importantly, we did not find any above-threshold decoding for the two cross-pairing classifiers at the same statistical threshold. Even at much more liberal thresholds (e.g., per-voxel p value at 0.05 with a cluster extent threshold of 30 voxels, which would lead to a whole-brain false positive rate much greater than 0.05), we still failed to find any significant voxels anywhere in the brain. These results suggest that the significant decoding observed in the main experiment and within-pairing classification cannot be attributed to attention to static features during the cue period.



Fig. 7. Multivariate pattern analysis: control experiment. Group-averaged whole-brain classification map visualized on the PALS atlas surface. (A) Performance of a classifier trained to classify Object 1 vs. Object 2. (B) Performance of a classifier trained to classify Object 3 vs. Object 4. (C) Summary of activation loci across three searchlight analysis.

We performed a simple overlap analysis, to find areas that showed significant decoding across the three sets of searchlight analyses (main experiment and the two within-pairing classifications). Common areas across the analyses were labeled schematically in Fig. 7C and they include bilateral anterior IPS, bilateral FEF, and right IFJ. We note that the right IFJ failed to reach significance in Object3/4 classification (Fig. 7B). However, given the overall low classification accuracy in this analysis (see Discussion), and prior research showing the importance of this area in top-down control (Baldauf and Desimone, 2014; Zanto et al., 2011), it is likely that the right IFJ plays an important role in object-based selection.

Discussion

We found that tracking a dynamic object through feature space evoked strong neural activity in dorsal frontoparietal and occipital visual areas. This finding is consistent with many prior studies showing the involvement of these areas during visual attention tasks. However, only in a subset of these active areas, notably FEF, IFJ, and anterior IPS, can we use neural activity to reliably decode the specific object that was attended on a given trial. This finding suggests that neural signals in these areas encode attentional priority for whole perceptual objects.

Isolating neural signals for object-based selection

It has been difficult to isolate object-level selection without influence from feature-level selection due to the fact that objects are always composed of features. As a result, many studies of object-based attention can be attributed to feature-based attention.

For example, in a popular paradigm where participants were instructed to attend to either a face or a house in a compound image (e.g., Baldauf and Desimone, 2014; O'Craven et al., 1999; Serences et al., 2004), they could potentially attend to different features present in the face vs. house (e.g., a face contains many curved parts whereas a house contains mostly rectilinear parts). In our own previous study of object-based attention, we attempted to equate the features between two objects by using the same orientated lines to construct simple geometric shapes (Hou and Liu, 2012). However, second-order features such as vertices and line junctions could not be equated between objects, raising the possibility that participants attended to these features during the task. Consistent with this interpretation, we found significant decoding of the attended object throughout early visual areas (V1-V4). Similarly, another recent study using the superimposed face-house stimuli found that neural activity in early visual cortex (V1-V4) can be used to decode the attended face or house (Cohen and Tong, 2015). This finding led the authors to suggest that object-based attention relies on selection of elemental features in early visual areas. While this conjecture is quite plausible in many ecological situations, as different objects are usually composed of different features, we believe that these results attest to the adaptive nature of selection, that is, object-based selection can be facilitated by the selection of features if such selection is helpful. However, we also know that attention is highly flexible such that it can select objects composed of identical and dynamic features. Indeed, many psychological theories have highlighted the importance of objects as the unit of selection (Duncan, 1984; Kahneman et al., 1992; Pylyshyn and Storm, 1988). Yet the neural basis of object-based selection has proved difficult to isolate given the potential contribution from feature-level selection.

Here we adopted a psychophysical paradigm that has been used previously to demonstrate "pure" object-level selection (Blaser et al., 2000). We used dynamic objects that evolved continuously along trajectories in three feature dimensions: color, orientation, and spatial frequency. Both objects traversed through exactly the same feature values along all feature dimensions, such that no single feature dimension or feature value can distinguish the two objects. This task thus required participants to attend to the whole object instead of elemental features, which was supported by strong psychophysical evidence from the original Blaser et al. (2000) study. For example, they found a same-object advantage in change detection, and also showed that participants could not attend to two objects simultaneously through the analysis of response patterns. Furthermore, we also performed a behavioral experiment with identical setting as the main experiment, except that participants were instructed to detect the target brightening on either the cued or uncued object. We found a strong validity effect such that targets were better detected on the cued than the uncued object, demonstrating that the cued object was assigned with a higher attentional priority than the uncued object (see Supplementary Material). In the fMRI experiment, we adapted this task and further restricted our imaging data analysis to trials in which participants did not make a response, eliminating contributions from decision- and motor-related neural signals associated with target detection. Thus, the neural signals we measured were closely related to attentive tracking of perceptual objects with minimal influence from other processes.

However, there was still a potential confound in that the objects always started with the same set of features on each trial, such that neural decoding could be due to attention to these features during the cue period. We think this explanation is unlikely for two reasons. First, we used average fMRI response over the entire trial to perform the classification analysis, and the cue period only accounted for a small fraction of time (<10%) in a trial. Second, feature-based attention is known to modulate early visual cortex in a consistent manner that should lead to robust classification of the attended feature in occipital visual areas (Kamitani and Tong, 2005, 2006; Liu et al., 2011; Serences and Boynton, 2007), which we did not observe. The lack of significant decoding in early visual cortex hence implies that selection in our task was not based on modulation of a particular feature dimension or value. To further assess any potential contribution of feature-level selection to our results, we conducted a control experiment. Here we introduced two additional dynamic objects that started in identical form as the original objects but evolved along the color and orientation trajectories in opposite directions (spatial frequency is not a circular dimension, thus preventing a similar manipulation). The control experiment largely replicated the main experiment in that the within-pairing classification showed significant decoding in the same set of frontoparietal areas. We also found significant decoding in the superior temporal region, which was presumably driven by different neural patterns evoked by the auditory cues. A somewhat unexpected finding is the stronger classification performance when decoding Object 1 vs. 2 than Object 3 vs. 4. This suggests that participants formed more distinct representations during Object 1/2 trials than during Object 3/4 trials. It is not clear what caused such a difference. It is possible that learning four temporally-varying objects was difficult, and participants prioritized the Object 1/2 pair over the Object 3/4 pair during training. Most importantly, however, we did not find any significant decoding for cross-pairing classification, suggesting that the featural difference during the initial cue period was not sufficient to support decoding. We thus conclude that featurebased attention cannot account for the decoding results in our experiment.

In addition to feature-level selection, we also need to consider whether classification can be driven by sensory responses to the cue. This possibility would predict significant decoding in sensory areas processing the cue. In the main experiment, we did not find significant decoding in early visual areas, presumably due to the small size and brief duration of the visual cue compared to the grating stimulus, such that cue-evoked response was masked by stimulus-evoked response in the visual cortex. In the control experiment, we did observe significant decoding in the auditory cortex, presumably because the cue was the only auditory event in a trial, and we captured relatively clean sensory responses to the cue in the auditory cortex. Using auditory cues eliminated any image-level differences caused by visual cues, and the fact that we observed significant decoding of the attended object in similar frontoparietal areas as the main experiment provided converging evidence that these areas represent attentional priority for visual objects, regardless of the modality of the instructional cue.

Neural representation of object-level attentional priority

Previous research on the neural basis of object-based attention has employed stimuli and task that were amenable to alternative selection strategy such as space-based and feature-based selection. Here we overcame the intrinsic difficulty in isolating object-based selection by using an experimental design that equated features and locations between objects, as well as careful control analyses and experiments. This study thus provides the first unambiguous evidence that several frontoparietal areas contain neural signals that encode attentional priority for perceptual objects.

What is the nature of the priority signals for the abstract objects in our study? There are at least two possibilities. On the one hand, because participants were trained to associate feature trajectories with objects, the neural signals could reflect sequential activation of different features along these trajectories in the feature space. On the other hand, these signals could represent feature-invariant identity information for the attended object, such as an "object file" representation, which was postulated based on behavioral studies (Kahneman et al., 1992), It seems unlikely that our results were due to sequential selection of features for two reasons. First, sequential activation of selected features in the course of a trial would be difficult to capture with the low sampling rate of fMRI-only three time points were measured per trial in which objects spanned many features. Second, our estimate of single-trial response essentially averaged across time points within a trial, likely blurring any within-trial dynamics. Thus we believe our results were more likely due to the maintenance of abstract identity information, as postulated by the theory of object file (Kahneman et al., 1992). Such a scenario is also consistent with single-unit recording studies showing analogous brain areas in the monkey brain can represent visual categories defined by arbitrary features (Freedman and Assad, 2006; Freedman et al., 2001). However, we should note that the above two scenarios are not mutually exclusive, and further research is needed to elucidate the contribution of dynamic vs. invariant representation during object-based selection. In addition, given that similar brain areas are also involved in spatial attention (Bisley and Goldberg, 2010), future work should also examine the relationship between neural encoding of spatial priority and object priority.

We found neural activity in three areas, FEF, IPS and IFI, can be used to decode the attended object. While previous studies have generally emphasized the involvement of dorsal areas FEF and IPS in top-down attention (Corbetta and Shulman, 2002), recent work has also highlighted the role of IFI in attentional control. For example, Baldauf and Desimone (2014) found increased functional connectivity between IFJ and posterior visual areas FFA and PPA during a selective attention task using superimposed face/house stimuli. Zanto et al., applied transcranial magnetic stimulation to IFJ and found impaired performance in a working memory task that required attention to color during encoding (Zanto et al., 2011). In our own previous study, we also found reliable decoding of the attended feature in IFJ (Liu et al., 2011). The IFJ is part of the "multiple-demand" system that is consistently activated in a diverse set of cognitive tasks (Duncan, 2010) and is also considered to be a shared node between the dorsal and ventral attention networks in an updated model of attentional control (Corbetta et al., 2008). Thus the IFJ could be a critical area in exerting top-down control of attention, presumably in coordination with more dorsal areas FEF and IPS. The precise functions subserved by these distinct cortical areas remain to be clarified in future studies.

Conclusions

In summary, in a task designed to isolate object-level selection, we found that neural patterns in a set of frontoparietal areas can be used to decode the attended object. These neural activities likely represent attentional priority for perceptual objects and exert top-down control in selecting whole perceptual objects.

Acknowledgment

I thank Scarlett Doyle, Sarah Young, Srinivasa Chaitanya, Chris Kmiec, and Nicci Russell for assistance in data collection. I also wish to thank the High Performance Computing Center at Michigan State University for access to computing services. This work was supported by a NIH grant (R01EY022727).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx. doi.org/10.1016/j.neuroimage.2016.01.034.

References

- Baldauf, D., Desimone, R., 2014. Neural mechanisms of object-based attention. Science 344 (6182), 424–427.
- Bisley, J.W., Goldberg, M.E., 2010. Attention, intention, and priority in the parietal lobe. Annu. Rev. Neurosci. 33, 1–21.
- Blaser, E., Pylyshyn, Z.W., Holcombe, A.O., 2000. Tracking an object through feature space. Nature 408 (6809), 196–199.
- Brewer, A.A., Liu, J., Wade, A.R., Wandell, B.A., 2005. Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat. Neurosci. 8 (8), 1102–1109.
- Carrasco, M., 2011. Visual attention: the past 25 years. Vis. Res. 51 (13), 1484–1525. Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27:21–27:27.
- Cohen, E.H., Tong, F., 2015. Neural mechanisms of object-based attention. Cereb. Cortex 25 (4) 1080–1092
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3 (3), 201–215.
- Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. Neuron 58 (3), 306–324.
- Deco, G., Rolls, E.T., 2004. A neurodynamical cortical model of visual attention and invariant object recognition. Vis. Res. 44 (6), 621–642.
- Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18 (1), 193–222.
- DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al., 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. U. S. A. 93 (6), 2382–2386.
- Duncan, J., 1984. Selective attention and the organization of visual information. J. Exp. Psychol. Gen. 113 (4), 501–517.
- Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. 14 (4), 172–179.
- Ekstrom, L.B., Roelfsema, P.R., Arsenault, J.T., Bonmassar, G., Vanduffel, W., 2008. Bottomup dependent gating of frontal signals in early visual cortex. Science 321 (5887), 414–417.
- 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 (2), 181–192.
- Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191.
- Freedman, D.J., Assad, J.A., 2006. Experience-dependent representation of visual categories in parietal cortex. Nature 443 (7107), 85–88.
- Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2001. Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291 (5502), 312–316.
- Gardner, J.L., Sun, P., Waggoner, R.A., Ueno, K., Tanaka, K., Cheng, K., 2005. Contrast adaptation and representation in human early visual cortex. Neuron 47 (4), 607–620.
- Hou, Y., Liu, T., 2012. Neural correlates of object-based attentional selection in human cortex. Neuropsychologia 50 (12), 2916–2925.
- Kahneman, D., Treisman, A., Gibbs, B.J., 1992. The reviewing of object files: object-specific integration of information. Cogn. Psychol. 24 (2), 175–219.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.
- Kamitani, Y., Tong, F., 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. 16, 1096–1102.
- Kastner, S., Ungerleider, L.G., 2000. Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci. 23, 315–341.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103 (10), 3863–3868.
- Liu, T., Hou, Y., 2013. A hierarchy of attentional priority signals in human frontoparietal cortex. J. Neurosci. 33 (42), 16606–16616.
- Liu, T., Hospadaruk, L., Zhu, D.C., Gardner, J.L., 2011. Feature-specific attentional priority signals in human cortex. J. Neurosci. 31 (12), 4484.
- Mendoza-Halliday, D., Torres, S., Martinez-Trujillo, J.C., 2014. Sharp emergence of featureselective sustained activity along the dorsal visual pathway. Nat. Neurosci. 17 (9), 1255–1262.

Moore, T., Armstrong, K.M., 2003. Selective gating of visual signals by microstimulation of frontal cortex. Nature 421 (6921), 370-373.

Moore, T., Fallah, M., 2004. Microstimulation of the frontal eye field and its effects on covert spatial attention. J. Neurophysiol. 91 (1), 152-162.

Nestares, O., Heeger, D.J., 2000. Robust multiresolution alignment of MRI brain volumes. Magn. Reson. Med. 43 (5), 705–715.

- O'Craven, K.M., Downing, P.E., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584-587.
- Pylyshyn, Z.W., Storm, R.W., 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. Spat. Vis. 3 (3), 179–197.
- Reynolds, J.H., Chelazzi, L., 2004. Attentional modulation of visual processing. Annu. Rev. Neurosci 27 611-647
- Scholl, B.J., 2001. Objects and attention: the state of the art. Cognition 80 (1–2), 1–46.
- Scolari, M., Ester, E.F., Serences, J.T., 2014. Feature- and object-based attentional modulation in the human visual system. In: Nobre, K., Kastner, S. (Eds.), Oxford Handbook of Attention. Oxford University Press, Oxford, UK.
- 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 objectbased attention in human cortex. Cereb. Cortex 14, 1346-1357.

- Sereno, M.I., Dale, A.M., Reppas, I.B., Kwong, K.K., Belliveau, I.W., Brady, T.I., et al., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268 (5212), 889-893.
- Todd, M.T., Nystrom, L.E., Cohen, J.D., 2013. Confounds in multivariate pattern analysis: theory and rule representation case study. NeuroImage 77, 157-165.
- Van Essen, D.C., 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. NeuroImage 28 (3), 635–662. Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neu-
- ron 56 (2), 366-383.
- Watson, J.D., Myers, R., Frackowiak, R.S., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., et al., 1993. Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb. Cortex 3 (2), 79–94.
- Watt, R.J., 1998. Visual Processing: Computational, Psychophysical and Cognitive Research. Erlbaum, Hillsdale, NJ.
- Wolfe, J.M., 1994. Guided Search 2.0: a revised model of visual search. Psychon. Bull. Rev. 1 202-238
- Zanto, T.P., Rubens, M.T., Thangavel, A., Gazzaley, A., 2011. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. Nat. Neurosci. 14 (5), 656-661.