frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 June 2012
doi: 10.3389/fnhum.2012.00187

Goal-directed attention alters the tuning of object-based representations in extrastriate cortex

Anthony J.-W. Chen[1,2,3,4]*, Michael Britton[1,4], Gary R. Turner[4,5], Jason Vytlacil[4], Todd W. Thompson[6] and Mark D'Esposito[1,3,4]

[1] Department of Neurology, Veteran's Administration Northern California Health Care System, Martinez, CA, USA
[2] Department of Neurology, Veteran's Administration Medical Center, San Francisco, CA, USA
[3] Department of Neurology, University of California, San Francisco, CA, USA
[4] Neuroscience Institute, University of California, Berkeley, CA, USA
[5] Department of Psychology, University of Toronto, Toronto, ON, Canada
[6] Massachusetts Institute of Technology, Cambridge, MA, USA

Edited by:
Srikantan S. Nagarajan, University of California, USA

Reviewed by:
Nikolaus Kriegeskorte, Medical Research Council Cognition and Brain Sciences Unit, UK
William A. Cunningham, The Ohio State University, USA

***Correspondence:**
Anthony J.-W. Chen, Helen Wills Neuroscience Institute, Program in Rehabilitation Neuroscience, 132 Barker Hall, University of California, Berkeley, CA 94720, USA.
e-mail: rehabneuroscience@gmail.com

Humans survive in environments that contain a vast quantity and variety of visual information. All items of perceived visual information must be represented within a limited number of brain networks. The human brain requires mechanisms for selecting only a relevant fraction of perceived information for more in-depth processing, where neural representations of that information may be actively maintained and utilized for goal-directed behavior. Object-based attention is crucial for goal-directed behavior and yet remains poorly understood. Thus, in the study we investigate how neural representations of visual object information are guided by selective attention. The magnitude of activation in human extrastriate cortex has been shown to be modulated by attention; however, object-based attention is not likely to be fully explained by a localized gain mechanism. Thus, we measured information coded in spatially distributed patterns of brain activity with fMRI while human participants performed a task requiring selective processing of a relevant visual object category that differed across conditions. Using pattern classification and spatial correlation techniques, we found that the direction of selective attention is implemented as a shift in the tuning of object-based information representations within extrastriate cortex. In contrast, we found that representations within lateral prefrontal cortex (PFC) coded for the attention condition rather than the concrete representations of object category. In sum, our findings are consistent with a model of object-based selective attention in which representations coded within extrastriate cortex are tuned to favor the representation of goal-relevant information, guided by more abstract representations within lateral PFC.

Keywords: selective attention, visual attention, prefrontal cortex, visual objects, pattern analysis, functional MRI, working memory

INTRODUCTION

Humans survive in environments that contain a vast quantity and variety of visual information. Much more information enters the nervous system than may be useful or within our capacity to process. To guide behavior based on one's goals or intentions (referred to as top-down processes), we must selectively process only what is relevant and ignore what is not relevant. However, information in the visual world may be parsed as relevant or non-relevant based on any number of different possible divisions. Relevant visual information may be determined by properties such as a particular spatial location or a specific object. The same location or object in our visual world may differ in its relevance depending on the context of the situation. For example, in one moment, a tourist may be interested in learning the faces of fellow tour members. Later she may be more interested in identifying scenic views for photo-taking opportunities. What neural mechanisms mediate selective attention to relevant visual information?

In this study, we focus on identifying the neural mechanisms underlying object-based selective attention. Several models regarding how the object category of perceived visual information is represented in the brain have been proposed, and these models for object perception provide possible foundations for investigating mechanisms by which attention guides information processing. One model proposes that visual objects are encoded in a modular architecture within visual cortex, with a limited number of identifiable regions that show differential response magnitudes to viewing of objects of different categories (Kanwisher et al., 1997; Aguirre et al., 1998; Epstein and Kanwisher, 1998; Downing et al., 2006). An alternative model proposes that object categories are encoded in visual cortex in widely distributed and spatially overlapping representations (Haxby et al., 2001; Hanson

et al., 2004; O'Toole et al., 2005). Based on human fMRI data, it was demonstrated that representations of object category could be identified in patterns of activity even within visual cortical regions that did not respond maximally to a particular category. Thus, object information was represented in codes not distinguished by localized regions, but by the combinatorial patterns of activity distributed across visual cortex. This type of organization of visual information offers an account regarding how an unlimited number of categories could be uniquely encoded in the brain.

Models of selective attention have put forth mechanisms in which top-down attentional signals bias competition between overlapping representations of visual information toward what is relevant for current goals (Desimone and Duncan, 1995; Beck and Kastner, 2009). In localized, object-selective visual areas (e.g., face or place areas), attention modulates both the magnitude (e.g., O'Craven et al., 1999; Gazzaley et al., 2005a,b) and selectivity (e.g., Murray and Wojciulik, 2004) of cortical activity. This presumably reflects some components of the effects of such top-down bias signals. However, given that the neural representations of category-specific information extends well beyond these localized regions in visual cortex, our aim in this study was to determine the effect of top-down attentional signals on spatially distributed object representations. Mechanisms of top-down control via distributed codes would afford the flexibility and precision required for specifying a wide range of possible targets of attention. Whether and by what mechanisms object-based attention acts at this level within visual cortex has not been directly tested.

It is proposed that top-down attentional signals direct distributed neural codes toward representing goal-relevant information. For example, if an individual pays selective attention only to a scene (and not a face) when taking a photograph, distributed neural codes would be altered by top-down signals to better represent scene information over the face information (illustrated in **Figure 1A**). Despite equivalent bottom-up input from the same external images, selective attention only to a face and not the scene would alter distributed neural codes to better represent face information over the scene information (**Figure 1B**). Without engagement of selective processing, such as when the face and scene are equally task-relevant, visual images would be represented as patterns of visual cortical activity that are a composite of the representations for both categories (**Figure 1C**). The effectiveness of this goal-based direction of distributed neural codes may influence the efficiency of higher level processes which depend on the clear "reading" of goal-relevant visual representations by other brain regions (Jazayeri, 2008).

One methodological limitation of previous fMRI or monkey single-unit studies of object-based attention is that they have predominantly relied on univariate measurements that cannot directly quantify information contained in spatially distributed neural representations. Functional MRI allows simultaneous measurements of activity throughout the brain, and pattern classifier and spatial correlation methods take into account the richness of information that may be encoded in multi-voxel patterns of activity. Thus, in this study we utilize these methods to determine the properties of responses that are distributed throughout extrastriate cortex, and not just localized to particular "category-specific" regions. We specifically set out to test the effect of selective attention on patterns of visual cortical activity from individuals engaged in tasks with differential information processing demands based on the relevance of an object

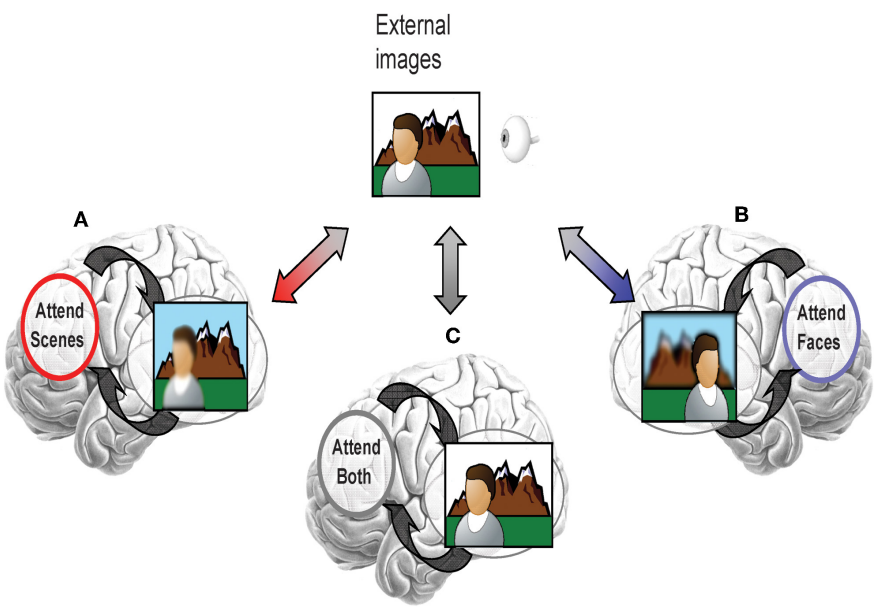

FIGURE 1 | Illustration of the effect of selective attention on directing distributed neural codes within extrastriate cortex to bias representation toward information that is goal-relevant. (A) If scenes were more relevant, then the network codes would be tuned to favor representation of scenes. However, if attention were selectively directed toward faces, then, **(B)** network codes would be tuned to favor representation of scenes. **(C)** If both faces and scenes are task-relevant, then there should be no bias toward one or another category.

category. In a behavioral task in which participants are presented with both scenes and faces in every trial (see **Figure 2**), three different attention conditions demanded either non-selective processing of presented stimuli (e.g., attend to both faces and scenes), or selective processing of one category of information over the other (e.g., attend to scenes and not faces). Selective attention toward one or the other category would be predicted to produce measurable activity patterns within extrastriate cortex that are distinctive from each other. Non-selective attention to both categories should result in an intermediate pattern composed of a non-biased combination of the representations for the two categories. Given that bottom-up demands are held constant in all task conditions, such findings would further support the existence of top-down attentional signals selectively acting upon these distributed neural representations.

Determination of the source of top-down signals is also necessary for understanding how processing of visual information is modulated by attention. Lateral prefrontal cortex (PFC) is one proposed source thought to encode abstract goals or states (Desimone, 1998; Miller and Cohen, 2001; Miller and D'Esposito, 2005; Fuster, 2009). If the modulation of object representation within extrastriate cortex is a result of direction by lateral PFC, then patterns of PFC activity should encode measurable information regarding top-down attention demands. However, we would predict that PFC codes representations of information that differs in nature from extrastriate cortex, consistent with a different role in the selection process.

METHODS

PARTICIPANTS

Twenty-six healthy adult subjects with no history of neurologic, psychiatric, or vascular disease participated in this study after informed consent as approved by the University of California, San Francisco and Berkeley, as well as VA Medical Center in San Francisco Institutional Review Boards. Data from six subjects were discarded due to motion or scanner-related artifacts, yielding 20 subjects (seven female, mean age 21.1 years), mean education 15.6 years (range 13–18).

BEHAVIORAL TASK DESIGN

In each trial of this event-related design, four stimuli (two faces and two scenes) were presented serially in the center of the screen (800 ms each, over 4 s) followed by a delay period (9 s) and then a recognition test consisting of a probe stimulus (face or scene, corresponding to the relevant stimulus category, or either one in Both) to which participants were required to respond with a button press whether the probe matched one of the previously presented stimuli (adapted from Gazzaley et al., 2005a,b; see **Figure 2**). The task was divided into four attention condition blocks. In the Selective conditions, subjects were instructed to attend and remember only one category of images (Attend Faces or Scenes). In the non-selective condition, subjects were instructed to attend and remember all images (Attend Both). Task conditions vary only in the instructions to attend scenes, attend faces, or both. Passive viewing was included as a non-selective task in which subjects viewed stimuli presented and then responded to the direction of an arrow during the probe period, but the current analyses focused on the three conditions that required working memory.

Data were acquired during six scanner runs, each containing all four block conditions in counter-balanced order. Each block contained five trials of a given task condition, each trial lasting 26 s, yielding a total of 30 trials per condition. Stimuli were grayscale images of human faces and natural scenes (subtending $5 \times 6^{\circ}$ of visual angle). For faces, details outside the hairlines were blurred. Stimuli were novel for every trial. Stimulus

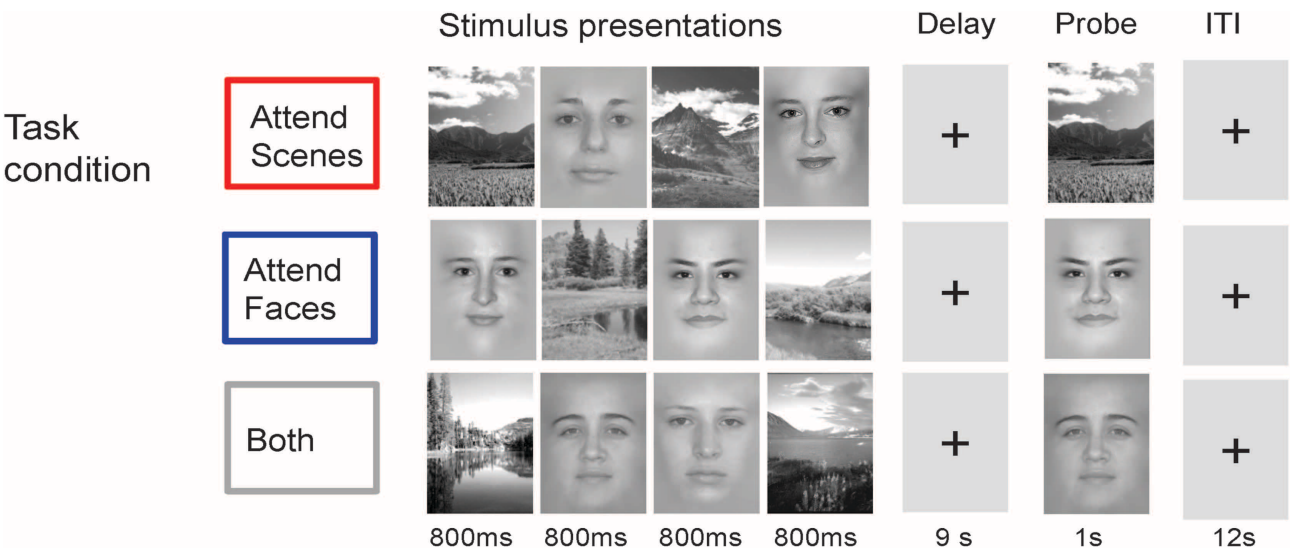

FIGURE 2 | Cognitive Task. In each trial, subjects view a centrally-presented series of visual stimuli consisting of two faces and two scenes over 4 s, followed by a delay period of 9 s and then a probe image to test delayed recognition. Task conditions vary only in the instructions to attend scenes, attend faces, or attend both scenes and faces.

presentation orders were pseudo-randomized and category orders were counter-balanced across trials. To ensure constant motivation and effort across the task, feedback, and point rewards were provided at the end of each task block.

Subsequent memory assessments

After scanning, participants were asked to complete a memory questionnaire. Subjects were asked to rate their certainty on a 1–4 scale that each of 150 probe images had or had not appeared during the experiment. The probe set consisted of evenly divided samples of face and scene stimuli, with samples from each attention condition as well as novel stimuli as foils, presented in a predetermined pseudo-random order. Stimuli derived from the scanner task were limited to those that had not presented twice in any trial (i.e., during both encoding and probe stages). Images were viewed one at a time and the presentation was self-paced. Familiarity ratings for relevant and non-relevant stimuli in the selective attention conditions were compared with each other, as well as with ratings for stimuli in the non-selective conditions. Familiarity ratings from the three classes of stimuli were initially compared in an ANOVA, and then we performed pair-wise comparisons between the ratings from each pair of conditions using paired *T*-tests.

FUNCTIONAL MRI ACQUISITION AND DATA PROCESSING

MRI data were acquired with a Varian INOVA 4T scanner equipped with a TEM send-and-receive RF head coil using a two-shot echo-planar imaging (EPI) sensitive to blood oxygen level dependent contrast (TR = 2000 ms, TE = 28 ms, FOV = 22.4 cm^2, matrix 64 × 64, for an in-plane resolution of 3.5 mm, eighteen 5 mm axial slices, 0.5 mm inter-slice gap). Data pre-processing included slice-timing correction, adjacent half k-space shots were interpolated to decrease the sampling interval to half of the total repetition time, re-alignment, correction for linear drift, normalization of mean signal intensity and variance, and high-pass filtering (period 128 s) using the Statistical Parametric Mapping (SPM2) package. As standard in many fMRI analyses, signal intensity for each voxel is normalized by global signal intensity and variance. In addition, we normalized signal intensity between the samples from the different task conditions, so that the mean signal intensities across the conditions were equal. Having the difference in mean signal intensity equal to zero allowed us to perform pattern analyses that would detect differences in the "patterns" and not mean signals. This helped to remove the contribution of mean signal differences to classifying the patterns from different conditions.

FUNCTIONAL MRI ANALYSES

Regions of interest

For all analytic methods, signal intensity patterns for each condition were extracted from regions of the lateral PFC and extratrastriate cortex. The extrastriate cortex mask was comprised of fusiform, parahippocampal, and lingual gyri, defined based on the following boundaries: superior (lateral ventricles), anterior (temporo-occipital fissure), posterior (intra-occipital sulcus) on the MNI-152 T1 template and back-projected to each individual's native space. For lateral PFC ROIs, we based samples on automated anatomical labeling maps of Brodmann's areas for reproducibility and these were back-projected from standardized space to each subject's native space (Chris Rorden, MRICro, http://www.sph.sc.edu/comd/rorden/template.html) (Tzourio-Mazoyer et al., 2002). Dorsolateral, ventrolateral, and frontopolar regions of PFC were composed of BA 9 and 46; BA 45 and 47, and BA 10, respectively. Reverse normalization was performed by first determining normalization transformation parameters for transforming from the native space EPI volumes for each individual to the standardized template using a 12-parameter affine transformation and non-linear estimation of deformations using SPM2, then reversing the normalization parameters to back project the standardized anatomical masks to native space (SPM http://www.fil.ion.ucl.ac.uk/spm/).

General multi-voxel pattern classification methods

These analyses utilized a multi-layer perceptron (MLP) with three layers (input layer corresponding to the voxel inputs of signal intensity, one hidden layer with 10 nodes, and three output nodes corresponding to the attention conditions to be classified) utilizing logistic activation functions in back-propagation, feedforward algorithms, implemented in a pattern classifier toolbox (Multi-voxel Pattern Analysis Toolbox, Ken Norman et al., Princeton University, with Netlab Neural Network Toolkit, Chris Bishop). This configuration was chosen based on pilot data showing higher classification generalization accuracies compared to zero hidden layers. However, it should be noted that the quantitative comparison of classification accuracies *per se* is not relevant to the hypothesis testing in this study. Rather, it will be the relative classification assignments (mis-classification in particular attention conditions) that are of greatest interest. Details are provided here for reproducibility, although it would be expected that the same conclusions would be drawn with different classifier configurations or methods, in that the qualitative relationships between patterns should be consistent even though the quantitative numbers may differ (O'Toole et al., 2007).

Inputs to the pattern classifier were multi-voxel signal intensity patterns from data restricted to the pre-defined masks (extrastriate cortex mask and left and right PFC masks, each analyzed separately). Data were not spatially smoothed and remained in each subject's native space. In order to select time points representative of the peak response to presented stimuli, stimulus onsets for the four stimuli during the 4 s encoding period for each trial were convolved with a double-gamma canonical hemodynamic response function and time points for which the result exceeded a threshold of 0.5 (normal units) were taken as samples for iterative training and testing procedures, while other time points were set to 0. This initial modeling resulted in standard time points being selected for sampling the experimental data—these time points corresponded to the peak of the modeled stimulus response for the 4 s encoding period, and other time points, including those in the probe period, were set to zero such that no other time points were included in the analyses. The voxel-by-voxel signal intensity values from these time points were then entered into the MVPA and spatial correlation analyses. An "N-1" leave-one-out bootstrap method for out-of-sample generalization was applied to the six runs as an unbiased test of the classifier (Hanson et al., 2004).

The procedure was repeated for 10 cycles (connection weights were initialized to random values for each cycle), and the average output was taken as a more precise estimate than single cycles.

Classification of attention condition

We first tested whether information that distinguishes attention conditions is present in extrastriate cortex and PFC by determining whether patterns from different attention conditions were distinguishable by the MLP. Classification accuracy was compared against rates when the condition samples were scrambled (on average occurring at chance, 33%). This simply confirmed that information relevant to attention condition was encoded in the evaluated activity patterns, and that the MLP could detect this information content and use it to discriminate activity patterns.

Determination of the distinctiveness of activity patterns

Determination of the distinctiveness of activity patterns of primary interest were measurements informative of the distinctiveness of activity patterns between the different attention conditions. One method is to calculate the rate at which patterns from a given attention condition are mis-classified as another by a pattern classifier (see **Figure 3**). Distinctiveness of two patterns is reflected in how likely the classifier is to classify two patterns as representing the same information, in this case, attention condition. In particular, the relative rates at which two patterns are mis-classified as one another reflects how distinctive the patterns are. This application differs from the more common usage of pattern classification methods, where the existence of information within the tested patterns is tested by classification rates above chance, rather than how distinctive different patterns are from each other. The lower the distinctiveness between two patterns, the more likely the pattern classifier will mis-classify one pattern as the other. The higher the distinctiveness between two patterns, the less likely the pattern classifier is to mis-classify one pattern as the other. Related concepts have been described as "confusability" (of patterns or information) (O'Toole et al., 2005) or "similarity" (Norman et al., 2006).

After the MLP was trained to recognize patterns associated with each attention condition, test sample patterns were provided as inputs to the MLP. When multi-voxel patterns from each attention condition are entered as test samples, the classifier either correctly guesses the attention condition or mis-classifies the pattern as one of the other conditions (see **Figure 3**). The rates at which the patterns were mis-classified as either of the other two conditions were quantified (outputs at right side of **Figure 3**).

Hypothesis testing using indices of distinctiveness

If selectively directing attention to one or the other category biases visual cortex patterns to favor the representation of the one that is goal-relevant, this should result different patterns of mis-classification rates in the two selective attention conditions, relative to the non-selective attention condition. Thus, patterns for opposing attention conditions (e.g., faces vs. scenes) would be relatively more distinctive from each other (**Figure 1B** vs. **1C**) than from the intermediate non-selective condition (e.g., Attend Both; **Figure 1A** vs. **1B** or **1A** vs. **1C**). Chi-Square tests were used to test for potential differences in classification assignments across the three attention conditions. In particular, we tested for differences in mis-classification rates for each pair of conditions, e.g., Faces mis-classified as Scenes vs. Faces mis-classified as Both. Comparisons were conducted in the PFC and extrastriate cortex, and to address the issue of multiple comparisons, we adjusted our statistical threshold of $p < 0.05$ to reflect the 13 comparisons

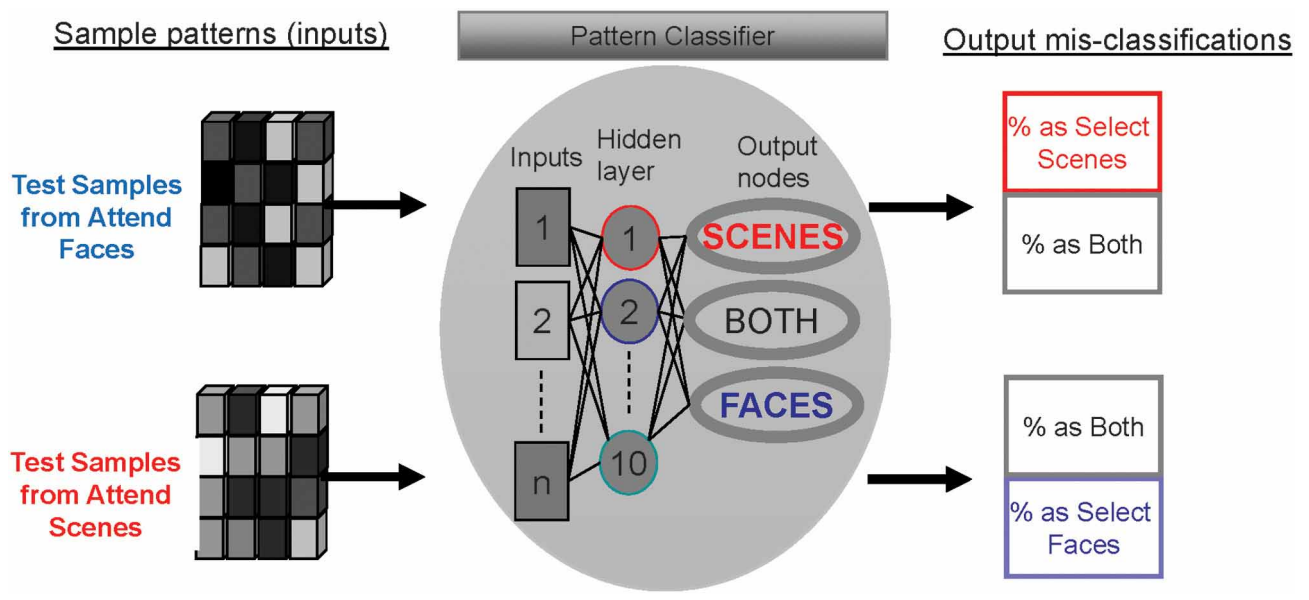

FIGURE 3 | Method for determining the distinctiveness of activity patterns for each attention condition using mis-classification rates of a multi-layer perceptron.

conducted (for 12 PFC masks and one extrastriate cortex mask), resulting in a corrected threshold of $p < 0.004$.

Exclusion of global differences in activation between attention conditions

In order to conclude that selective attention modulates spatially distributed activity patterns it is necessary to exclude the possibility that these activity patterns differ simply due to global differences in magnitude of activation. We therefore tested visual cortex activity patterns after specifically excluding the potential contributions of global signal intensity differences across conditions by normalizing signal intensities across attention conditions. Thus, we calculated *z*-scores for each blocked condition and each anatomical mask by normalizing every voxel signal intensity by the mean and standard deviation of signal intensities for all the voxels in each mask for each block. We performed pattern classification analyses in both the PFC and extrastriate cortex using these *z*-scored data following the same procedures described above.

Exclusion of category-selective extrastriate cortex regions

We also determined to what extent attentional modulation of extrastriate cortex was driven by information carried by localized object "category-selective" regions (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2006) as compared to spatially, distributed regions. Based on prior findings from a localization-based ROI approach (e.g., Gazzaley et al., 2005a,b), it might be the case that attention alters information codes simply by modifying regions with the highest category-selectivity. In order to test for this possibility, we first calculated the object selectivity of voxels (i.e., differential activation to viewing of faces vs. scenes), following procedures as previously published (Gazzaley et al., 2005a,b). Subjects performed an independent block design 1-back matching task involving presentations of blocks (16 s) of scene stimuli alternating with blocks of face stimuli. Contrasts of the faces and scenes blocks were calculated and the resulting *t*-statistics were used as an index of category selectivity. This analysis led to the identification of localized category-selective regions corresponding to the "parahippocampal place area" (PPA) and "fusiform face area" (FFA). These regions were individually functionally defined as the top seven voxels showing the greatest selectivity in responses for faces or scenes (corresponding to the FFA and PPA, respectively). These regions were then excluded from the pattern classification analyses, and the classification procedures were repeated.

A second approach we implemented for investigating the extent attentional modulation of extrastriate cortex is driven by localized object "category-selective" regions was to stratify extrastriate cortex voxels based on their degree of category selectivity. Thus, extrastriate cortex voxels were sorted into 10th percentile strata based on the degree of differential activation for faces and scenes (stratified from "most selective" to "least selective" based on *t*-value in the contrast of faces vs. scenes). The "FFA" and "PPA" ROIs described above contained voxels from the most selective bin. Pattern classification was performed when each 10th percentile stratum were excluded, thereby testing the contribution of each stratum to the measured information representations. The resulting classifier accuracy rates were tested by repeated measures ANOVA, with one factor (strata) with 10 levels. Follow-up analysis of significance was based on dependent sample t tests of accuracy rates for pairs of strata.

Spatial correlations

We confirmed that the results from the pattern classifier methodology were valid across methods by also determining the distinctiveness of activity patterns based on measurement of the correlations of distributed spatial patterns. Correlational analyses of spatial patterns provide methods for determining the relative distinctiveness of patterns, where a lower correlation reflects greater distinctiveness. This method is sensitive primarily to changes in variations across space (i.e., patterns), and are insensitive to widespread, spatially uniform "activation" differences across conditions or, conversely, changes that are only highly localized and do not change the relationships between voxels. Regressors that modeled the stimulus presentation period were included in a general linear model, and beta values for the contrast of image presentation periods minus inter-trial fixation periods were calculated for each voxel in the respective anatomical mask resulting in a multi-voxel matrix of post-stimulus activation. These matrices were translated into a linear vector (Aguirre, 2007) for each attention condition, and correlations between each pair of conditions were tested using a non-parametric test [Kendall rank order correlation (tau)]. The statistical significance of the differences between resulting τ (tau) was assessed using dependent sample *T*-test, with an a priori threshold of $p < 0.05$ for significance, Bonferroni corrected for three comparisons yielding a threshold of $p = 0.016$.

RESULTS

BEHAVIORAL RESULTS

When participants were asked to rate familiarity of images in an unexpected recognition task after the scanning session, familiarity was significantly higher for attended images in the selective conditions (Attend Faces; Attend Scenes) than for non-attended images ($p < 0.0001$). Familiarity for images from the Attend Both condition was intermediate, less than for attended ($p < 0.05$) and greater than for non-attended images ($p < 0.001$) in the selective conditions.

IMAGING RESULTS

Extrastriate cortex

Confirmation of information encoding. Classification of extrastriate cortex activity patterns as corresponding to one of the attentional conditions was accurate significantly above chance (overall accuracy 57%, $p < 0.01$ compared to 33% for scrambled). This confirmed that information regarding the attention condition (target of attention) is encoded in the interrogated activity patterns.

Determination of distinctiveness between activity patterns with a pattern classifier analysis. Testing our main hypotheses required determining the relative distinctiveness of activity patterns between attention conditions. We calculated rates for which activity patterns from each given attention condition was classified as representing each of the other conditions.

Lower rates of mis-classification as a particular condition reflect that the particular pair of patterns is more distinctive (less likely to be confused). We found that the classifier was more likely to mis-classify samples from the Attend Scenes condition as being from the Attend Both than the Attend Faces condition ($\chi^2(1, N = 20) = 46.2, p < 0.001$, **Figure 4A**). Likewise, the classifier was more likely to mis-classify samples from the Attend Faces condition as being from Attend Both than the Attend Scenes condition ($\chi^2(1, N = 20) = 21.73, p < 0.001$, **Figure 4A**). In other words, patterns from the opposing selective attention conditions (e.g., attend face or attend scene) were more distinctive from each other than from the non-selective condition (e.g., attend both).

Control for "global" activation differences between attention conditions. To exclude the possibility that the activity patterns for different attention conditions simply differed because of global activation magnitude differences between the conditions, we normalized regional mean signal intensities across attention conditions and repeated the pattern classifier analyses. We found that the pattern of results did not change. The classifier was more likely to mis-classify samples from the Attend Scenes condition as being from the Attend Both than the Attend Faces condition ($\chi^2(1, N = 20) = 88.36, p < 0.001$). Similarly, the classifier was more likely to mis-classify samples from the Attend Faces condition as being from the Attend Both than the Attend Scenes condition ($\chi^2(1, N = 20) = 27.81, p < 0.001$). This suggests that the findings reflect changes in multi-voxel, spatially varying activity patterns and not generalized differences in activation levels between attention conditions.

Determination of distinctiveness between activity patterns with a spatial correlation analysis. Spatial correlations were calculated as an alternative method for estimating activity pattern distinctiveness between attention conditions. Consistent with the pattern classification findings, spatial correlations between the two selective attention conditions (Attend Faces-to-Attend Scenes $\tau = 0.77$) were lower ($p < 0.01$) than the correlations between each selective attention condition and the non-selective attention condition ($\tau = 0.79$ for Attend Faces-to-Attend Both and Scenes-to-Attend Both).

Exclusion of category-selective extrastriate cortex regions. In order to determine the contributions of regions corresponding to the "FFA" and "PPA" to the primary findings, we excluded these regions (as described in "Methods") from the pattern classification analyses. This exclusion resulted in only a small decrement in the classification of attention condition (0.7%), suggesting that top-down attentional signals act on spatially distributed representations beyond these localized category-specific regions.

We also divided voxels from the whole extrastriate cortex mask into 10 percentile strata based on differential univariate responses to viewing of faces or scenes during the performance of

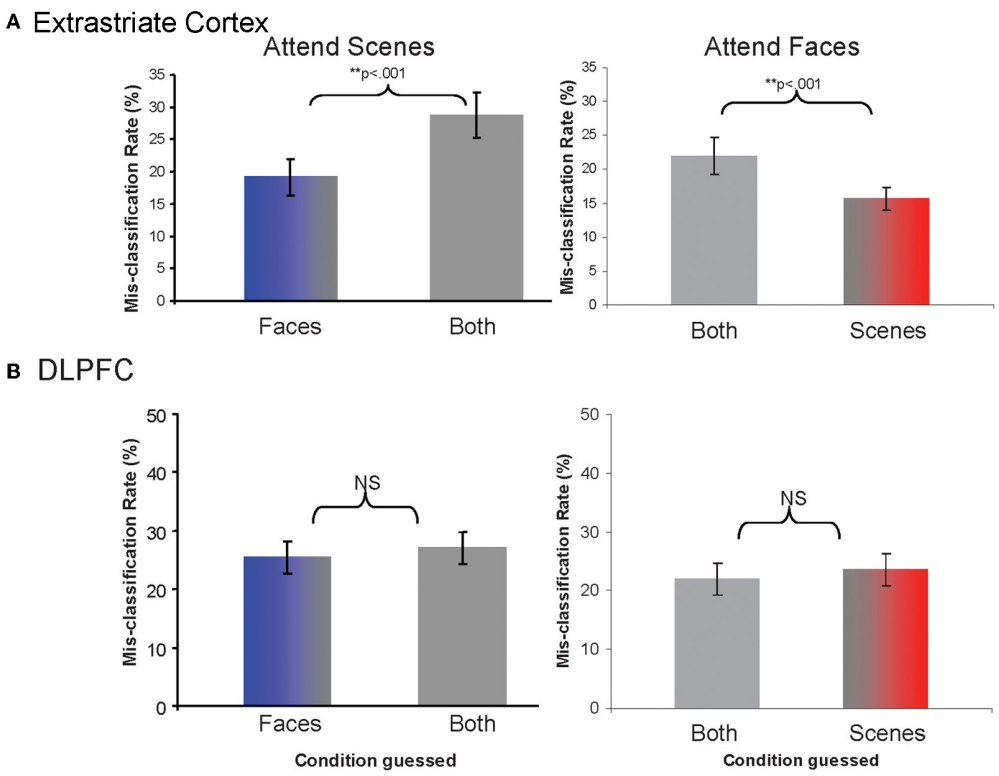

FIGURE 4 | Rates of mis-classification as each of the alternative conditions when the pattern samples are derived from Attend Scenes or Attend Faces trials, for (A) extrastriate cortex and (B) DLPFC.

an independent face and scene "localizer" task (see "Methods"). Next, we performed pattern classification with exclusion of single strata to determine the sensitivity of the pattern classification to the information encoded in each stratum. We found that classification rates remained significantly above chance regardless of which stratum was excluded (range 54–57%). A slightly larger decrement in classification accuracy was observed when the top most selective stratum was excluded. There was no linear relationship between category-selectivity and the classification rates. (See **Figure 5**).

Next, we determined pattern classification accuracy when using data limited to each specific stratum (excluding data from the other nine strata). Classification accuracy was significantly above chance in all strata, including the strata with the lowest category-selectivity. The range of classification accuracy rates was 47–54%. The strata with the highest category selectivity showed a small but significantly higher classification accuracy (54%, compared to 49% for the next stratum, $p < 0.05$, with no significant differences in the remaining strata).

Lateral prefrontal cortex

Confirmation of information encoding. Classification of PFC activity patterns as corresponding to one of the attentional conditions was accurate significantly above chance in all PFC ROIs (dorsolateral PFC—overall accuracy 51% for both left and right, $p < 0.01$ compared to 33% for scrambled; ventrolateral PFC—48% and 49%, left and right, respectively, each $p < 0.01$ compared to 33% for scrambled; frontal polar ROI, 49% and 46%, each $p < 0.01$ compared to 33% for scrambled) This confirmed that information regarding the attention condition (target of attention) is encoded in the interrogated activity patterns.

Determination of distinctiveness between activity patterns with a pattern classifier and spatial correlation analysis. The relative distinctiveness between activity patterns was determined as described above. When samples from each selective attention condition were entered as test patterns, mis-classification as the other two conditions did not differ significantly for either condition in any lateral PFC ROI. Results from the DLPFC ROI are presented in **Figure 4B**. Thus, although the PFC activity patterns from the three attention conditions were distinguishable from the scrambled image condition, there was no difference in the relative distinctiveness of the selective and non-selective attention conditions. Control analyses excluding global activation magnitude differences between attention conditions did not change these findings. The spatial correlation analyses also did not reveal significant differences in correlations for any of the attention condition pairs.

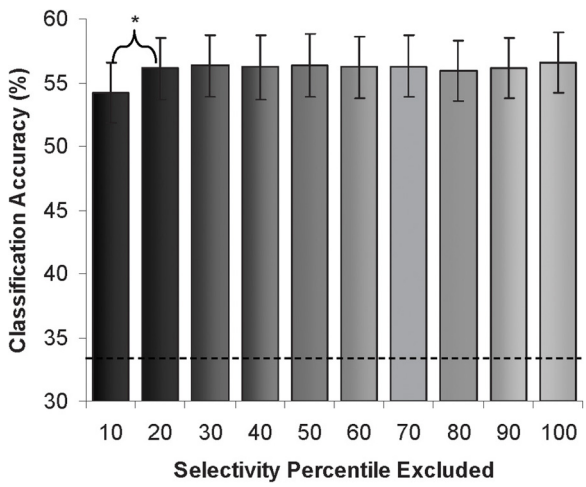

FIGURE 5 | Attention condition classification rates (% accuracy) after exclusion of each "object-selectivity" stratum. (Dotted line demarcates chance rate of classification).

DISCUSSION

Mechanisms for selective processing of visual images based on the relevance of their information content are crucial for goal-directed behavior. There is an encoding advantage for selectively attended stimuli in the context of competition, as evident in the selectivity of subsequent learning and memory. In order to better understand the neural mechanisms of selective information processing, we investigated the modification of visual networks to favor the representation of goal-relevant information during working memory, and putative mechanisms of goal-based guidance of the selection process involving lateral PFC.

MECHANISMS UNDERLYING TUNING OF CATEGORY-SPECIFIC REPRESENTATIONS IN EXTRASTRIATE CORTEX

How are neural representations of visual object information modified by selective attention? Previously proposed mechanisms for object-based attention have been based on the assumption of a modular architecture of information processing. For example, magnitudes of stimulus-driven activation in specific extrastriate cortex regions that respond differentially to object categories of perceived stimuli have been shown to be modulated by attention (O'Craven et al., 1999; Gazzaley et al., 2005a,b). However, extending this finding to a more generalizable model of object-selective attention is problematic. Any model of attention founded solely on localized top-down signaling would quickly reach a roadblock in that a separate module (i.e., a separate category-selective region) would be needed for every possible target of attention. Such a model could not accommodate the diversity and complexity of possible targets of attention. A richer code, such as might be accommodated by a distributed coding structure, would logically be necessary to represent information to specify a diversity of possible targets.

We examined to what extent attention modifies the information represented in spatially distributed neural codes, and in what ways these codes may be modified. A distributed code structure could be measurable as a combinatorial of local activities across space, describable as "patterns." This should be applicable not only at the microscopic scale (e.g., combinations of neurons), but also a macroscopic scale (e.g., across areas of cortex, as measured with fMRI). At this level, the effects of attention may be conceptualized and operationally measured as a reconfiguration of the codes (multi-voxel patterns) to favor representation of goal-relevant information. First, we tested the most basic concept that distributed neural codes are altered (tuned) based on selective

attention to different targets during working memory. We tested the effect of selective attention on information coded in extrastriate activity patterns from individuals engaged in tasks with differential information processing demands based on the relevance of an object category. Three attention conditions demanded either non-selective processing of presented stimuli, or selective processing of one category of information over the other.

At the level of these distributed codes, mechanisms of attention need to act to select one representation over other spatially overlapping representations. Therefore, we predicted that shifting selective attention to one vs. another competing visual object should tune activity patterns in extrastriate cortex in different directions. This prediction was based on a conceptual model in which neural representations of object category information are coded in distributed extrastriate cortex patterns. Basic stimulus-driven perception of two images of different object categories would be encoded in distributed, overlapping representations in extrastriate networks. The overall measurable pattern of visual cortical activity would be a composite of the representations for both categories. With engagement of more in-depth processing, such as for any task that requires the image information to be maintained during working memory, these codes would then be tuned by the top-down direction of attention to favor representation of goal-relevant information. In this experimental protocol, this tuning would make patterns for the opposing selective attention conditions (Attend Scenes vs. Attend Faces) more distinctive from each other (see **Figure 1**). If representations of both categories were equally relevant for the task at hand, both would be represented without a top-down bias in neural activity patterns during working memory and subsequent processing, resulting in an intermediate pattern. We found that brain activity patterns during the contrasting task conditions fit these predictions. These findings support the contention that object-based selective attention is indeed implemented as a shift in the tuning of information represented within distributed neural codes. We then proceeded to address follow-up questions regarding the main finding of this tuning effect.

If any given condition was associated with a generally higher level of activation, due to non-specific factors such as "mental effort," this would provide a non-specific source of information to the pattern classifier. Therefore, we excluded the possibility that differences in activity patterns between conditions could be driven by simple generalized (spatially non-varying) signal intensity differences. The core pattern of findings did not change when global signal intensity differences between conditions were excluded. That is, even when mean signal intensities were equivalent across conditions, the main findings held. This supports our secondary conclusion that the neural representations are coded in spatially-varying, multi-voxel patterns. Having established this, we could then systematically examine potential contributions to the spatially-varying codes, such as the object category selectivity of voxels (discussed below).

Furthermore, the findings generalized across methods of pattern analysis. We performed a separate corroborative analysis using a spatial correlations technique to determine the distinctiveness of the patterns. We calculated the strength of the spatial correlations of multi-voxel patterns representing each attention condition. This method is sensitive primarily to changes in the relationships between voxels in the overall pattern. That is, this method is insensitive to differences that are only widespread (spatially invariant) or localized but that do not change the relationships between voxels in the multi-voxel patterns. For example, if only the most highly activated voxels by a visual task are simply more activated, then the overall spatial correlation would not actually be altered (noting that correlations are magnitude-independent). In other words, only changes that affect the overall *combination* that comprises the topographical patterns would affect the spatial correlation values. Consistent with the pattern classifier findings, the spatial correlation data support the conclusion that the top-down attention signals modulate spatial distributed patterns of visual cortex in which information is coded.

Could the results be specific to the pattern classification method we utilized? We utilized a non-linear classifier in order to avoid assuming that the distinguishing features in the multi-voxel patterns would be linear, but it is possible that the results might not generalize to different types of classifiers. At a general level, it has been argued that the same conclusions would be drawn with different classifier configurations or methods (O'Toole et al., 2007; Misaki et al., 2010), but this presumption should be tested for any given question. Importantly, our results did generalize across methods of pattern analysis (spatial correlations as well as pattern classification with a non-linear classifier), arguing against a classifier-specific result.

DOES OBJECT-BASED ATTENTION ACT ON SPATIALLY-DISTRIBUTED CATEGORY-SPECIFIC CODES IN EXTRASTRIATE CORTEX?

We propose that attention codes are spatially distributed. One may ask whether there is any organizational structure to these codes, and if so, what the organizing principle might be. As mentioned, there is no intuitively simple topographical mapping analogous to that for spatial attention, translating between spatial location in the environment and topography in the brain. One possibility could be based on a hierarchy of object category selectivity of various brain areas, a concept that has provided an underpinning for localization based approaches to object category perception (e.g., Epstein and Kanwisher, 1998). One possibility is that top-down signals (as specified by the attention conditions) could act solely through category-selective nodes. Another possibility is that top-down signals act on spatially-distributed, category-specific codes. In this case, although localized attentional effects would still be observed, they do not reveal the true underlying mechanism underlying object-based attention. To reconcile prior findings of modulation in localized category-selective regions with our proposal of a spatially distributed mechanism of attention, we systematically examined information coded in segments of extrastriate cortex, organized by the "object-selectivity" of the stimulus-driven activation responses in each voxel. We first investigated whether the modulatory effect of attention extends to distributed information codes outside of "object-selective" regions of extrastriate cortex. To determine the contributions of face-selective (FFA) and scene-selective (PPA) regions, we excluded these regions from the classification procedures, and found only a small decrement in the classification of attention

condition. This finding suggests that object-based attention does not act solely on localized, category-specific regions but rather acts on spatially-distributed category-specific codes.

Next, we examined to what extent there might be a relationship between a topography of "category-selectivity" (based on univariate response profiles) and the distributed attentional effects. We divided extrastriate cortex voxels into "strata" of voxels based on differential responses to viewing of faces or scenes in an independent task. We successively excluded each stratum from classification of attention condition in order to quantify how much information was contributed by each stratum. Classification rates remained significantly above chance regardless of which stratum was excluded. In other words, no particular stratum was *necessary* for the attentional effects, which were distributed across extrastriate cortex. In a complementary approach, we determined pattern classification accuracy when using data limited to each specific stratum. All strata, even those showing no selectivity for object-category (the lowest 10% strata in the whole extrastriate cortex mask), represented information coding the attention condition. This finding again supports the contention that object-based attention acts on spatially distributed category-specific codes, not detected by univariate analyses of local regional response changes.

Other sources of modulation, such as motivation or affect, may also be mediated by similar mechanisms of tuning of information codes. The proposed analytic methods could be combined with conditions varying other sources of modulation, such as affect or motivation, to test hypotheses regarding modulatory input from sources involved in other forms of control.

THE NATURE OF OBJECT-BASED ATTENTIONAL SIGNALS IN LATERAL PREFRONTAL CORTEX

If tuning of category-specific codes in extrastriate cortex is mediated by top-down signals emanating from PFC, then patterns of PFC activity should encode measurable information regarding the different attention conditions of our task. We found this to be the case in that information distinguishing the three attention conditions was represented in PFC, as evident by the successful classification of PFC activity patterns for the different conditions. However, we sought to also determine the nature of the representations encoded in lateral PFC supporting top-down attentional signals. For example, does lateral PFC code target-specific information (in this case, specific to object category) to guide attention, a more abstract representation of the goal-relevance of perceptual events, or a representation of a more general cognitive state, completely abstract from the specific external targets of attention?

If PFC codes representations for object categories (concrete, stimulus-based properties of the visual images), then findings for PFC activity patterns should re-capitulate those of extrastriate cortex. This was not the case, since we found that the changes measured in PFC regions differed from those in extrastriate cortex (see **Figure 4**). There was no evidence that the opposing selective attention demands tuned PFC patterns in opposite directions, in contrast with the findings in extrastriate cortex. Rather, classification rates were equivalent for the selective and non-selective attention conditions (see **Figure 4B**). At the other extreme, if PFC codes represent something as abstract as a general cognitive state (with no information specifying the direction of selective attention), then the brain activity patterns in the different selective attention conditions should not be distinguishable at all. Again, this was not the case since we found that information distinguishing the attention conditions was represented in PFC. Our findings do suggest that lateral PFC codes representations of information that are more abstract than the concrete stimulus property of object category, but less abstract than some form of general cognitive state. This finding is consistent with the possibility that PFC codes represent the goal-relevance of perceptual events which could provide guidance for the selective processing of visual information. This is consistent with the conception that PFC is part of a system that adapts to represent currently relevant information (Hampshire et al., 2007).

CONCLUSIONS

This study examined mechanisms of selective attention of representations of object categories that are coded in extrastriate cortex in widely distributed and spatially overlapping representations. Our findings support the hypothesis that extrastriate cortex networks are tuned by attention to favor representation of relevant targets, at the *population level* (at the macroscopic scale measured with fMRI). Where information is coded in combinatorials within the same finite set of nodes (voxels or otherwise), attention may re-configure these combinatorials to tune the codes to better represent goal-relevant information. This provides a mechanism by which top-down signals may bias the competition that occurs when overlapping codes compete to represent different information within finite networks. This study adds to other studies that have examined mechanisms of attention at the level of distributed codes. For example, attention to lower-level visual features has been shown to propagate outside of attended spatial locations (Serences and Boynton, 2007), consistent with attention signals being feature-specific but broadly spatially distributed (Serences et al., 2009). Mental imagery likely relies on attentional top-down signals and also appears to be mediated through spatially distributed codes (Stokes et al., 2009). Electrophysiologic methods will be necessary to disentangle different types of attentional mechanisms (e.g., gain vs. tuning) at the neuronal level. For example David et al. showed that feature-based attention does alter neuronal tuning in V4, although spatial attention only alters gain (David et al., 2008). Object-based attention is likely to share mechanisms with other forms of feature-based attention, though object processing requires the integration of multiple features. Our study provides support for the logic that attention would need to tune representations in complex codes to favor goal-relevant object information.

ACKNOWLEDGMENTS

We appreciate funding support from the VA Rehabilitation Research and Development, UCSF Division of Geriatrics, and NIH grants MH63901 and NS40813. We appreciate the assistance of Michael Souza and Drew Fegen with data analyses, and Jeff Cooney, Joshua Hoffman, and Adam Gazzaley for task stimuli and advice, and Andrew Kayser, MD/PhD, for programming.

REFERENCES

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. *Neuroimage* 35, 1480–1494.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* 21, 373–383.

Beck, D. M., and Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res.* 49, 1154–1165.

David, S. V., Hayden, B. Y., Mazer, J. A., and Gallant, J. L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59, 509–521.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1245–1255.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., and Kanwisher, N. (2006). Domain specificity in visual cortex. *Cereb. Cortex* 16, 1453–1461.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601.

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., and D'Esposito, M. (2005a). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J. Cogn. Neurosci.* 17, 507–517.

Gazzaley, A., Cooney, J. W., Rissman, J., and D'Esposito, M. (2005b). Top-down suppression deficit underlies working memory impairment in normal aging. *Nat. Neurosci.* 8, 1298–1300.

Hampshire, A., Duncan, J., and Owen, A. M. (2007). Selective tuning of the blood oxygenation level-dependent response during simple target detection dissociates human frontoparietal subregions. *J. Neurosci.* 27, 6219–6223.

Hanson, S. J., Matsuka, T., and Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001). revisited: is there a "face" area? *Neuroimage* 23, 156–166.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.

Jazayeri, M. (2008). Probabilistic sensory recoding. *Curr. Opin. Neurobiol.* 18, 431–437.

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

Miller, B. T., and D'Esposito, M. (2005). Searching for "the top" in the top-down control. *Neuron* 48, 535–538.

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.

Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118.

Murray, S. O., and Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat. Neurosci.* 7, 70–74.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430.

O'Craven, K. M., Downing, P. E., and Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature* 401, 584–587.

O'Toole, A. J., Jiang, F., Abdi, H., and Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17, 580–590.

O'Toole, A. J., Jiang, F., Abdi, H., Penard, N., Dunlop, J. P., and Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J. Cogn. Neurosci.* 19, 1735–1752.

Serences, J. T., and Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55, 301–312.

Serences, J. T., Saproo, S., Scolari, M., Ho, T., and Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage* 44, 223–231.

Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J. Neurosci.* 29, 1565–1572.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 December 2011; accepted: 04 June 2012; published online: 21 June 2012.

Citation: Chen AJ-W, Britton M, Turner GR, Vytlacil J, Thompson TW and D'Esposito M (2012) Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. Front. Hum. Neurosci. ***6****:187. doi: 10.3389/fnhum.2012.00187*