nature neuroscience

Attention during natural vision warps semantic representation across the human brain

Tolga Çukur[1], Shinji Nishimoto[1,4], Alexander G Huth[1] & Jack L Gallant[1–3]

Little is known about how attention changes the cortical representation of sensory information in humans. On the basis of neurophysiological evidence, we hypothesized that attention causes tuning changes to expand the representation of attended stimuli at the cost of unattended stimuli. To investigate this issue, we used functional magnetic resonance imaging to measure how semantic representation changed during visual search for different object categories in natural movies. We found that many voxels across occipito-temporal and fronto-parietal cortex shifted their tuning toward the attended category. These tuning shifts expanded the representation of the attended category and of semantically related, but unattended, categories, and compressed the representation of categories that were semantically dissimilar to the target. Attentional warping of semantic representation occurred even when the attended category was not present in the movie; thus, the effect was not a target-detection artifact. These results suggest that attention dynamically alters visual representation to optimize processing of behaviorally relevant objects during natural vision.

Attention is thought to increase information processing efficiency throughout the brain through several convergent mechanisms[1]. Neurophysiology studies in early visual areas have shown that spatial attention changes response baseline, response gain and contrast gain[2–4]. However, because the brain pools information across successive stages of processing, attentional modulation of baseline and gain at early stages likely causes changes in neuronal tuning in higher sensory and cognitive brain areas[3,5]. Indeed, feature-based attention can cause modest changes in tuning of single neurons even as early as V4 (refs. 5,6), but tuning changes of single neurons in prefrontal cortex can be substantial[7–9]. Tuning shifts in single neurons change the way that information is represented across the neural population, warping the representation to favor certain signals at the expense of others[5]. Thus, it has been proposed that tuning shifts reflect the operation of a matched-filter mechanism that optimizes task performance by expanding the cortical representation of attended targets[5,10].

Attentional warping of cortical representation might be particularly valuable during demanding tasks such as natural visual search. Recent evidence suggests that the brain represents thousands of object categories by organizing them into a continuous semantic similarity space (**Fig. 1a**) that is mapped systematically across visual cortex[11]. Because natural scenes are cluttered with many different objects, they may elicit patterns of brain activity that are widely distributed across this semantic space, making target detection difficult. Attention could markedly increase sensitivity for the target and improve target detection under these demanding conditions[5] by expanding the cortical representation of behaviorally relevant categories and compressing the representation of irrelevant categories (**Fig. 1b**,**c**).

It is currently unknown whether attention warps the cortical representation of sensory information in the human brain. To search for evidence for this complex attentional effect, we exploited the fact that attention would expand the representation of an attended category by causing neural populations throughout visual and nonvisual cortex to shift tuning toward the target[5–9] (**Supplementary Fig. 1**). We hypothesized that visual search for a single object category should cause tuning shifts in single voxels measured by functional magnetic resonance imaging (fMRI; **Fig. 1d**–**f**).

To identify semantic tuning shifts, we measured category tuning in single voxels during a natural category–based visual search task (**Fig. 2**). We recorded whole-brain fMRI data from five human subjects while they viewed 60 min of natural movies (Online Methods). Subjects maintained steady fixation while covertly searching for ‘humans’ or ‘vehicles’. These categories were used because they are quite distinct from one another, they occur commonly in real-world scenes and they are common targets of visual search[12,13].

Category-based attention tasks have been used in several previous fMRI experiments[12,14,15]. However, these earlier studies used a small set of object categories and region-based data analysis procedures. Thus, they did not explore voxel-based tuning and could not distinguish voxel-based changes in tuning from changes in response baseline or gain. To maximize our ability to detect tuning changes in single voxels, we used complex natural movie stimuli containing hundreds of different object and action categories[16,17]. To remove attentional effects on response baseline and gain, we normalized the blood oxygen level–dependent (BOLD) responses of each voxel to have zero mean and unit variance individually in each attention condition before further modeling. This procedure allowed us to clearly separate tuning changes from simple modulation of response baseline or gain.

We then employed a previously developed voxel-wise modeling approach to obtain accurate estimates of category tuning in single

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA. [2]Program in Bioengineering, University of California, Berkeley, California, USA. [3]Department of Psychology, University of California, Berkeley, California, USA. [4]Present address: Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita, Osaka, Japan. Correspondence should be addressed to J.L.G. (gallant@berkeley.edu).

Received 17 January; accepted 18 March; published online 21 April 2013; doi:10.1038/nn.3381

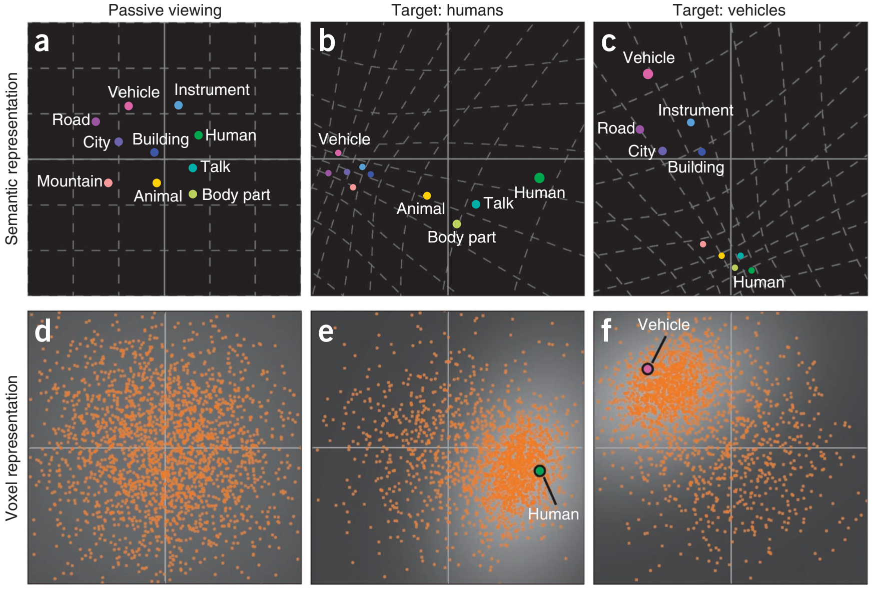

Figure 1 Tuning-shift hypothesis predicts that attention warps semantic representation. (**a**–**c**) Hypothesized changes in semantic representation. Previous studies have suggested that the brain represents categories by organizing them into a continuous space according to semantic similarity. (**a**) During passive viewing, semantically similar categories project to nearby points in the semantic space. (**b**,**c**) The tuning-shift hypothesis predicts that attention to one specific category expands the representation of both the attended and nearby categories in the semantic space and compresses the representation of distant categories. (**d**–**f**) Attentional warping of semantic representation implies corresponding changes in voxel-wise semantic tuning. (**d**) During passive viewing cortical voxels (orange dots) are tuned for different categories, and can also be visualized in the semantic space as in **a**. (**e**,**f**) During visual search, many voxels should shift their tuning toward the attended category to expand representation of the corresponding part of semantic space. This causes fewer voxels to be tuned for distant categories.

cortical voxels and in each individual subject[11,18–21]. The WordNet lexicon[22] was used to label 935 object and action categories in the movies (**Supplementary Fig. 2**). Regularized linear regression was used to fit voxel-wise models that optimally predicted the measured BOLD responses from the categorical indicator variables (**Supplementary Fig. 3**). We estimated separate models using data acquired during visual search for humans and for vehicles. The resulting model weights give the category tuning vectors for each voxel under each attention condition.

RESULTS

Attentional changes in semantic representation can be inferred by comparing category tuning vectors across attention conditions (**Fig. 3**). However, inferences drawn from this comparison will only be justified and functionally important if the fit category models can successfully predict BOLD responses to novel natural stimuli. To address this issue, we validated the prediction performance of category models on separate data reserved for this purpose. Prediction scores were defined as the Pearson's correlation between the BOLD responses measured in the validation data set and those predicted by the fit models (Online Methods). All statistical significance levels were corrected for multiple comparisons using false discovery rate (FDR) control[23].

We found that category models provided accurate response predictions across many regions of visual and nonvisual cortex (**Supplementary Fig. 4**). Overall, 83.7 ± 5.12% (mean ± s.d. across subjects) of cortical voxels were significantly predicted by the category model (*t* test, $P < 0.05$). The category model explained more than 20% of the response variance in 11.60 ± 5.84% (mean ± s.d.) of these voxels across subjects. These results suggest that category tuning vectors accurately reflect category responses of many cortical voxels during visual search.

If attentional tuning changes are statistically significant, then category models for individual attention conditions should yield better response predictions than a null model fit by pooling data across conditions. To assess significance, we therefore compared the prediction scores obtained from category models to those obtained using null models. We found that 59.57 ± 8.31% (mean ± s.d. across subjects) of cortical voxels exhibited significant tuning changes (*t* test, $P < 0.05$). Across subjects, 17.13 ± 0.97% (mean ± s.d.) of these voxels also had high prediction scores (greater than 1 s.d. above the mean), yielding 4,245–7,785 well-modeled voxels in individual subjects. Control analyses revealed that these tuning changes could not be attributed to nuisance factors, including eye movements, head motion, physiological noise and spatial attention (Online Methods). Furthermore, because all responses were *z* scored individually in each attention condition, these results cannot be explained by additive or multiplicative modulations of responses in single voxels. Thus, they suggest that category-based attention causes substantial tuning shifts in many cortical voxels.

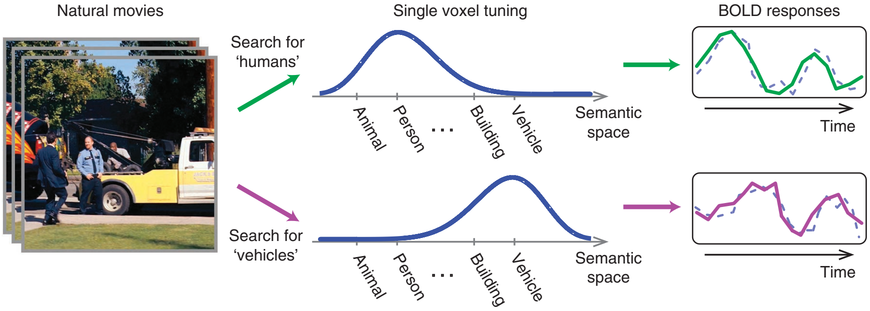

Figure 2 Voxel-wise tuning vectors are measured from BOLD responses evoked by natural movies. Tuning changes in single voxels are a unique, diagnostic aspect of the tuning-shift hypothesis. To test this hypothesis, we measured changes in voxel tuning during covert visual search for either humans or vehicles in complex natural movies. A separate category model was fit to each voxel in each attention condition to optimally predict evoked BOLD responses (dashed lines indicate predicted response, solid lines indicate measured response). The category model gives voxel tuning under each condition, and tuning shifts can be identified by comparing tuning across conditions.

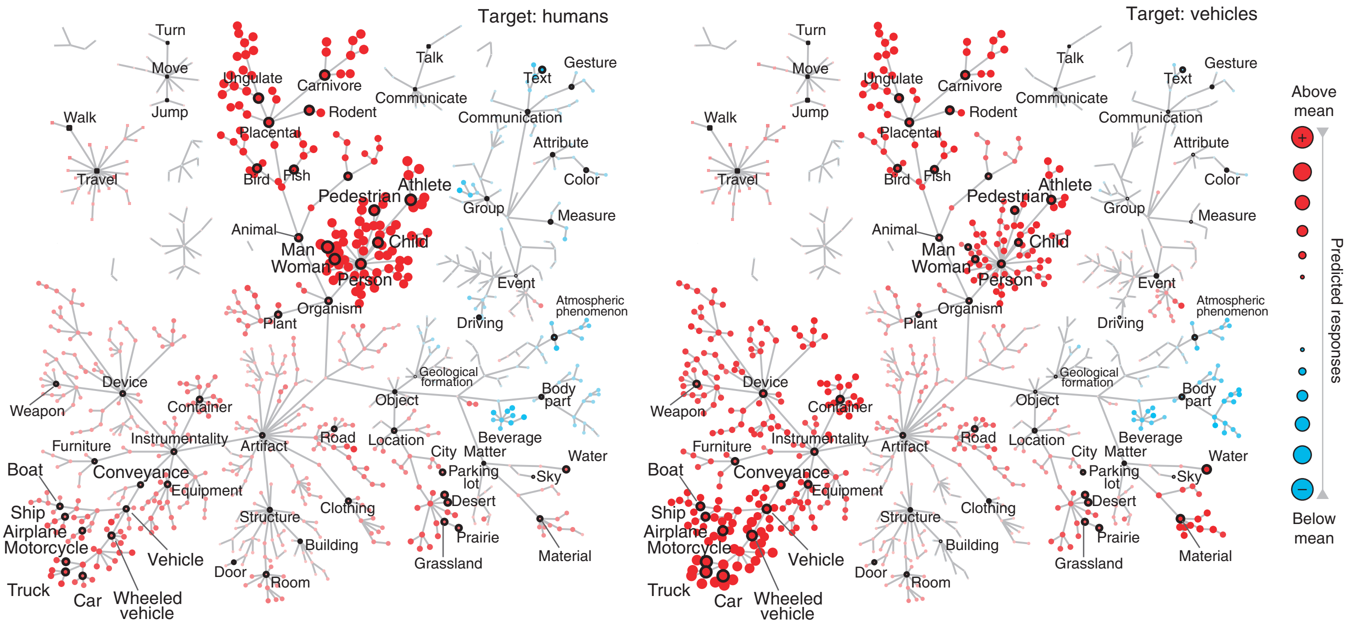

Figure 3 Attentional tuning changes for a single voxel in lateral occipital complex. Tuning for 935 object and action categories in a single voxel selected from lateral occipital complex (LO) in subject S1, during search for humans (left) and for vehicles (right). Each node in these graphs represents a distinct object or action, and a subset of the nodes has been labeled to orient the reader. The nodes have been organized using the hierarchical relations found in the WordNet lexicon. Red versus blue nodes correspond to categories that evoked above- and below-mean responses. The size of each node shows the magnitude of the category response. This well-modeled lateral occipital complex voxel (a prediction score of 0.401) exhibited significant tuning changes across attention conditions (*t* test, $P < 10^{-6}$). The voxel was strongly tuned for the attended category in both conditions, and weaker tuning was observed for the unattended categories.

Our experiment used a category-based attention task that required attention to humans or vehicles. However, complex natural movies may contain low-level features that are correlated with these semantic categories. Do the attentional tuning shifts shown here reflect category-based attention or are they a result of attention to correlated low-level features? To address this issue, we fitted simpler structural encoding models that reflect tuning for elementary features, such as spatio-temporal frequency, orientation and eccentricity (Online Methods). We then compared predictions of category models and structural models across well-modeled voxels that showed significant tuning shifts in the category-based attention task.

We found that the average prediction score of structural models was only 0.22 ± 0.03 (mean ± s.d. across subjects), which is significantly lower than that of category models (0.54 ± 0.11, randomized *t* test, $P < 10^{-4}$). We also found that the percentage of response variance explained by structural tuning shifts was only 1.53 ± 0.68% (mean ± s.d.), which is significantly lower than that explained by category-based tuning shifts (13.57 ± 7.65%, Wilcoxon signed-rank test, $P < 10^{-4}$). These findings suggest that tuning for elementary visual features cannot account for the category-based tuning shifts measured here.

Next, we asked whether these changes in category tuning are consistent with the tuning-shift hypothesis[5], in which attention warps semantic representation to favor behaviorally relevant categories at the expense of irrelevant categories. The tuning-shift hypothesis makes three explicit and diagnostic predictions about how attention alters semantic representation. First, it predicts that attention causes tuning shifts toward the attended category when the targets are present, expanding the representation of the attended category. Second, it predicts that attention causes tuning shifts toward the attended category even when no targets are present. Finally, it predicts that attention expands the representation of unattended categories that are semantically similar to the target and compresses the representation of categories that are semantically dissimilar to the target. We tested the tuning-shift hypothesis by evaluating each of these predictions in turn.

Tuning shifts in the presence of targets

To determine whether attention causes tuning shifts toward the attended category when the targets are present, we first projected voxel-wise tuning vectors measured during visual search into a continuous semantic space. The semantic space was derived from principal components analysis of tuning vectors measured during a separate passive-viewing task (Online Methods). Different voxels that are tuned for semantically similar categories will project to nearby points in this space. We then visualized the distribution of tuning across well-modeled voxels that had significant category models (*t* test, $P < 0.05$). We found that most well-modeled voxels were selectively tuned for the attended category, and attention caused tuning shifts in most of these voxels (**Fig. 4a**).

We quantified the magnitude and direction of tuning shifts across attention conditions by measuring the selectivity of voxel tuning for humans or vehicles under each condition (Online Methods and **Supplementary Figs. 5** and **6a–e**). We then computed a tuning shift index (TSI) that summarizes the difference in selectivity for the attended versus unattended category (Online Methods). Under this scheme, a voxel that shifts toward the attended category will have a positive TSI. We found that the mean TSI across well-modeled voxels was significantly greater than 0 in all subjects (Wilcoxon signed-rank test, $P < 10^{-6}$; **Supplementary Fig. 7**). Because all responses were *z* scored individually in each attention condition before TSI values were calculated, these tuning shifts cannot be explained by changes in voxel response baseline or gain (see Discussion). Thus, these results are consistent with the view that attention changes tuning to expand the representation of the attended category.

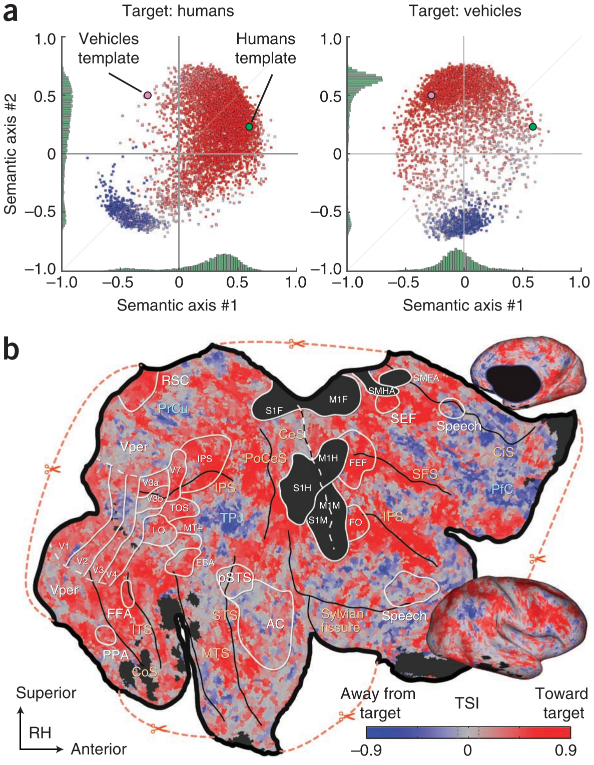

Figure 4 Attention causes tuning shifts in single voxels. (**a**) Semantic tuning of single voxels during two attention conditions: search for humans (left) or vehicles (right). To assess attentional changes, we projected voxel-wise tuning vectors into a continuous semantic space. The semantic space was derived from principal components analysis (PCA) of tuning vectors measured during a separate passive-viewing task. Horizontal and vertical axes correspond to the second and third principal components (the first principal component distinguishes categories with high versus low stimulus energy and so is not shown here). A total of 7,785 well-modeled voxels with significant model weights (*t* test, $P < 0.05$) and high prediction scores (greater than 1 s.d. above the mean) are shown for subject S1. Each voxel is represented with a dot whose color indicates the TSI: red or blue for shifts toward or away from the target, respectively. The positions of the idealized templates for attended categories are shown as colored circles. The marginal distributions are displayed with separate histograms (green). Most well-modeled voxels strongly shifted toward the attended category (Wilcoxon signed-rank test, $P < 0.05$). (**b**) The TSIs for subject S1 are shown on a cortical flat map of the right hemisphere (RH). The color bar represents the 95% central range of TSIs, and voxels with insignificant TSIs appear in gray ($P > 0.05$, between dashed black lines). Regions of fMRI signal dropout and motor areas excluded from all analyses are shown with dark gray patches. The boundaries of cortical areas identified by standard localizers are indicated with solid (functionally inferred) and dashed (anatomically inferred) white lines (**Supplementary Table 1**). Major anatomical landmarks (blue font) and sulci (orange font and black lines) are also labeled (**Supplementary Table 2**). CeS, central sulcus; CiS, cingulate sulcus; CoS, collateral sulcus; IFS, inferior frontal sulcus; IPS, intraparietal sulcus; ITS, inferior temporal sulcus; LO, lateral occipital complex; MTS, middle temporal sulcus; PfC, prefrontal cortex; PoCeS, postcentral sulcus; PrCu, precuneus; SFS, superior frontal sulcus; STS, superior temporal sulcus; TPJ, temporo-parietal junction. Voxels in many brain regions shift their tuning toward the attended category. These include most of ventral-temporal cortex, lateral occipital complex, IPS, IFS, SFS, and dorsal bank of CiS. In contrast, PrCu, TPJ, PfC and areas along the anterior CiS shifted their tuning away from the search target. Data sets are available at http://gallantlab.org/brainviewer/cukuretal2013.

Cortical distribution of tuning shifts

Previous neurophysiology studies have suggested that tuning shifts should be widespread across the brain, extending from higher order visual areas into frontal cortex[5–9]. To visualize the distribution of tuning shifts across cortex, we projected TSI values onto cortical flat maps. We found that voxels in many different brain regions shifted

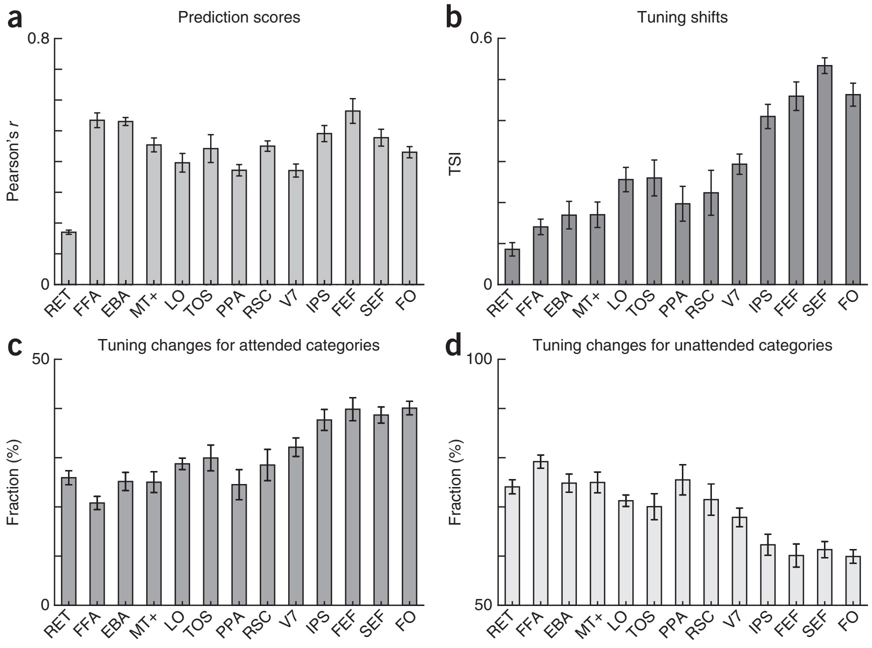

Figure 5 Attention causes different degrees of tuning shifts in functional ROIs. (**a**) Prediction scores (Pearson's *r*, mean ± s.e.m. results averaged across all five subjects). RET, early visual areas V1–3; FFA, fusiform face area; EBA, extrastriate body area; MT+, human MT; LO, lateral occipital complex; TOS, transverse occipital sulcus; PPA, parahippocampal place area; RSC, retrosplenial cortex; FEF, frontal eye fields; SEF, supplementary eye fields; FO, frontal operculum. The average prediction score in category-selective areas in occipito-temporal cortex (FFA, EBA, lateral occipital complex and TOS) was 0.48 ± 0.07 (mean ± s.d.), and the average prediction score in more anterior brain areas in frontal cortex (FEF, SEF and frontal operculum) was 0.49 ± 0.07 (mean ± s.d.). (**b**) Tuning shift indices (mean ± s.e.m.) in functional ROIs. TSIs were significantly greater than 0 in all ROIs (Wilcoxon signed-rank test, $P < 10^{-6}$). Furthermore, TSI increased toward later stages of visual processing. (**c**) Fraction of the overall tuning change (mean ± s.e.m.) explained by tuning changes for attended categories. (**d**) Fraction of the overall tuning change (mean ± s.e.m.) explained by tuning changes for unattended categories (that is, excluding both humans and vehicles). The degree of tuning shift (that is, TSI) was positively correlated with the fraction of variance explained by tuning changes for attended categories ($r = 0.86 \pm 0.02$, *t* test, $P < 10^{-6}$).

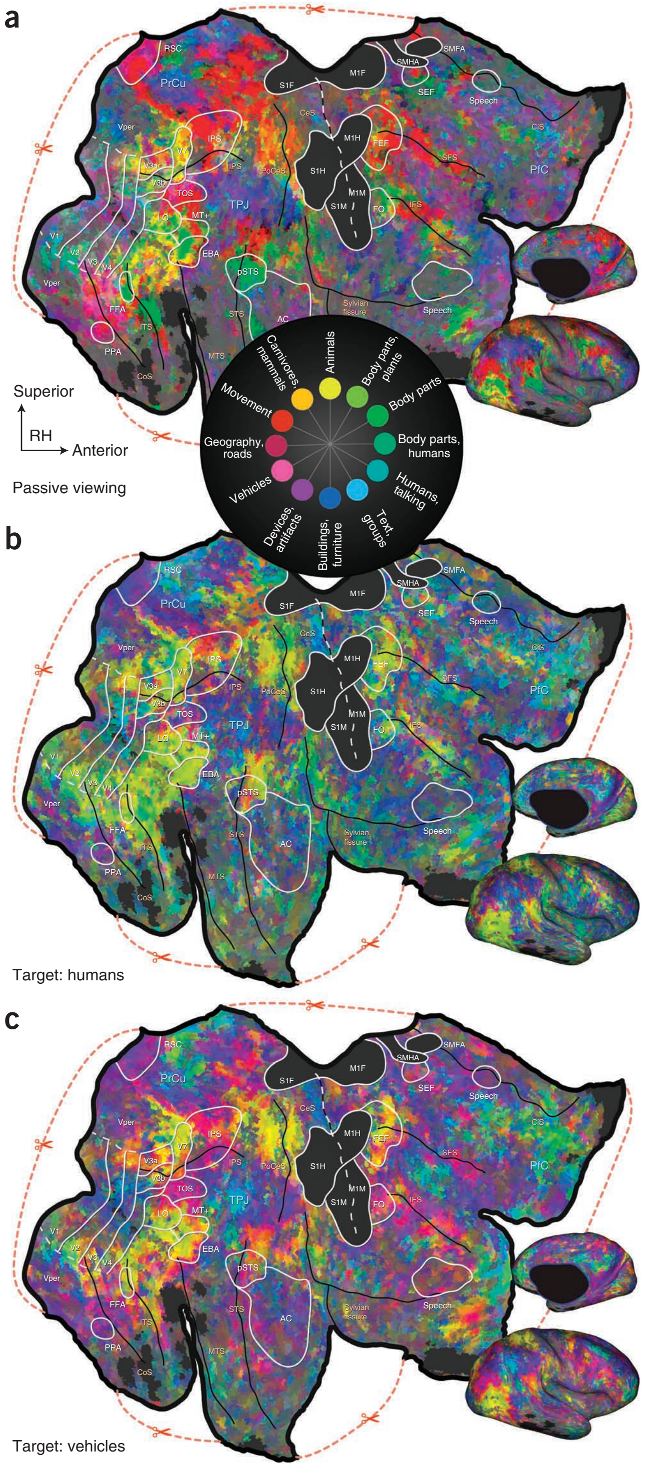

Figure 6 Semantic tuning for unattended categories shifts toward the attended category even when no targets are present. (**a**) Distribution of semantic tuning across the cortex (subject S1, right hemisphere) during passive viewing. Tuning was estimated from responses to all available movie clips. A four-dimensional semantic space was derived from these data using PCA. The tuning vector for each cortical voxel was then projected into this space and the projections onto the second, third and fourth principal components were assigned to the red, green and blue channels. Voxels with similar tuning projected to nearby points in the semantic space and so they are assigned similar colors. Insignificant voxels are shown in gray. Yellow-green voxels were more selectively tuned for animals and body parts and purple-red voxels were more selectively tuned for geographic locations and movement. Anatomical landmarks are labeled as in **Figure 4b**. (**b**) Distribution of semantic tuning for the subject shown in **a**, but during search for humans. Tuning was estimated only from responses evoked by movie clips in which the target did not appear. Data are presented as in **a**. Yellow-green voxels that are tuned for animals and body parts predominated during search for humans. Many voxels in posterior areas that were tuned for vehicles under passive viewing (for example, PPA, RSC and TOS) shifted their tuning away from vehicles, and many voxels that were not tuned for humans under passive viewing (in FEF, frontal operculum, IPS, PfC and insular cortex) shifted their tuning toward humans. (**c**) Distribution of semantic tuning for the subject shown in **a**, but during search for vehicles. Tuning was estimated only from responses evoked by movie clips in which the target did not appear. Data are presented as in **a**. Purple-magenta voxels that were tuned for geographic locations and artifacts predominated during search for vehicles. Many voxels in posterior areas that were tuned for humans under passive viewing (for example, EBA, FFA, TPJ and PrCu) shifted their tuning away from humans, and many voxels that were not tuned for vehicles under passive viewing (in FEF, frontal operculum, IPS, PfC and insular cortex) shifted their tuning toward vehicles. Data sets are available at http://gallantlab.org/brainviewer/cukuretal2013.

their tuning toward the attended category (**Fig. 4b** and **Supplementary Fig. 6a–e**; data sets are available at http://gallantlab.org/brainviewer/cukuretal2013). These include most of ventral-temporal cortex, the lateral-occipital and intraparietal sulci, the inferior and superior frontal sulci, and the dorsal bank of the cingulate sulcus (**Supplementary Figs. 8** and **9**). In contrast with most brain regions, voxels in the precuneus, temporo-parietal junction, anterior prefrontal cortex and anterior cingulate sulcus shifted their tuning away from the attended category. This finding suggests that these brain areas are involved in distractor detection and in error monitoring during visual search[24,25].

To examine how specific brain areas change their representations of attended and unattended categories, we performed detailed analyses of tuning shifts in several common regions of interest (ROIs). We found that regions in higher order visual cortex and more anterior brain areas had high prediction scores, indicating that tuning shifts in these regions are functionally important (**Fig. 5a**). TSI was small in retinotopic early visual areas, but was significantly larger in more anterior brain areas that correspond to later stages of visual processing (Wilcoxon signed-rank test, $P < 10^{-6}$; **Fig. 5b**). This result implies that attentional tuning shifts become progressively stronger toward later stages of processing. We also found that these tuning shifts occurred for both attended (that is, humans and vehicles; **Fig. 5c**) and unattended categories (Wilcoxon signed-rank test, $P < 10^{-6}$; **Fig. 5d**). This finding is consistent with an attentional mechanism that alters the representation of the entire semantic space during visual search (**Supplementary Fig. 1d**). Finally, we found that tuning changes for attended categories accounted for a relatively larger fraction of the overall tuning change in more anterior brain areas compared with earlier visual areas (**Fig. 5c**), whereas those for unattended categories accounted for a relatively smaller fraction of tuning changes (**Fig. 5d**). Taken together, these results suggest that more anterior brain areas are primarily involved in representing the attended category and that visual representations in more frontal areas are relatively more dependent on the search task than those at earlier stages of visual processing[5–9,26].

Tuning shifts in the absence of targets

The second prediction of the tuning-shift hypothesis is that attention causes tuning changes even when no targets are present. To address this issue, we estimated voxel tuning using only those segments of the movies that did not contain humans or vehicles. Note that, as any

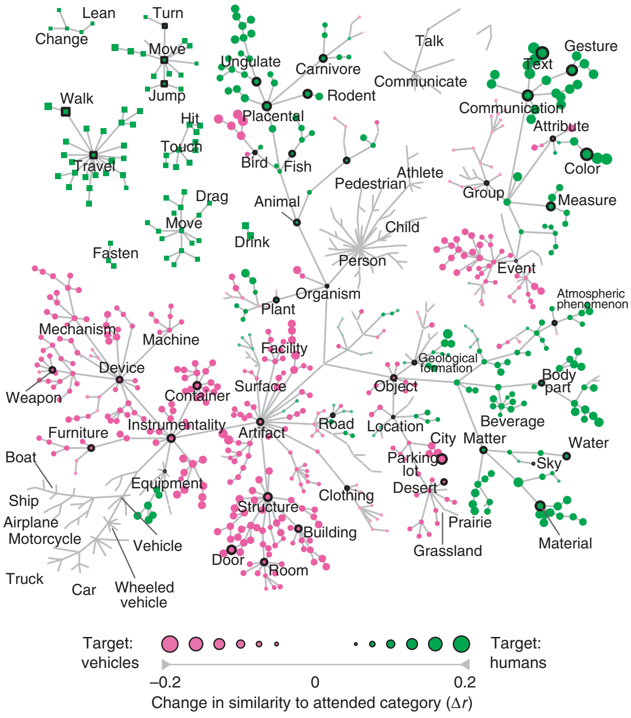

Figure 7 Attention expands the representation of unattended categories that are semantically similar to the attended category. The tuning-shift hypothesis predicts that attention expands the representation of unattended categories that are nearby the attended category in the semantic space. This implies that the representation of unattended categories that are semantically similar to the target will shift toward the representation of the attended category. To address this issue, we measured the similarity of BOLD response patterns evoked by unattended categories to those evoked by the attended category. In each subject, response patterns were estimated across a total of 4,245–7,785 well-modeled voxels that were used in the main analysis. The response patterns for unattended and attended categories were estimated using target-absent and target-present movie segments, respectively. The similarity of response patterns was quantified using Pearson's correlation (*r*) and the results were averaged across subjects. Each node represents a distinct object or action, and some nodes have been labeled to orient the reader. The nodes have been organized using the hierarchical relations found in the WordNet lexicon. The size of each node shows the magnitude of change in similarity (Wilcoxon signed-rank test, $P < 10^{-4}$; see legend at the bottom). During search for humans, representations of semantically similar categories (for example, animals, body parts, action verbs and natural materials) shifted toward the representation of humans (green nodes). During search for vehicles, representations of semantically similar categories (for example, tools, devices and structures) shifted toward the representation of vehicles (magenta nodes).

systematic differences in arousal, respiration and spatial attention across attention conditions are most likely to occur when the targets are present, this analysis also serves as a powerful control against such nuisance factors (Online Methods). Because data recorded when the targets were present were excluded from analysis, tuning for the attended categories could not be assessed directly. However, our modeling framework allowed us to measure tuning shifts for the remaining categories and to infer the direction of shifts with respect to the attended categories from these measurements.

To assess the direction of tuning shifts in the absence of the targets, we projected the tuning vectors estimated in the absence of the targets into the semantic space. We found that voxels in many brain regions shifted their tuning toward the attended category even when no targets were present (**Fig. 6** and **Supplementary Fig. 10a**–**e**). The mean TSI across the population of well-modeled voxels was significantly greater than 0 in all subjects (Wilcoxon signed-rank test, $P < 10^{-6}$; **Supplementary Fig. 11**). These results indicate that attention causes tuning shifts toward the attended category even when no targets are present and that attentional tuning shifts are not a mere consequence of target detection.

Semantic representation of unattended categories

The third prediction of the tuning-shift hypothesis is that attention expands the representation of categories that are semantically similar to the attended category, even when no targets are present. If the representation of an unattended category is expanded, its representation should shift toward the representation of the attended category (that is, the region of the semantic space that many voxels are tuned for). To address this issue, we assessed how the similarity between representations of unattended and attended categories changed across attention conditions. The similarity between representations of two categories was measured using Pearson's correlation between corresponding BOLD response patterns across well-modeled voxels[27]. Responses for unattended and attended categories were estimated using target-absent and target-present movie segments, respectively.

We found that, during search for humans, representations of animals, body parts, action verbs and natural materials shifted toward the representation of humans. In contrast, during search for vehicles, representations of tools, devices and structures shifted toward the representation of vehicles (Wilcoxon signed-rank test, $P < 10^{-4}$; **Fig. 7**). This result suggests that attention expands the representation of unattended categories that are semantically similar to the target at the expense of categories that are semantically dissimilar to the target.

DISCUSSION

Our results indicate that category-based attention during natural vision causes semantic tuning changes that cannot be explained by additive or multiplicative response modulations in single voxels. These tuning changes altered the cortical representation of both attended and unattended categories. Furthermore, attentional changes in tuning for unattended categories occurred even when the attended categories were not present in the movie. These effects are consistent with an attentional mechanism that acts to expand the representation of semantic categories nearby the target in the semantic space at the cost of compressing the representation of distant categories.

Because we measured hemodynamic changes, we cannot make direct inferences about the underlying neural mechanisms mediating tuning shifts. Several possible neural mechanisms might conceivably contribute to semantic tuning changes in single voxels. When the targets are present in the display, then it is possible that changes in response baseline or gain of single neurons that are tuned to the attended targets contribute to tuning changes. However, tuning changes for unattended categories that are observed when no targets are present cannot be explained by this mechanism; because the attended categories were never present in these cases, neurons tuned only to the attended categories never entered into the model estimation procedure and therefore could not have any effect on estimated voxel-wise tuning curves.

Our results are consistent with existing neurophysiology studies that have demonstrated tuning shifts in single neurons as early as area V4 (refs. 5,6), and that have shown far stronger tuning shifts at relatively higher levels of visual and cognitive processing[7–9]. Some of these single neuron studies have reported that tuning shifts are consistent with a matched-filter mechanism that shifts tuning toward the attended target, expanding the representation of attended stimuli at the cost of unattended stimuli. Our results are also consistent with theoretical expectations based on the anatomical structure of the cortical hierarchy; because neurons pool information across successive stages of processing, attentional modulation of baseline or gain at one level must inevitably cause tuning changes at subsequent levels[3,5]. Thus, it is reasonable to expect that changes in voxel tuning at least partly reflect tuning shifts in individual neurons in the underlying neural population.

Although natural movies have strong face validity, correlations inherent in natural movies could potentially complicate interpretation of the results. We took several measures to ensure that stimulus correlations did not confound our results. First, the collection of movies used in the experiments was highly diverse. Second, we used a regression-based modeling approach that minimizes the effect of residual correlations on the fit models. Finally, we performed control analyses on raw BOLD responses to rule out biases resulting from correlations between attended and unattended categories (Online Methods).

Given that our data are finite, there is always some chance that residual correlations may introduce some bias in the results. However, artificial stimuli that contain only a small number of categories introduce much more substantial and pernicious bias, and are therefore more likely to lead to misinterpretation. Interpretation of experiments that use limited stimulus sets inevitably relies on a strong assumption of linearity, that is, that responses to multiple objects in a natural context will be predictable from responses to isolated objects. In contrast, natural stimuli do not require any such linearity assumptions. Note, however, that this important issue is really not relevant to this study. The main goal of this study was not to measure tuning, but rather to measure changes in tuning between different search tasks. Because natural stimuli have high ecological relevance for natural visual search, natural movies appear to be better suited for these measurements.

An important question to be answered is the role of bottom-up processing versus top-down feedback in measured tuning changes. Because we used the same movie stimulus for the two separate search tasks in our experiment, all attentional tuning changes between the two tasks must necessarily reflect top-down modulatory effects. We found small tuning shifts in retinotopic early visual areas and significantly larger tuning shifts in higher visual areas in occipitotemporal cortex and relatively more anterior brain areas. We also found that tuning shifts could not be explained by response modulations for lower-level visual features that are known to be represented in early visual areas. These results imply that attentional modulations primarily warp semantic representation at later stages of visual processing. However, the slow nature of BOLD responses makes it difficult for any fMRI study to measure the temporal relationship between signals arising in different brain areas at these later stages of processing.

The way that attention optimizes target detection depends not only on the target, but also on the similarity between the target and the distractors[28]. If the target is very different from the distractors, then target detection can be optimized by shifting tuning toward the target[5]. However, if the target is very similar to the distractors, target detection can be improved by enhancing the representation of task-irrelevant features that optimally distinguish the target from the distractors[29]. Here, the attentional targets were highly distinct (humans and vehicles), so it is natural to expect that tuning should shift toward the target. An important topic for future research will be to determine whether attention causes tuning shifts toward task-irrelevant features when the target and distractors are very similar.

In conclusion, we found that natural visual search for a single category warps the entire semantic space, expanding the representation of nearby semantic categories at the cost of more distant categories. This effect suggests a more dynamic view of attention than is assumed under the conventional view that attention is a simple mechanism that merely modulates the baseline or gain of labeled lines. This dynamic mechanism can improve the effective resolution of the visual system for natural visual search, and it likely enables the use of limited neural resources to perform efficient search for many different object categories. Overall, these findings help explain the astounding human ability to perform complex visual tasks in an ever-changing natural environment.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Whitney for discussions regarding this manuscript. We also thank J. Gao, N. Bilenko, T. Naselaris, A. Vu and M. Oliver for their help in various aspects of this research. This work was supported by the National Eye Institute (EY019684 and EY022454) and the Center for Science of Information, a National Science Foundation Science and Technology Center, under grant agreement CCF-0939370.

AUTHOR CONTRIBUTIONS

T.Ç. and S.N. designed the experiments. T.Ç. and A.G.H. operated the scanner. T.Ç. conducted the experiments and analyzed the data. T.Ç. and J.L.G. wrote the manuscript. J.L.G. provided guidance on all aspects of the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Olshausen, B.A., Anderson, C.H. & Van Essen, D.C. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**, 4700–4719 (1993).
2. Luck, S.J., Chelazzi, L., Hillyard, S.A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
3. McAdams, C.J. & Maunsell, J.H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
4. Reynolds, J.H., Pasternak, T. & Desimone, R. Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703–714 (2000).
5. David, S.V., Hayden, B.Y., Mazer, J.A. & Gallant, J.L. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* **59**, 509–521 (2008).
6. Connor, C.E., Preddie, D.C., Gallant, J.L. & Van Essen, D.C. Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214 (1997).
7. Asaad, W.F., Rainer, G. & Miller, E.K. Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* **84**, 451–459 (2000).
8. Warden, M.R. & Miller, E.K. Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* **30**, 15801–15810 (2010).
9. Johnston, K. & Everling, S. Neural activity in monkey prefrontal cortex is modulated by task context and behavioral instruction during delayed-match-to-sample and conditional prosaccade-antisaccade tasks. *J. Cogn. Neurosci.* **18**, 749–765 (2006).
10. Mazer, J.A. & Gallant, J.L. Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron* **40**, 1241–1250 (2003).
11. Huth, A.G., Nishimoto, S., Vu, A.T. & Gallant, J.L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
12. Peelen, M.V., Li, F.F. & Kastner, S. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* **460**, 94–97 (2009).
13. Li, F.F., VanRullen, R., Koch, C. & Perona, P. Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. USA* **99**, 9596–9601 (2002).
14. O'Craven, K.M., Downing, P.E. & Kanwisher, N. fMRI evidence for objects as the units of attentional selection. *Nature* **401**, 584–587 (1999).
15. Reddy, L. & Kanwisher, N. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr. Biol.* **17**, 2067–2072 (2007).

16. Bartels, A. & Zeki, S. Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* **21**, 75–85 (2004).
17. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).
18. Kay, K.N., Naselaris, T., Prenger, R.J. & Gallant, J.L. Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).
19. Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
20. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M. & Gallant, J.L. Bayesian reconstruction of natural images from human brain activity. *Neuron* **63**, 902–915 (2009).
21. Mitchell, T.M. *et al.* Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
22. Miller, G. WordNet: a lexical database for English. *Commun. ACM* **38**, 39–41 (1995).
23. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
24. Corbetta, M. & Shulman, G.L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
25. Carter, C.S. Anterior cingulate cortex, error detection and the online monitoring of performance. *Science* **280**, 747–749 (1998).
26. Womelsdorf, T., Anton-Erxleben, K., Pieper, F. & Treue, S. Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat. Neurosci.* **9**, 1156–1160 (2006).
27. Haxby, J.V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
28. Turin, G. An introduction to matched filters. *IEEE Trans. Inf. Theory* **6**, 311–329 (1960).
29. Navalpakkam, V. & Itti, L. Search goal tunes visual features optimally. *Neuron* **53**, 605–617 (2007).

ONLINE METHODS

Subjects. Five healthy adult volunteers (five males) with normal or corrected to normal vision participated in this study: S1 (age 30), S2 (age 32), S3 (age 25), S4 (age 25) and S5 (age 26). The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley, and written informed consent was obtained from all subjects.

Stimuli. For each attention condition in the main experiment, 1,800 s of continuous color natural movies (24° × 24°, 512 × 512 pixels) were presented without repetition in a single session. The stimuli were compiled by combining many short clips (10–20 s) from a diverse selection of natural movies[19]. Only humans or only vehicles were each presented for 450 s, the two categories co-occurred for 450 s, and both categories were absent for 450 s. Humans and vehicles appeared in highly diverse scenes and in many different positions, sizes and viewpoints. A fixation spot (0.16° square) was superimposed on the movies and its color was alternated at 1 Hz, rendering it continuously visible. The stimuli were presented at a rate of 15 Hz using an MR-safe projector (Avotec) and a custom-built mirror system.

Experimental procedure. Each subject participated in a total of seven scan sessions. Functional localizer, retinotopic mapping and anatomical data were collected in two sessions. Functional scans for the main experiment were collected in a single scan session. To increase sensitivity for the analysis performed in the absence of the target stimuli, we collected another session of functional data using the same experimental design, but with a different set of movie clips. To construct the continuous semantic space, we presented 7,200 s of natural movies in three separate sessions while subjects performed a passive-viewing task.

In the main experiment, subjects fixated continuously while covertly searching for humans or vehicles in natural movies. To ensure continuous vigilance, subjects depressed a response button continuously whenever an exemplar of the attended category was present in the movies. The data for each attention condition were recorded in three separate 10-min runs. The movie clips in each run were selected randomly without repetition. To avoid sampling bias, we presented an identical set of movie clips for both attention conditions. The presentation order of these clips was counterbalanced across the conditions. Four mutually exclusive classes of stimuli (that is, only humans, only vehicles, both humans and vehicles, and neither humans nor vehicles) were randomly interleaved and evenly distributed in and across the runs. The attended category was fixed in each run. The attention conditions were alternated in consecutive runs. A cue word, humans or vehicles, was displayed before each run to indicate the attended category. To compensate for hemodynamic transients caused by movie onset, each run was preceded by the last 10 s of that run. Data collected during the transient period were discarded.

MRI protocols. MRI data were acquired on a 3 T Siemens scanner located at the University of California, Berkeley using a 32-channel head coil. Functional data were acquired using a T_2*-weighted gradient-echo EPI sequence customized with a water-excitation radiofrequency pulse to prevent contamination from fat signal. The following parameters were prescribed: repetition time = 2 s, echo time = 34 ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm^3, field of view = 224 × 224 mm^2, and 32 axial slices to cover the entire cortex. Head motion was minimized with foam padding. To reconstruct cortical surfaces, we collected anatomical data with 1 × 1 × 1 mm^3 voxel size and 256 × 212 × 256 mm^3 field of view using a three-dimensional T_1-weighted MP-RAGE sequence. The anatomical and retinotopic mapping data for subjects S2 and S3 were obtained on a 1.5 T Philips Eclipse (Philips Medical Systems) scanner.

Data pre-processing. Functional scans were intra- and inter-run aligned using the Statistical Parameter Mapping toolbox (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). All volumes were aligned to the first image from the first functional run for each subject. Non-brain tissue was excluded from further analysis using the Brain Extraction Tool (BET, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET). Voxels whose BOLD responses are primarily driven by button presses were identified using a motor localizer that included a button-press task. The identified voxels were contained in the primary motor, somatosensory motor and premotor cortices, and voxels in these regions were excluded from analysis. The cortical surface of each subject was reconstructed from anatomical data using Caret5 (http://www.nitrc.org/projects/caret/). Cortical voxels were identified as the set of voxels in a 4-mm radius of the cortical surface. Subsequent analyses were restricted to 47,125–53,957 cortical voxels identified for the various subjects.

The low-frequency drifts in voxel responses were estimated using a 240-s-long cubic Savitzky-Golay filter for each run (10 min). The drifts were removed from the responses, which were then normalized to have zero mean and unit variance. Neither spatial nor temporal averaging was performed on the data during pre-processing and model-fitting stages. The data from separate subjects were not transformed into a standard brain space.

Functional localizer and retinotopic mapping data were used to assign voxels to the corresponding ROIs[11]. All functional ROIs were defined on the basis of relative response levels to contrasting stimuli (t test, $P < 10^{-5}$, uncorrected).

Category model. The object and action categories in each 1-s clip of the natural movie stimulus were manually labeled using terms from the WordNet lexicon[22]. Three naive raters performed the labeling, and potential conflicts were resolved by conferral among all raters. In WordNet, words are grouped into sets of synonyms according to the concepts they describe and are organized into a hierarchical network of semantic relations on the basis of word meaning. By definition, the existence of a category in a given scene indicates the existence of all of its superordinate categories. For example, if a clip is labeled with 'child', it also contains the following categories: offspring, relative, person, organism, living thing, whole, object and entity. To facilitate labeling, the raters exploited these hierarchical relationships in WordNet. The raters initially labeled 604 object and action categories, and inferred the presence of 331 superordinate categories from these initial labels.

A stimulus time course (categories × seconds) was then formed using a binary variable to indicate the presence or absence of each category in each 1-s movie clip. The category model fit to each voxel describes evoked responses as a weighted linear combination of these indicator variables. The predicted response of each voxel to any category is the sum of weights for all the categories it encompasses (including itself). In other words, the weight for each category is the estimated difference between the response to that specific category and the cumulative response to all of its superordinate categories.

Retinotopically organized early visual areas (V1–4) are selective to structural characteristics of visual stimuli[19]. To ensure that model fits were not biased by structural differences in movie clips, one additional regressor was included in the model that characterizes the total motion energy in each 1-s clip. This regressor was computed as the average response of 2,139 space-time quadrature Gabor filter pairs to the movie stimuli. The filters were selected to cover the entire image space (24° × 24°) and reflected a wide range of preferred receptive-field sizes, orientations and spatiotemporal frequencies. In addition, to ensure that semantic tuning changes do not simply reflect tuning changes for elementary visual features, a separate structural model (with all 2,139 filter pairs) was fit to each voxel.

To ensure that the results were not biased by the hierarchical relationships in WordNet, reduced category models were fit using the subset of regressors for the 604 initially labeled categories. The data presented here was also analyzed in this separate framework, and no substantial discrepancies were observed in the obtained results. Furthermore, the original full category model outperformed this reduced category model in terms of prediction accuracy of BOLD responses (**Supplementary Fig. 12**). This indicates that the full category model provides a better description of category selectivity in cortical voxels.

Model fitting. The model for each attention condition was fit separately to 1,800 s of stimuli and responses. The stimulus time course was down-sampled by a factor of 2 to match the sampling rate of the measured BOLD responses. To model the slow hemodynamic response, each category was assigned a distinct time-inseparable finite impulse response filter with delays restricted to 2–6 s before the BOLD responses. All model parameters were simultaneously fit using L2-regularized linear regression.

To assess the significance of attentional tuning changes, a jackknifed model training and validation procedure was repeated 1,000 times. At each turn, 20% of the samples were randomly held out to validate the model performance. The regularization parameter (λ) for regression was selected with tenfold cross-validation on the remaining 80% of training samples. These samples were further split into 10% testing and 90% training sets at each fold. The trained models were tested on the 10% held-out sets by computing prediction scores. Prediction score was

doi:10.1038/nn.3381

taken as the correlation coefficient (Pearson's r) between the actual and predicted BOLD responses. The optimal λ was determined for each voxel by maximizing the average prediction score. To prevent potential bias in the models, we selected a final λ value as an intermediate between the optima for models from two attention conditions. The model-fitting procedures were performed with in-house software written in Matlab (MathWorks).

Characterizing tuning shifts. Attentional tuning shifts toward the target will increase the degree of tuning selectivity (tuning strength) for the attended category. Thus, the magnitude and direction of tuning shifts can be assessed by measuring the tuning strengths for humans and vehicles separately during each attention condition. Tuning strengths for humans and vehicles were quantified as the similarity between voxel tuning and idealized templates tuned solely for humans and vehicles, respectively. The templates were constructed by identifying the set of labels that belong to these categories (**Supplementary Fig. 5**). Tuning strength for each category was then quantified as Pearson's correlation between voxel tuning and the corresponding template.

$$s_{i,\mathrm{H}} = \mathrm{corr}(w_i, t_\mathrm{H})$$
$$s_{i,\mathrm{V}} = \mathrm{corr}(w_i, t_\mathrm{V})$$

Here, $s_{\mathrm{i,H}}$ is the tuning strength for humans and $s_{\mathrm{i,V}}$ is the tuning strength for vehicles during attention condition i (i = H: search for humans, and i = V: search for vehicles). Meanwhile, w_i is the voxel-wise tuning vector during condition i, and t_H and t_V are the templates for humans and vehicles, respectively. Finally, TSI was quantified using the measured tuning strengths for each voxel.

$$\mathrm{TSI} = \frac{(s_{\mathrm{H,H}} - s_{\mathrm{H,V}}) + (s_{\mathrm{V,V}} - s_{\mathrm{V,H}})}{2 - \mathrm{sign}(s_{\mathrm{H,H}} - s_{\mathrm{H,V}})s_{\mathrm{H,V}} - \mathrm{sign}(s_{\mathrm{V,V}} - s_{\mathrm{V,H}})s_{\mathrm{V,H}}}$$

Here, the numerator measures the difference in tuning strength for the attended versus unattended category, summed across two attention conditions. The denominator scales the TSI to range in [−1, 1]. Tuning shifts toward the attended category will yield positive TSIs, with a value of 1 in the case of a perfect match between voxel tuning and idealized template for the attended category. In contrast, tuning shifts away from the attended category will yield negative TSIs, with a value of −1 in the case of a total mismatch between voxel tuning and idealized template for the attended category. Finally, a TSI of zero indicates that the voxel tuning did not shift between the two attention conditions. Complementary tuning-shift analyses were performed in individual ROIs. For each attention condition, the mean tuning shift in each ROI was computed by averaging the TSI values of the corresponding set of voxels with significant models (t test, $P < 0.05$, FDR corrected) and positive prediction scores.

Eye-movement and behavioral controls. Eye movements are a legitimate concern in many experiments on visual perception and attention, especially when naive subjects are tested. However, four lines of evidence suggest that eye movements were not a problem in our experiment and that they could not have accounted for our results. First, all of the subjects tested in this experiment were highly trained psychophysical observers who had extensive experience in fixation tasks. Based on our previous work with trained and naive subjects, we fully expected that our trained observers would fixate much better than the naive subjects used in many attention experiments.

Second, we found no statistical evidence that fixation differed across attention conditions in the main and control analyses for any of the observers. Subjects' eye positions were monitored at 60 Hz throughout the scans using a custom-built camera system equipped with an infrared source (Avotec) and the ViewPoint EyeTracker software suite (Arrington Research). The eye tracker was calibrated before each run of data acquisition. A nonparametric ANOVA test was used to determine systematic differences in the distribution of eye positions. The eye position distributions were not affected by attention condition ($P > 0.24$), or by target presence or absence ($P > 0.61$). To determine whether the results were biased by explicit eye movements during target or distractor detection, we also analyzed the distribution of eye positions during 250-ms, 500-ms and 1-s windows around target onset and target offset. The eye position distributions were not affected by target onset ($P > 0.26$) or offset ($P > 0.49$). Furthermore, there were no significant interactions between any of the aforementioned factors ($P > 0.14$). To determine whether the results were biased by rapid moment-to-moment variations in eye position, we examined the moving-average s.d. of eye position in a 200-ms window (to capture potential saccades). There were no effects of attention condition ($P > 0.13$), target presence or absence ($P > 0.52$), target onset ($P > 0.47$), or target offset ($P > 0.17$), and there were no significant interactions between these factors ($P > 0.22$).

Third, although there may be some micro-saccade scale eye movements during covert visual search, we found no statistical evidence for a bias in the recorded BOLD responses across attention conditions. Specifically, there were no significant differences in BOLD responses resulting from interactions between the search task and scenes likely to contain the attended category or scenes that contained objects that share visual features with the attended category (two-way ANOVA, $F < 1.8$, $P > 0.18$, FDR corrected). Because we measured attentional tuning shifts using BOLD responses, this analysis indicates that small eye movements could not have accounted for our results.

Finally, to further ensure that the results were not confounded by eye movements, we regressed the moving-average s.d. of eye position out of the BOLD responses, and then we repeated the entire modeling procedure on these filtered data. Including this nuisance regressor did not affect the model fits or the results in any brain regions in which the category model provided significant response predictions.

Behavioral responses were also recorded during the scans with a fiber-optic response pad (Current Designs). A hit was defined as a button response detected within 1 s of the target onset in the movies. A false alarm was defined as a button response when the target was absent from the movies. The behavioral performance, as measured by the sensitivity index (d'), was compared across the two attention conditions using Wilcoxon rank-sum tests. Participants performed equally well when searching for either category, indicating that the task difficulty was balanced across attention conditions (**Supplementary Fig. 13**).

Head-motion and physiological-noise controls. To ensure that our results were not biased by head motion or physiological noise, we used estimates of these nuisance factors to regress them out of the BOLD responses, and we then repeated the entire modeling procedure on these filtered data. The moment-to-moment variations in head position were estimated during motion correction pre-processing. These six-parameter affine transformation estimates of head position were used to create head-motion regressors. The cardiac and respiratory states were recorded using a pulse oximeter and a pneumatic belt. These recordings were used to create pulse-oximetry and respiratory regressors as low-order Fourier series expansions of the cardiac and respiratory phases. The inclusion of these various nuisance regressors did not affect the model fits or the results in any brain region in which the category model provided significant response predictions.

Spatial-attention controls. Given the stimulus correlations inherent in natural movies, differences in spatial attention across attention conditions might have confounded our results, even in the absence of targets. We performed two additional control analyses to ensure that the results derived from target-absent movie clips were not biased by stimulus correlations. First, all target-absent movie clips were coded to indicate whether they contained objects that shared visual features with humans (that is, scenes that contain animals, body parts or animate motion) or with vehicles (that is, scenes that contain inanimate objects such as artifacts, buildings or devices). Thereafter, a two-way ANOVA was performed on the evoked BOLD responses to determine whether there was any interaction between scene content and attended category. There were no significant interactions between scene content and attended category ($F < 1.8$, $P > 0.18$, FDR corrected).

Second, all target-absent movie clips were coded to indicate whether humans were likely to appear (that is, scenes that contain animate motion, tools for human use, buildings or rooms) or whether vehicles were likely to appear (that is, scenes of urban areas or cities, and scenes containing roads or highways). Another two-way ANOVA was performed on the evoked BOLD responses to determine whether there was any interaction between scene type and attended category. There were no significant interactions between scene type and attended category ($F < 2.0$, $P > 0.16$, FDR corrected). Thus, we found no evidence for an interaction between scene content or type and the attentional target. These results suggest that the tuning shifts reported here are not biased by systematic differences in spatial attention across attention conditions.

Construction of the semantic space. To construct the continuous semantic space, we collected functional data while subjects passively viewed 7,200 s of natural movies. Voxel-wise tuning vectors were estimated using these data and following identical procedures to the main experiment. A semantic space of cortical representation was then derived using PCA across the tuning vectors of cortical voxels (following procedures described in ref. 11). PCA ensures that voxels tuned for similarly represented categories project to nearby points in the semantic space, whereas voxels tuned for dissimilarly represented categories project to distant points. Each principal component represents a distinct dimension of the semantic space, ordered according to percentage of variance explained. To maximize the quality of the semantic space, we selected only the first six principal components that captured approximately 30% of the variance. To perform analyses of attentional tuning changes in the semantic space, voxel-wise tuning vectors obtained under different attention conditions were first projected onto these principal components. The results did not substantially vary with the number of principal components used to define the semantic space.

Control analysis in the absence of target stimuli. A control analysis was performed to assess tuning changes for unattended categories in the absence of target stimuli. To increase sensitivity, additional functional data were collected in all subjects using the same experimental procedure. A total of 1,800 s of stimuli were compiled from a different selection of movie clips than those used in the main experiment. To estimate tuning, the BOLD responses to movie clips in which no humans or vehicles appeared were pooled across this additional session and the main experiment (yielding 900 s total). Tuning during each attention condition was estimated separately.

Tuning changes for unattended categories were measured, and the direction of tuning shifts with respect to the attended categories was then inferred from these measurements. For this purpose, we used the semantic space that assesses the similarity between the attended categories and remaining ones in terms of cortical representation. Specifically, if a single voxel's tuning shifts toward categories similar to humans, then we should find that its tuning vector is closer to the humans template than the vehicles template in the semantic space. To test this prediction, voxel-wise tuning vectors in the control analysis and the template vectors for the attended categories were projected into the semantic space. Thereafter, TSI was quantified following procedures in the main analysis, but, to increase sensitivity, we first computed the tuning change between the two attention conditions. We then computed an idealized tuning change between the template vectors in the semantic space. Finally, TSI was taken as the correlation between the actual and idealized tuning changes. As in the main analysis, tuning shifts toward the attended category will yield positive TSI values, whereas tuning shifts away from the target will yield negative TSI values.

The mean TSI was significantly greater than 0 in all subjects (Wilcoxon signed-rank test, $P < 10^{-6}$; **Supplementary Fig. 11**). This result clearly shows that attention shifts tuning of unattended categories toward the attended category even when the targets are not present. Furthermore, the tuning shifts were in consistent directions across the main (**Supplementary Fig. 7**) and control analyses for an average of 65.42 ± 7.73% (mean ± s.d., averaged across subjects) of cortical voxels. Although the direction of tuning shifts was highly consistent, the mean TSI was larger in the main analysis (when targets were present) than in the control analysis (when targets were excluded). TSI distributions were also less bimodal in the main analysis than in the control analysis. These differences are caused by two factors. First, the attentional effects on BOLD responses were strongest for the attended categories. Thus, tuning changes obtained when the targets were present (main analysis) were naturally stronger than the tuning changes that occurred when the targets were absent (control analysis). This reduced the TSI values in the control analysis.

Second, different metric spaces were used to estimate the TSI distributions in the main and the control analyses. In the main analysis, TSI was computed across 935 dimensions of the category model, and each category was treated as a separate dimension. As such, a tuning shift in the direction of humans represents tuning changes for humans alone. Thus, the main analysis only considers tuning changes for the attended categories, which account for 38.79 ± 0.07% (mean ± s.d.) of tuning changes in cortical voxels. However, in the control analysis, TSI was computed across six dimensions of the semantic space that organizes categories according to semantic similarity. A tuning shift in the direction of humans represents tuning changes for both humans and nearby categories in this space. Thus, the control analysis considers tuning changes for both attended and unattended categories, and tuning shifts toward the attended categories account for 72.70 ± 0.04% (mean ± s.d.) of tuning changes in cortical voxels. This causes TSI distributions to be more bimodal in the control analysis.

Cortical flat map visualization. The cortical surface of each hemisphere was flattened after five relaxation cuts were applied to reduce distortions. For surface-based visualization, functional data were aligned to the anatomical data using in-house Matlab scripts (MathWorks). The functional data were then projected onto the cortical surface. Each point in the generated flat maps corresponded to an individual voxel.

A custom color map was designed to simultaneously visualize the cortical distribution of tuning strength for the attended categories. The tuning strengths (that is, s_H for humans and s_V for vehicles) were measured as the correlations between the voxel-wise tuning vectors and the idealized templates tuned solely to these attended categories. Distinct colors were assigned to six landmark values of the pair (s_V, s_H): red for (0.75, 0), turquoise for (−0.75, 0), green for (0, 0.75), magenta for (0, −0.75), gray for (0, 0), and black for (−0.75, −0.75). The colors for the remaining values were linearly interpolated from these landmarks. A gray color was assigned to voxels with insignificant model weights.

A separate color map was designed to visualize the cortical distribution of semantic tuning. For this purpose, voxel-wise tuning vectors for each attention condition were projected into the semantic space. The first four principal components that captured approximately 20% of the variance were selected. The first principal component mainly distinguishes categories with high versus low stimulus energy and so was not visualized. The projections onto the second, third and fourth principal components were assigned to the red, green and blue channels. Voxels with similar semantic tuning project to nearby points in the semantic space and so they were assigned similar colors. In this color map, voxels tuned for humans and communication verbs appeared in shades of green-cyan. Voxels tuned for animals and body parts appeared in yellow-green, whereas those tuned for movement verbs appeared in red. Voxels tuned for locations, roads, devices and artifacts appeared in shades of purple, whereas those tuned for buildings and furniture appeared in blue. Finally, voxels tuned for vehicles appeared in magenta. A gray color was assigned to voxels with insignificant model weights.

Statistical procedures. Statistical comparisons of prediction scores were based on raw correlation coefficients between the predicted and actual responses. Prediction scores were Fisher transformed, and one-sided *t* tests were applied to assess significance. Although this procedure is appropriate for significance testing, noise in the measured BOLD responses biases raw correlation values downward[30]. Thus, to attain reliable estimates of model performance across subjects, correlation values were corrected for noise bias[11].

Unless otherwise noted, all other comparisons were performed using one-sided nonparametric Wilcoxon signed-rank tests. All statistical significance levels were corrected for multiple comparisons using FDR control[23].

30. David, S.V. & Gallant, J.L. Predicting neuronal responses during natural vision. *Network* **16**, 239–260 (2005).