Behavioral/Cognitive

Task-Dependent Warping of Semantic Representations during Search for Visual Action Categories

Mo Shahdloo,[1,2,3] Emin Çelik,[2,5] Burcu A. Urgen,[2,4,5] Jack L. Gallant,[6] and Tolga Çukur[2,3,5,6]

[1]Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Oxford OX3 9DU, United Kingdom, [2]National Magnetic Resonance Research Centre, Bilkent University, 06800 Ankara, Turkey, [3]Departments of Electrical and Electronics Engineering and, [4]Psychology, Bilkent University, 06800 Ankara, Turkey, [5]Neuroscience Program, Aysel Sabuncu Brain Research Centre, Bilkent University, 06800 Ankara, Turkey, and [6]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, California 94720

Object and action perception in cluttered dynamic natural scenes relies on efficient allocation of limited brain resources to prioritize the attended targets over distractors. It has been suggested that during visual search for objects, distributed semantic representation of hundreds of object categories is warped to expand the representation of targets. Yet, little is known about whether and where in the brain visual search for action categories modulates semantic representations. To address this fundamental question, we studied brain activity recorded from five subjects (one female) via functional magnetic resonance imaging while they viewed natural movies and searched for either communication or locomotion actions. We find that attention directed to action categories elicits tuning shifts that warp semantic representations broadly across neocortex and that these shifts interact with intrinsic selectivity of cortical voxels for target actions. These results suggest that attention serves to facilitate task performance during social interactions by dynamically shifting semantic selectivity toward target actions and that tuning shifts are a general feature of conceptual representations in the brain.

Key words: **attention; fMRI; natural movies; visual actions; voxelwise modeling**

Significance Statement

The ability to swiftly perceive the actions and intentions of others is a crucial skill for humans that relies on efficient allocation of limited brain resources to prioritize the attended targets over distractors. However, little is known about the nature of high-level semantic representations during natural visual search for action categories. Here, we provide the first evidence showing that attention significantly warps semantic representations by inducing tuning shifts in single cortical voxels, broadly spread across occipitotemporal, parietal, prefrontal, and cingulate cortices. This dynamic attentional mechanism can facilitate action perception by efficiently allocating neural resources to accentuate the representation of task-relevant action categories.

Introduction

The ability to swiftly perceive the actions and intentions of others is a crucial skill for all social animals. In the human brain this ability has been attributed to a network of occipitotemporal, parietal, and premotor areas collectively called the action observation network (AON; Oberman et al., 2007; Caspers et al., 2010; Molinari et al., 2013; Rozzi and Fogassi, 2017). Other reports suggest that the AON hierarchically represents diverse information pertaining to actions, ranging from shape and kinematics to action–effector interactions and action categories (Grafton and de C Hamilton, 2007; Oosterhof et al., 2010, 2012, 2013; Handjaras et al., 2015; Lingnau and Downing, 2015; Wurm et al., 2017; Cavina-Pratesi et al., 2018; Urgen et al., 2019). Low-level shape and movement kinematics are represented in occipitotemporal cortex and in the posterior bank of inferior temporal cortex (Jastorff and Orban, 2009). Effector type (e.g., foot, hand) is represented in ventral premotor cortex (Jastorff et al., 2010; Corbo and Orban, 2017), whereas parietal cortex represents higher level action categories (Abdollahi et al., 2013; Ferri et al., 2015).

Evidence suggests that selective attention alters population responses to actions across this representational hierarchy. Prior electrophysiology (Muthukumaraswamy et al., 2004; Muthukumaraswamy and Singh, 2008; Schuch et al., 2010; Puglisi et al., 2017, 2018) and neuroimaging studies (Rowe et al., 2002; de Lange et al., 2008; Safford et al., 2010; Herrington et al., 2012; Nicholson et al., 2017) have examined attention to low-level action features. Schuch et al. (2010) reported

Received July 1, 2021; revised June 29, 2022; accepted July 6, 2022.

Author contributions: T.C. designed research; M.S. and E.C. performed research; M.S., B.A.U., and T.C. analyzed data; M.S., J.L.G., and T.C. wrote the paper.

This work was supported in part by a Marie Curie Actions Career Integration Grant (PCIG13- GA-2013-618101), a European Molecular Biology Organization Installation Grant (IG 3028), a Turkish Academy of Sciences Young Scientist Outstanding Achievement Award Program 2015 Fellowship, and a Science Academy Young Scientists Program 2017 Fellowship.

The authors declare no competing financial interests.

Correspondence should be addressed to Tolga Çukur at cukur@ee.bilkent.edu.tr.

https://doi.org/10.1523/JNEUROSCI.1372-21.2022

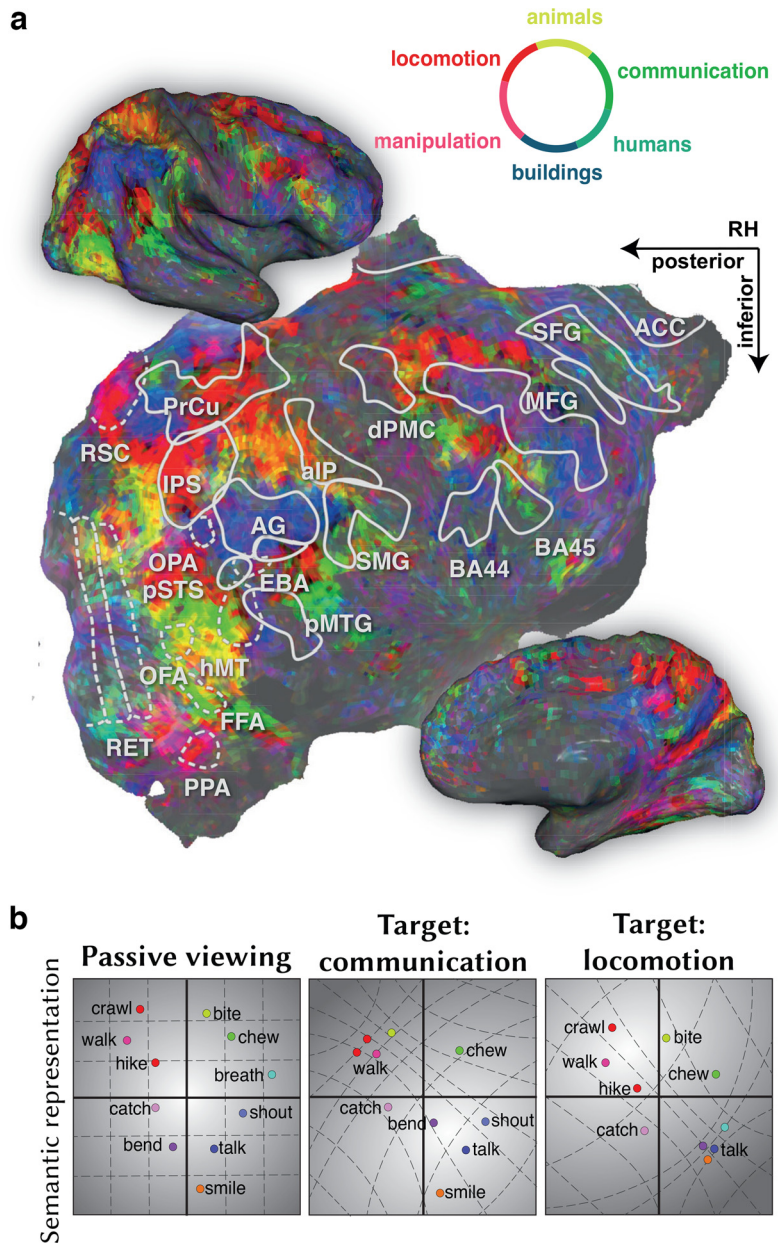

Figure 1. Hypothesized changes in semantic representation of action categories. Recent evidence suggests that the human brain organizes hundreds of object and action categories in a semantic space that is distributed systematically across the cerebral cortex (Huth et al., 2012). ***a***, Semantic representation for a single subject from Çukur et al. (2013) is shown on flattened cortical surface and on inflated hemispheres. Colors indicate tuning for different object or action categories (top right, color legend). Regions of interest identified using conventional functional localizers are denoted by white borders. For abbreviations for regions of interest, see below, Materials and Methods. ***b***, In the semantic space, action categories that are semantically similar to each other are mapped to nearby points, and semantically dissimilar actions are mapped to distant points. There is evidence that visual search for object categories warps semantic representation in favor of the targets by shifting single-voxel tuning for object categories toward target objects (Çukur et al., 2013). Thus, we hypothesized that visual search for a given action category should similarly expand the semantic representation of the target and semantically similar categories.

increased electroencephalography (EEG) responses in AON with attention to action kinematics. Safford et al. (2010) reported enhanced blood oxygen level-dependent (BOLD) responses in superior temporal sulcus (STS) with attention to animate actors (i.e., humans) presented via simplified point-light displays (Johansson, 1973). Nicholson et al. (2017) reported enhanced responses in inferior frontal gyrus (IFG), occipitotemporal cortex, and middle frontal gyrus (MFG) with attention to action goals and in parietal cortex and fusiform gyrus with attention to manipulated objects. Few reports have further investigated the effects of attention to higher level action features (Nastase et al., 2017, 2018). Presenting movie clips from various animal taxonomies performing several actions, Nastase et al. (2017) reported that attending to performed actions versus taxonomy alters multivariate response patterns across anterior intraparietal sulcus (IPS) and premotor cortex.

Current electrophysiology and neuroimaging findings on visual actions suggest that attention increases AON responses to target features ranging from action kinematics and goals to actors. That said, high-level semantic representations during visual search for specific action categories remain understudied. Furthermore, prior studies did not question whether attending to action features causes baseline and gain changes or rather elicits dynamic tuning shifts that can alter cortical representation. Evidence indicates that visual search for object categories shifts single-voxel category tuning toward target objects (Çukur et al., 2013). Therefore, it is likely that attention to action categories also causes tuning shifts to facilitate visual search. Here, we hypothesized that natural visual search for action categories induces semantic tuning shifts in single cortical voxels toward targets. Tuning shift toward target categories elevates the local sampling density near the target actions and expands target-action representations while compressing behaviorally irrelevant action representations by increasing the discriminability in the semantic neighborhood of the finely sampled action categories (Fig. 1).

To test the tuning-shift hypothesis, we recorded whole-brain BOLD responses while human subjects viewed 60 min of natural movies and covertly searched for either 14 communication actions or 30 locomotion actions among 109 action categories in the movies (Fig. 2, Extended Data Fig. 2-1). Using spatially informed voxelwise modeling (Çelik et al., 2019), we measured category responses for hundreds of objects and actions in the movies separately for each individual subject and for each search task. We estimated a semantic space underlying action category responses, and semantic tuning for action categories were measured by projecting voxelwise model weights onto this space. Finally, semantic tuning profiles during the two search tasks were compared to quantify the magnitude and direction of tuning shifts in single voxels.

Materials and Methods

Subjects

Five healthy adult volunteers with normal or corrected-to-normal vision who participated in this study were subject (S)1 (male, age 31), S2 (male, age 27), S3 (female, age 32), S4 (male, age 33), and S5 (male, age 27). Data were collected at the University of California, Berkeley. The

experimental protocol was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. All participants gave written informed consent before scanning.

Stimuli and experimental design

Data for the main experiment were collected in six 10 min 50 s runs in a single session. Continuous natural movies were used as the stimulus in the main experiment. Three distinct 10 min movie segments were compiled from short movie clips (10–20 s) without sound. Movie clips were selected from a diverse set of natural movies (Nishimoto et al., 2011). Movie clips were cropped into a square frame and downsampled to 512 × 512 pixels. The movie stimulus was displayed at 15 Hz on an MRI-compatible projector screen that covered a 24 × 24° visual angle. Subjects were instructed to covertly search for target categories in the movies while maintaining fixation. A set of instructions regarding the experimental procedure and exemplars of the search targets were provided to the subjects before the experiment. A color square of 0.16 × 0.16° at the center with color changing at 1 Hz was used as the fixation spot. A cue word was displayed before each run to indicate the attention target: communication or locomotion. The communication target contained actions with the intent of communication, including both verbal communication actions and nonverbal gestural communication actions (e.g., talking, shouting, smirking). The locomotion target contained locomotion-related actions with the intent of moving animate entities, including humans and anthropomorphized animals (e.g., moving, running, driving). The same movie stimuli were used during each of the two attention tasks. The order of attention conditions was interleaved across runs to minimize subject expectation bias. This resulted in the presentation of 1800 s of movies without repetition in each attention condition. Data from the first 20 s and last 30 s of each run were discarded to minimize effects of transient confounds. Following these procedures, 900 data samples for each attention condition were obtained.

A separate set of functional data were collected while the same set of subjects passively viewed 120 min of natural movies., passive-viewing data; this dataset was also used in Huth et al., 2012 but here it was reanalyzed with a focus on action categories). This dataset was used to construct the semantic space and to select voxels subjected to further analyses. Data for the passive-viewing experiment were collected in 12 10 min 50 s runs in which 12 separate movie segments were displayed. Presentation procedures were the same between the main experiment and passive-viewing experiment, save for the number of runs. The passive-viewing dataset contained 3600 data samples.

fMRI data collection

Data were collected on a 3T Siemens Tim Trio MRI scanner (Siemens Medical Solutions) via a 32-channel receiver coil. Functional data were collected using a T2*-weighted gradient-echo echo-planar imaging pulse sequence with the following parameters: TR = 2 s, TE = 33 ms, water-excitation pulse with flip angle = 70°, voxel size = 2.24 mm ×

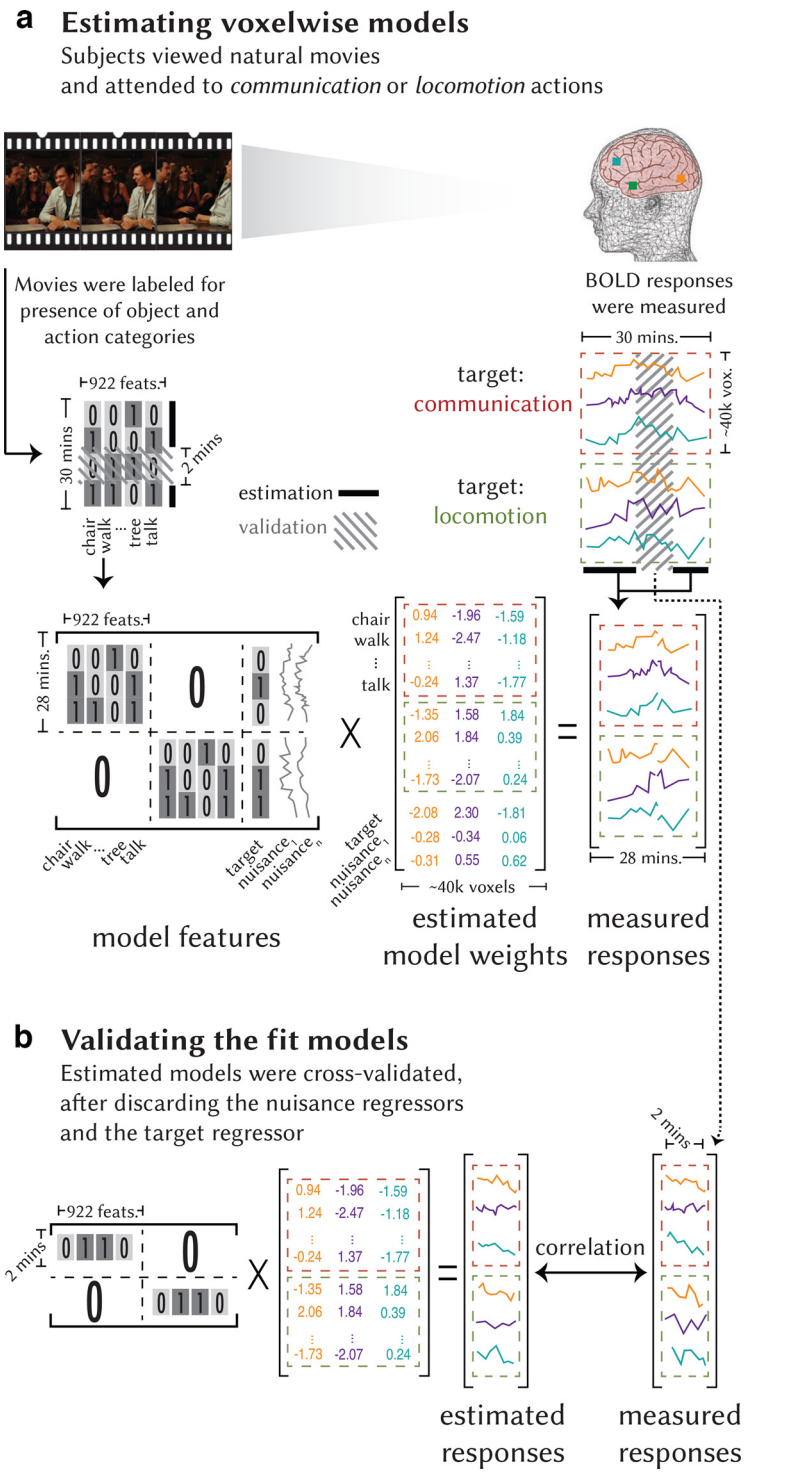

Figure 2. Model fitting and validation procedure. Undergoing fMRI, human subjects viewed 60 min of natural movies and covertly searched for communication or locomotion action categories while fixating on a central dot. ***a***, An indicator matrix was constructed that identified the presence of each of the 922 object and action categories in each 1 s clip of the movies (Extended Data Fig. 2-1). Nuisance regressors were included to account for head motion, physiological noise, and eye movement confounds. An additional nuisance regressor was included to account for target detection confounds. In a CV procedure, regularized linear regression was used to estimate separate category model weights (i.e., category responses) for each search task that mapped each category feature to the recorded BOLD responses in single voxels. ***b***, Accuracy of the fit models was cross-validated by measuring prediction performance on the held-out data in each CV fold after discarding the nuisance regressors and the target regressor. The prediction score of the fit models was taken as the product-moment correlation coefficient between estimated and measured BOLD responses, averaged across the two search tasks.

2.24 mm × 4.13 mm, field of view = 224 mm × 224 mm, 32 axial slices. To construct cortical surfaces, anatomic data were collected using a three-dimensional T1-weighted magnetization-prepared rapid-acquisition gradient-echo sequence with the following parameters: TR = 2.3 s, TE = 3.45 ms, flip angle = 10°, voxel size = 1 mm × 1 mm × 1 mm, field of view = 256 mm × 212 mm × 256 mm. Surface flattening and visualization were done via FreeSurfer and PyCortex software (Dale et al., 1999; Reuter et al., 2012; Gao et al., 2015).

fMRI data preprocessing

Motion correction was performed using Statistical Parametric Mapping (SPM12) toolbox (Friston et al., 1995). Functional volumes were aligned to the first image from the first run in each subject. Brain tissue was identified using the brain extraction tool (BET) from the Functional MRI of the Brain Software Library software package (Smith, 2002). Low-frequency response components were detected using a third-order Savitzky–Golay low-pass filter with 240 s temporal window and were removed from voxel responses. Voxel responses were then z scored to attain zero mean and unit variance. Voxels within the 2 mm neighborhood of the cortical sheet were identified as cortical voxels in each subject (S1, 37,791 voxels; S2, 32,671 voxels; S3, 36,942 voxels; S4, 42,090 voxels; S5, 39,254 voxels).

Definition of regions of interest

To define the anatomic regions of interest (ROIs) in each subject, the cortical surface was segmented into 156 regions of the Destrieux atlas (Destrieux et al., 2010) via FreeSurfer. Segmentation results were projected from the anatomic space onto the functional space using PyCortex, and each voxel was assigned an anatomic label based on the projections. Functional ROIs were identified in each subject using visual category and retinotopic localizers (Huth et al., 2012). Localizer experiments for visual category-selective areas [fusiform face area (FFA), occipital face area (OFA), parahippocampal place area (PPA), retrosplenial cortex RSC)] were performed in six 4.5 min runs of 16 blocks (Huth et al., 2012). Subjects passively viewed 20 random static images from one of the objects, scenes, body parts, faces, or spatially scrambled object groups in each block. Each image was shown for 300 ms following a 500 ms blank period. PPA and RSC were identified as voxels with positive scene versus objects contrast (t test, $p < 10^{-4}$, uncorrected). FFA and OFA were defined using face-versus-object contrast (t test, $p < 10^{-4}$, uncorrected). The boundaries of these areas were hand drawn on the cortical surfaces along the contours at which the contrast level reached half of the maximum. A localizer experiment for retinoptic early visual areas (RET; V1, V2, V3) contained four 9 min runs. Subjects viewed clockwise and counterclockwise rotating polar wedges in two runs. In the remaining two runs, subjects viewed expanding and contracting rings. Visual angle and eccentricity maps were used to define visual areas V1–3. Finally, ROIs were refined to voxels inside the drawn boundaries near a 2 mm neighborhood of the cortical sheet.

Abbreviations for regions of interest and important sulci

Several regions of interest and important sulci were labeled on the flattened cortical surfaces to guide the reader.

Regions of interest. Abbreviations are pMTG (posterior middle temporal gyrus), pSTS (posterior superior temporal sulcus), AG (angular gyrus), SMG (supramarginal gyrus), IPS (intraparietal sulcus), aIP (anterior intraparietal cortex), PrCu (precuneous), dPMC (dorsal premotor cortex), BA44/45 (Brodmann area 44/45), MFG (middle frontal gyrus), SFG (superior frontal gyrus), ACC (anterior cingulate cortex), RET (retinoptic early visual areas V1–3), FFA (fusiform face area), OFA (occipital face area), PPA (parahippocampal place area), RSC (retrosplenial cortex).

Sulci. Abbreviations are TOS (temporo-occipital sulcus), STS (superior temporal sulcus), SF (Sylvian fissure), IFS (inferior frontal sulcus), MFS (middle frontal sulcus), SFS (superior frontal sulcus).

Head motion, eye movement, and physiological noise

To prevent head motion and physiological noise confounds, estimates of these nuisance factors were regressed out of the BOLD responses. Six affine motion time courses estimated during the motion correction stage were taken as the head motion regressors. The cardiac and respiratory activity during the main experiment was recorded using a pulse oximeter and a pneumatic belt. These data were then used to estimate two regressors to capture respiration and nine regressors to capture cardiac activity (Verstynen and Deshpande, 2011).

To ensure that eye movements did not unduly bias the results, several control analyses were performed. ViewPoint EyeTracker (Arrington Research) was used to monitor subjects' eye positions at 60 Hz after getting calibrated at the beginning of each experimental run. Kruskal–Wallis tests were used to detect systematic differences in the distribution of eye position and movement. The distribution of eye position during search for communication and locomotion tasks were examined. We find that the distribution of eye position is not affected by search task ($p = 0.17$) or by target presence or absence ($p = 0.74$), and no significant interactions are present between these two factors ($p = 0.60$). To test whether eye movement is affected by target or distractor detection, the distribution of eye position during a 1 s window around target onset and target offset was studied. The eye position distribution is not affected by target onset ($p = 0.73$) or offset ($p = 0.17$), and there is no significant interaction between the aforementioned factors ($p = 0.83$). Furthermore, the moving-average SD of eye position was studied in a 200 ms window to determine systematic differences in rapid moment-to-moment variations in eye position across the two search tasks. There are no significant effects of search task ($p = 0.11$), target presence or absence ($p = 0.32$), target onset ($p = 0.49$), or target offset ($p = 0.36$), and there are no significant interactions among these factors ($p = 0.16$). Finally, moving-average SD of eye position was included in the model as a nuisance regressor and was regressed out of the BOLD responses.

To maintain subject vigilance, the subjects were instructed to depress a button whenever they detected a member of the target category in the stimulus (i.e., either a communication or a locomotion action depending on the search task). The behavioral responses were initially analyzed to ensure that subjects performed the tasks and that task difficulty was balanced across search targets. The target detection rate was 89 ± 9% for the communication and 91 ± 8% for the locomotion targets (mean ± SD across subjects), with no significant difference between the two tasks (bootstrap test, $p > 0.05$).

Category features

A category feature space was constructed to encode the information pertaining to object and action categories in the movies. Each second of the movie stimulus was manually labeled using the WordNet lexicon (Miller, 1995) to find the time course for the presence of 922 different object and action categories in the movie stimulus. This yielded an indicator matrix where each row represents a 1 s clip of the movie stimulus, and each column represents a category. Finally, category features were obtained by downsampling the indicator matrix to 0.5 Hz to match the acquisition rate of fMRI.

Motion-energy features

To infer cortical selectivity for low-level scene features, local spatial frequency and orientation information of each frame of the movie stimulus were quantified using a motion-energy filter bank. The filter bank contained 2139 Gabor filters that were computed at eight directions (0–350° in 45° steps), three temporal frequencies (0, 2, and 4 Hz), and six spatial frequencies (0, 1.5, 3, 6, 12, and 24 cycles/image). Filters were placed on a square grid spanning the 24 × 24° field of view. The luminance channel was extracted from the movie frames and passed through the filter bank. The outputs were then passed through a compressive nonlinearity to yield the motion-energy features (Nishimoto et al., 2011; Lescroart and Gallant, 2019). Finally, the motion-energy features were temporally downsampled to match the fMRI acquisition rate.

Space-time interest points features

Intermediate-level kinematic information of the movies were quantified by constructing the Space-Time Interest Point (STIP) features using STIP toolbox (Laptev, 2005; Laptev et al., 2008). STIP features have been successfully leveraged in many computer vision applications to recognize

human actions. As detailed in Laptev (2005) and Laptev et al. (2008), Harris operators were used to identify spatiotemporal interest points in the movie stimulus at multiple scales $(\sigma_i^2, \tau_j^2) = (2^{1+i}, 2^j)$, $i \in \{1, ..., 6\}$, $j \in \{1, 2\}$, where σ and τ are the standard deviations of the Gaussian kernels in spatial and temporal domains, respectively. Histograms of oriented gradients (Dalal and Triggs, 2005), and histograms of optical flow (Holte et al., 2010) were calculated in the $(\Delta_{x,i}, \Delta_{y,i}, \Delta_{t,j})$ spatiotemporal neighborhood of each interest point, where $\Delta_{x,i} = \Delta_{y,i} = 2k\sigma_i$ and $\Delta_{t,j} = 2k\tau_j$, and k is the scale factor. The scale factor was set to 9 according to the default configuration of the toolbox. Finally, normalized histograms were concatenated to construct the collection of 162 STIP features and were downsampled to match the acquisition rate of fMRI.

Model estimation and testing

Separate linearized models were fit in each voxel to estimate model weights that map each set of features (i.e., category, motion-energy, or STIP features) to the measured BOLD responses in each search task in individual subjects. Banded-ridge regression (Nunez-Elizalde et al., 2019) was used to fit the models. To capture the hemodynamic response, delayed feature time courses were concatenated. Delays of two, three, and four samples, corresponding to 4, 6, and 8 s were used. To account for potential correlations between target detection and BOLD responses, a nuisance target-presence regressor was included in the model. The target-presence regressor contained the category regressor for communication during search for the communication task and the category regressor for locomotion during search for the locomotion task. Model fitting for the two search tasks was performed concurrently by concatenating the features and BOLD responses across search tasks (Fig. 2). This procedure ensured consistency between the assigned regularization parameters across search tasks and enabled use of the target regressor (Shahdloo et al., 2020).

A nested cross-validation (CV) procedure was used to choose the regularization parameters and estimate model weights. Data from the main experiment were segmented into 60 30 s blocks. In each of the 10 outer folds, four randomly chosen blocks were held out as validation data. Then, in each of the 10 inner folds, 54 randomly chosen blocks were used as training data, and the 2 remaining blocks were used as test data. To fit models for the passive-viewing data, data were segmented into 144 50 s blocks. In each fold, eight randomly chosen blocks were held out as validation data, 132 randomly chosen blocks were used as training data, and the four remaining blocks were used as test data. For each feature set, regularization parameters were selected with a random search; a thousand normalized regularization parameter candidates were sampled from a Dirichlet distribution and were scaled by 30 log-spaced values ranging from 10^{-5} to 10^{20}. Training data were used to fit models for each set of regularization parameters independently. Model weights were then used to predict responses in the test data, and prediction scores of the fit models were assessed. Prediction scores were taken as the product-moment correlation coefficient between measured and predicted voxel responses. The set of regularization parameters maximizing the average prediction score across inner CV folds was chosen in each voxel. Finally, the optimal set of parameters was used to fit models on the union of training and test data in each outer fold, and model weights were averaged across the outer folds.

Finally, prediction performance of the fit models were evaluated. In each outer fold, after discarding the nuisance regressors, responses were predicted for the validation data using the fit models, and prediction scores were averaged across the search tasks. Prediction scores were then averaged across the outer folds.

For each voxel, separate linearized models were estimated to relate each feature representation to the BOLD responses. Specifically, category models were fit to estimate category responses that represented the contribution of each category to single-voxel BOLD responses separately for the data in the main experiment and the passive-viewing data in individual subjects. Furthermore, a motion-energy model and a STIP model were fit in each voxel to represent the contribution of the low- and intermediate-level stimulus features to the responses. These alternative models were further used to select analysis voxels (i.e., semantic voxels).

Variance partitioning

Object-action categories can be correlated with low-level visual features of natural movies (Lescroart and Gallant, 2019), and there is evidence for representation of intermediate-level action features (e.g., action kinematics) across cortex (Jastorff et al., 2010). Therefore, there is a possibility that the estimated category responses are confounded by selectivity for low- and intermediate-level scene features. To control for potential confounds, we performed a variance partitioning analysis. This analysis estimates the response variance that is uniquely explained by the category model after accounting for variance that can be attributed to low- and intermediate-level features captured by the motion-energy and STIP models. To do this, we separately measured the variance explained when all three models (category, motion energy, and STIP) are fit simultaneously (i.e., combined model), and variance explained when only motion-energy and STIP models are fit simultaneously (i.e., control model). Banded ridge regression was used to fit the combined and control models to prevent bias in assigning regularization parameters across different feature sets. The explained variance (R^2) was calculated as squared prediction scores, separately for the combined and control models. Note that from a model-fitting perspective, negative prediction scores that correspond to zero explained variance. Finally, unique variance explained by the category model was calculated as follows:

$$\hat{R}^2_{cat} = R^2_{comb} - R^2_{cont}. \quad (1)$$

Here, $\hat{R}^2_{cat}$ is the variance uniquely explained by the category model after accounting for low- and intermediate-level features, R^2_{comb} is the variance explained by the combined model, and R^2_{cont} is the variance explained by the union of motion-energy and STIP models in each voxel.

Action category responses

The fit category responses reflect voxel tuning for each of the 922 object and action categories in the movie stimulus. To infer tuning for action categories, 922-dimensional category responses were masked to select only the 109 action categories. This yielded the voxelwise 109-dimensional action category responses.

Semantic representation of actions

Passive-viewing data were used to construct a continuous semantic space for action category representation. In this space, semantically similar action categories would project to nearby points, whereas semantically dissimilar categories would project to distant points (Huth et al., 2012). Category models were fit, and action category responses during passive viewing were estimated. A group semantic space was then obtained using principal component analysis (PCA) on the action category responses of cortical voxels pooled across all subjects. To maximize the quality of the semantic space, voxels in which the category model predicted unique response variance after accounting for the variance attributed to low- and intermediate-level stimulus features were selected. These voxels were further refined to include only the top 3000 best predicted voxels within each subject. The top 12 principal components (PCs) that explained >95% of the variance in responses were selected. Subsequent analyses were also repeated using the top eight PCs that explained >90% of the response variance, but the results remained consistent. The semantic tuning profile for each voxel under each search task was then obtained by projecting the respective action category responses onto the PCs. To illustrate the semantic content of the PCs, characteristic actions of the movie stimulus were clustered in the semantic space, and cluster centers were projected onto the PCs after getting labeled (see Fig. 6).

Consistency of the semantic space across subjects

To test whether the estimated semantic space is consistent across subjects, we used a leave-one-out cross-validation procedure. In each cross-validation fold, voxels from four subjects were used to derive 12 PCs to construct a semantic space. In the left-out subject, the semantic tuning profile for each voxel was obtained by projecting action category responses during passive viewing onto the derived PCs. Next, the

product-moment correlation coefficient was calculated between the tuning profiles in the derived space and the tuning profiles in the original semantic space. Results were averaged across semantic voxels in the left-out subject. The cross-validated semantic spaces consistently correlate with the original semantic space (see Fig. 7).

Characterizing tuning shifts

Attentional tuning shifts toward or away from targets would be reflected in modulation of semantic selectivity for communication or locomotion action categories. Thus, the magnitude and direction of tuning shifts can be assessed by comparing the semantic selectivity for these categories between the two search tasks. Semantic selectivity for the two target categories was quantified as the similarity between semantic tuning profiles and idealized templates tuned solely for communication or locomotion action categories. First, idealized category responses were constructed as 109-dimensional vectors that contained ones for target categories (either communication or locomotion categories) and zeros previously. Idealized templates were then obtained by projecting these idealized category responses onto the semantic space. Semantic selectivity for each target category was quantified as the product-moment correlation coefficient between the voxel semantic tuning profile and the corresponding template as follows:

$$T_{i,C} = corr(s_i, s'_C) \tag{2}$$

$$T_{i,L} = corr(s_i, s'_L), \tag{3}$$

where $T_{i,C}$ and $T_{i,L}$ are the tuning strength for communication (C) and locomotion (L) during condition $i \in \{C, L\}$ denoting attend to communication or attend to locomotion; s_i is the semantic tuning profile during condition i, and s'_C and s'_L denote the idealized semantic tuning templates for communication and locomotion, respectively. Finally, voxelwise tuning shift index (TSI_{all}) was quantified as follows:

$$TSI_{all} = \frac{(T_{C,C} - T_{C,L}) + (T_{L,L} - T_{L,C})}{2 - sign(T_{C,C} - T_{C,L})T_{C,L} - sign(T_{L,L} - T_{L,C})T_{L,C}}. \tag{4}$$

The numerator of TSI captures the difference in semantic selectivity for the attended versus unattended category, summed over the two attention tasks (i.e., search for communication and search for locomotion). Observing that the maximum possible selectivity for the attended category is 1, obtained when voxel tuning is equivalent to the idealized template, the denominator is cast to normalize the potential range of the TSI metric between 1 and −1 without affecting its sign. Tuning shifts toward the attended category would yield positive values where a TSI_{all} of 1 indicates a complete match between voxel semantic tuning and idealized templates, whereas negative values would indicate shifts away from the attended category where a TSI_{all} of −1 indicates a complete mismatch between voxel tuning and idealized templates. A TSI_{all} of 0 would indicate that the voxel tuning did not shift between the two search tasks.

The TSI metric in Equation 4 can also be adopted to calculate tuning changes for any given set of action categories. To do this, the 922-dimensional category responses measured during attention tasks were masked to keep only the responses for the given set of actions. The masked tuning vectors and the idealized template for the given set were then projected onto the 12-dimensional semantic space. Semantic selectivity of a voxel to the given set was taken as the correlation coefficient between the projections of voxel tuning and the idealized template in the semantic space. Attentional modulation of semantic tuning for nontarget categories was examined by calculating a separate tuning shift index (TSI_{nt}). Note that this index can be calculated based on Equation 3 but by zeroing out the category responses for communication and locomotion actions before projection onto the semantic space. To study the tuning shifts in an ROI, TSIs were averaged across semantic voxels within the ROI.

The change in voxelwise tuning during attending to the first target (e.g., communication) versus the second target (e.g., locomotion) was defined as the l_1-norm of the tuning difference between the two conditions. This calculated tuning change can be linearly decomposed into a component explained by the target features (i.e., the union of communication and locomotion features) and a component explained by the nontarget features (i.e., all features excluding the target features). The fraction of tuning change for target/nontarget features was computed by taking the ratio of the respective component to the overall tuning change.

Characterizing target preference during visual search

To investigate the interaction between tuning shifts and intrinsic selectivity for individual target action categories, we quantified a target preference index (PI; $PI \in [-1,1]$) separately during the search for communication actions (PI_{com}) and during the search for locomotion actions (PI_{loc}). PI during the search for each target action was taken as the difference in selectivity for the attended versus the unattended target as follows:

$$PI_{com} = \frac{T_{C,C} - T_{C,L}}{1 - sign(T_{C,C} - T_{C,L})T_{C,L}} \tag{5}$$

$$PI_{loc} = \frac{T_{L,L} - T_{L,C}}{1 - sign(T_{L,L} - T_{L,C})T_{L,C}}, \tag{6}$$

where PI_{com} denotes the relative tuning preference for communication actions during the search for communication, and PI_{loc} denotes the relative tuning preference for locomotion actions during search for locomotion. In this scheme, a PI of 1 indicates a complete match between voxel semantic tuning and the idealized template for the target, whereas a PI of −1 indicates a complete mismatch between voxel tuning and the idealized template for the target. Finally, a PI of zero indicates that the voxel semantic tuning does not shift toward any of the target actions.

Characterizing action category preference during passive viewing

To investigate the interaction between the calculated preference index for individual targets and intrinsic selectivity for action categories, we quantified a selectivity index (SI; $SI \in [-1,1]$), separately for communication actions (SI_{com}) and for locomotion actions (SI_{loc}). SI for each target in each voxel was calculated as the product-moment correlation coefficient between the voxel category response and idealized template category tuning for the given target.

Action clustering

To facilitate interpretation of stimulus information captured by individual PCs, the characteristic action content of the movies was clustered and labeled. Action content ($\bar{C}$) for each short clip was calculated as the number of frames where each of the 109 actions were present ($\overline{N_a}$) and was normalized by the total number of clip frames (N) as follows:

$$\bar{C} = \frac{\overline{N_a}}{N}. \tag{7}$$

This yielded a 109-dimensional action content vector for each clip. The action content vectors were then projected onto the semantic space and were grouped into 10 clusters using k means. The number of clusters was optimized using the elbow method (Thorndike, 1953). Average action content of each clip (A) was calculated as the mean of the clip's action content vector as follows:

$$A = \frac{\sum \bar{c}}{109}, \tag{8}$$

where $\bar{C} = [c_1, c_2, c_3, \ldots, c_{109}]$. To label the clusters, five clips with the highest average action contents within each cluster were selected. Four candidate labels for each cluster were manually assigned, and 15 evaluators were asked to score (from 1 to 5) the correspondence of the selected

clips to each of the four candidate labels. Finally, the label with the highest score was selected to represent each cluster.

Statistical analyses

Bootstrap tests were used to assess statistical significance. To assess significance of the prediction scores, single-voxel predicted responses were resampled 5000 times with replacement. For each bootstrap sample, the prediction score was computed. The significance level (p value) of the prediction scores was taken as the fraction of bootstrap samples in which the prediction scores were >0. The significance level of the unique response variance (Eq. 1) was taken as the fraction of bootstrap samples in which the unique variance explained by the category model was >0. All single-voxel significance levels were corrected to account for multiple comparisons using the false discovery rate correction (FDR; Benjamini and Hochberg, 1995).

Significance of TSI_{all}, TSI_{nt}, PI_{com}, and PI_{loc} was assessed for each ROI across subjects. To do this, ROI-wise metrics were resampled across subjects with replacement 10,000 times. Significance level was taken as the fraction of bootstrap samples where the test metric averaged across resampled subjects is <0 (for right-sided tests) or >0 (for left-sided tests). This procedure was performed in a total of 21 functional ROIs separately. All ROI significance levels were corrected to account for multiple comparisons using FDR.

In ROIs with a significant metric across subjects, the metric was further tested within individual subjects. To do this, semantic voxels within a given ROI were resampled with replacement 10,000 times. For each bootstrap sample, mean value of a given metric was computed across resampled voxels. The significance level was taken as the fraction of bootstrap samples in which the tested metric was <0 (for right-sided tests) or >0 (for left-sided tests).

Data availability

Data supporting the findings of this study are available from the corresponding authors on request. Results can be explored online via an interactive brain viewer at http://www.icon.bilkent.edu.tr/brainviewer/shahdloo_etal/. The codes used to estimate spatially informed voxelwise model weights are freely available on GitHub at https://github.com/icon-lab/SPIN-VM.

Results

Visual search modulates category responses

Little is known on whether and where in the brain natural visual search for action categories warps semantic representations. To answer this question, we investigated voxelwise tuning for hundreds of object and action categories across cortex. Human subjects viewed natural movies and covertly searched for communication or locomotion actions. Category regressors were constructed to label the presence of 922 distinct object and action categories in the movies. Separate category models were then fit in each voxel for each search task. These models enabled us to measure single-voxel category responses during each search task (Fig. 2*a*; see above, Materials and Methods).

As natural stimuli contain correlations among various levels of features, there is a possibility that estimated category responses are confounded by voxel tuning for low- and intermediate-level scene features. To rule out this potential confound, we measured the response variance explained by low-level motion-energy features, and intermediate-level STIP features. Motion-energy features were constructed using a pyramid of spatiotemporal Gabor filters (Nishimoto et al., 2011). STIP features, providing an intermediate representational basis for human actions, were constructed by measuring optical flow over interest points with significant spatiotemporal variation (Laptev et al., 2008). We identified voxels in which the category model explained unique response variance after accounting for these alternative features via variance partitioning, and subsequent analyses were conducted on this set of uniquely explained voxels. To prevent bias in voxel selection because of attention, variance partitioning was performed on a separate dataset collected for this purpose (i.e., passive-viewing dataset; see above, Materials and Methods). We find that the category model explains unique response variance after accounting for low- and intermediate-level features in 25.7 $\pm$ 1.6% of cortical voxels [mean $\pm$ SEM across five subjects; bootstrap test, $q(FDR) < 0.05$; see Figs. 4, 5], yielding 8613–13,435 voxels in individual subjects (henceforth referred to as the semantic voxels).

Comparison of estimated category responses across search tasks would be justified only if the fit models can accurately predict BOLD responses that were held out during model fitting. To assess prediction performance of the fit category models, we measured average prediction scores across the two search tasks, taken as the product-moment correlation coefficient between the predicted and measured held-out responses (Fig. 2*b*). Category models have high prediction scores (>1 SD above the mean) in 46.9 $\pm$ 0.6% of the semantic voxels. These include many voxels spread across the AON comprising occipitotemporal, parietal, and premotor cortices, as well as voxels in prefrontal and cingulate cortices (Fig. 3).

A previous study provided the first evidence that attention can alter single-voxel category tuning profiles during search for object categories (Çukur et al., 2013). We thus hypothesized that visual search for action categories can also cause changes in voxelwise category tuning. If attentional tuning changes are significant, the category models fit to individual search tasks should yield higher prediction scores than a null model fit by pooling data across the two search tasks. To test this prediction, we compared the prediction scores obtained from the category and null models. We find that the category model significantly outperforms the null model in 46.1 $\pm$ 1.8% of semantic voxels [bootstrap test, $q(FDR) < 0.05$]. Additional control analyses further ensured that these attentional changes cannot be attributed to residual eye movements, head motion, physiological noise, or target-detection biases (see above, Materials and Methods). Together, these results suggest that many cortical voxels in occipitotemporal, parietal, and prefrontal cortices encode high-level category information and that action-based visual search significantly modulates category responses in single voxels.

Visual search warps semantic representation of actions

Previous studies suggest that the human brain represents visual categories by embedding them in a continuous semantic space (Huth et al., 2012). Here, we used linear encoding models to map category features of natural movies onto the recorded BOLD responses in single voxels. The model features, namely actions, are fundamental semantic concepts in both language and vision. The models successfully predict brain activity in cortical voxels, after controlling for lower levels of features (i.e., motion energy and STIP features). Thus, from a quantitative perspective, it could be argued that there is an explicit representation of the semantic categories of actions in the voxel responses (Naselaris et al., 2011). Note that a theoretical characterization of relationships among semantic concepts is difficult. In computational semantics, an empirical approach is adopted instead that is rooted in the distributional hypothesis. This hypothesis states that concepts with similar statistical distributions have similar meanings. Accordingly, co-occurrence statistics of concepts in corpora are used as a proxy metric for similarity of meaning in many methods for learning semantic relationships (Jurafsky and Martin, 2021). Here, to derive a semantic space underlying

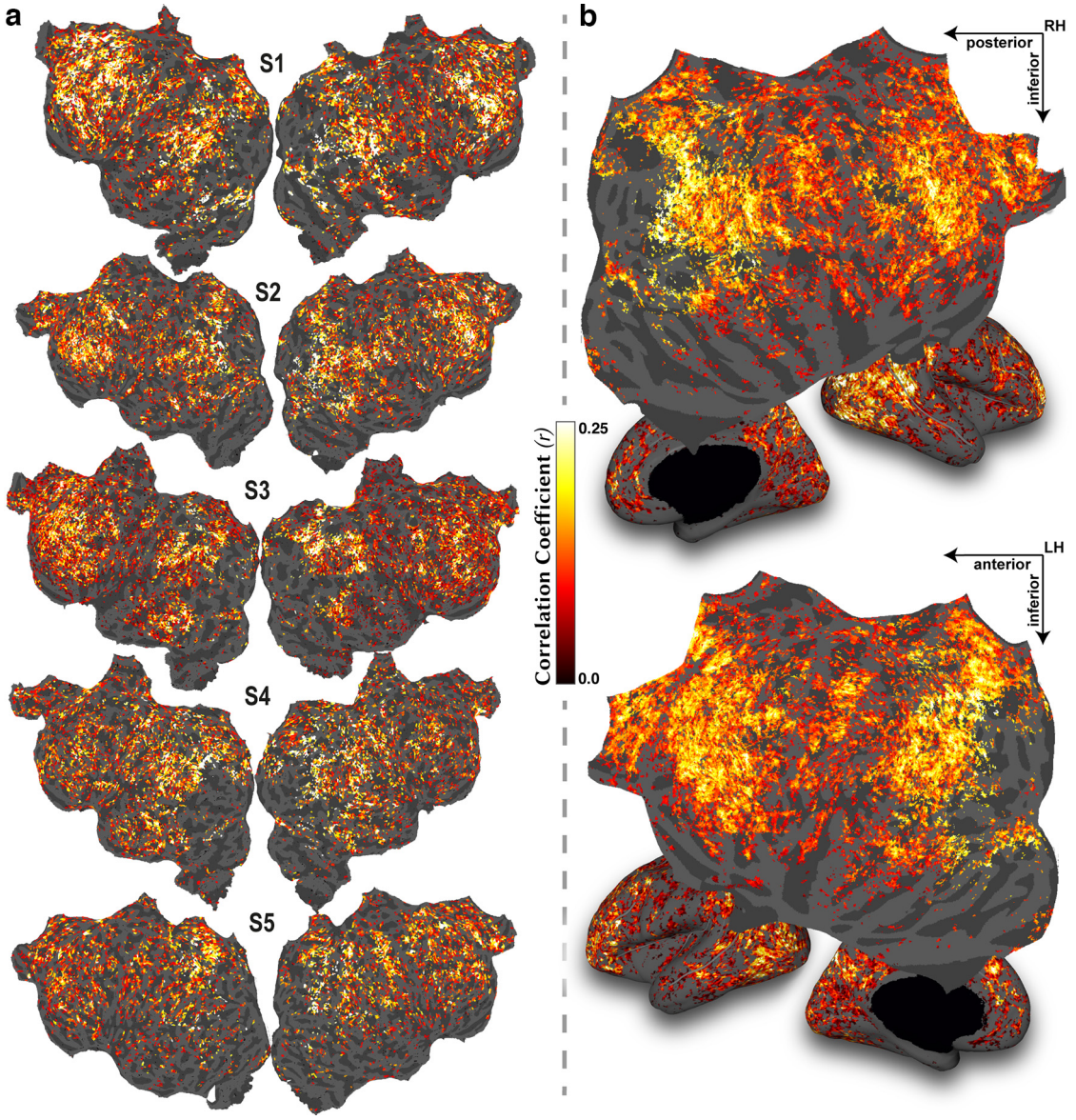

Figure 3. Prediction performance of the category model. To test the performance of fit category models, the prediction score was calculated on held-out data as the product-moment correlation coefficient between the predicted category responses and measured BOLD responses, and it was averaged across the two search tasks. ***a***, Prediction scores of the category model are plotted on flattened cortical surfaces of individual subjects. A variance partitioning analysis was used to quantify the response variance that was uniquely predicted by the category model after accounting for low- and intermediate-level stimulus features (see above, Materials and Methods; Fig. 4). Voxels where the category model did not explain unique response variance after accounting for these features were masked [bootstrap test, $q(\text{FDR}) < 0.05$; Fig. 11]. ***b***, To visualize single-subject results in a common space, prediction score values are shown following projection onto the standard brain template from FreeSurfer and averaging across subjects, after getting thresholded in single subjects. Only voxels that were identified as semantic in all individual subjects were averaged and displayed in the template. Regions of interest are illustrated by white borders. Several important sulci are illustrated by dashed gray lines. (For abbreviations for regions of interest and sulci, see above, Materials and Methods.) The category model predicts responses across ventral-temporal, parietal, and frontal cortices well, suggesting that visual categories are broadly represented across visual and nonvisual cortex. Results can be explored via an interactive brain viewer at http://www.icon.bilkent.edu.tr/brainviewer/shahdloo_etal/.

action category representations, we performed PCA on the model weights for action categories. Visual search for actions alters category model weights as reported here, so performing PCA on data from search tasks can bias estimates of the semantic space. Instead, we derived the semantic space using the passive-viewing dataset. Action categories that are semantically close to each other should project to nearby points in this space, whereas semantically dissimilar categories should project to distant points. The top 12 PCs that explained >95% of the variance in responses were selected, which showed a high degree of intersubject consistency ($r = 0.52 \pm 0.02$ mean ± SEM across subjects; see Fig. 7). To visually examine the semantic information captured by this space, we projected action categories onto the PCs (Fig. 6*a*, see Fig. 9, projections onto the first three dimensions that accounted for 72.8% of the response variance; see Fig. 10, loadings for all PCs). All further quantitative analyses regarding tuning shifts were instead conducted in the full semantic space of 12 dimensions, including all the identified PCs.

Previous evidence suggests that visual search shifts single-voxel tuning profiles to expand the representation of the targets (Çukur et al., 2013). Thus, it is possible that action-based visual search also shifts semantic tuning in single voxels toward the target category. To investigate this possibility, we projected action category responses onto the semantic space. The first and third PCs maximally differentiated between actions belonging to the target categories (i.e., communication vs locomotion categories;

see Fig. 8). Therefore, we visually compared the projections onto these PCs across the two search tasks. We observe that attention causes semantic tuning modulations broadly across cortex (Fig. 6*b*, Extended Data Figs. 6-1, 6-2, 6-3, 6-4, 6-5, results in individual brain spaces). Specifically, voxels in inferior posterior parietal cortex (PPC), cingulate cortex, and anterior inferior prefrontal cortex shift their tuning toward communication during search for communication actions. Meanwhile, voxels in superior PPC and medial parietal cortex shift their tuning toward locomotion during search for locomotion actions. Several reports suggest involvement of superior PPC in representing locomotion actions (Corbo and Orban, 2017) and inferior PPC in representing communication actions (Rizzolatti and Matelli, 2003; Abdollahi et al., 2013). Therefore, our findings suggest that during search for a given action category, tuning shifts toward the target category are most prominent in voxels that are primarily selective for the target.

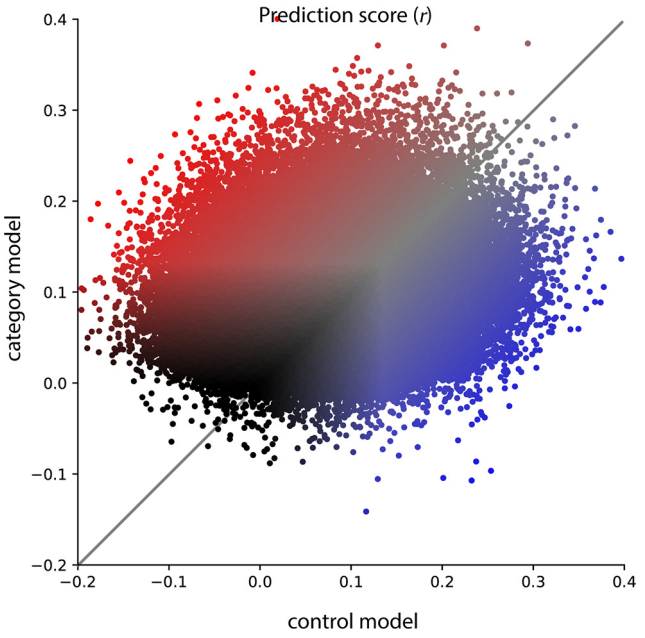

Figure 4. Comparison of category and control models. The prediction scores (raw product-moment correlation coefficient) of the category and control (the collection of motion energy and STIP regressors) models were measured for all cortical voxels. Voxels across all subjects are displayed. Each voxel is represented with a dot. Red versus blue dots indicate whether the category model or the control model yields higher prediction scores. Black dots indicate voxels where none of the models has high prediction scores. The category model outperforms the control model in 53.75 ± 3.29% of cortical voxels (mean ± SEM; average over 5 subjects).

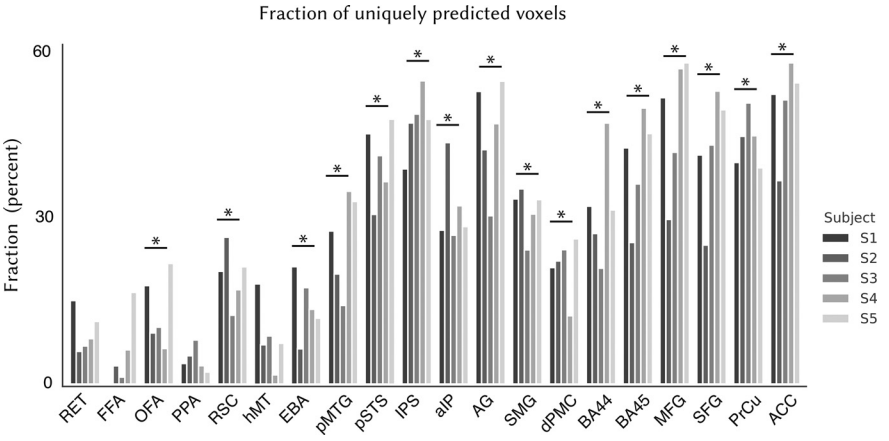

Figure 5. Fraction of uniquely predicted voxels in ROIs. We identified voxels in which the category model explained unique response variance after accounting for low-level motion energy and intermediate-level STIP stimulus features by performing a variance partitioning analysis (see above, Materials and Methods). Fraction of these semantic voxels is shown across ROIs in individual subjects. Asterisk indicates across-subject significance [bootstrap test, $q(FDR) < 0.05$].

Visual search for action categories shifts single-voxel semantic tuning profiles

Our inspection of semantic representations during visual search reveals that attention broadly modulates high-level action representations by shifting semantic tuning profiles in single voxels. To quantify the magnitude and direction of these tuning changes, we separately measured semantic selectivity for communication and locomotion action categories in each search task. The 922-dimensional category responses for individual voxels measured during attention tasks and the idealized template vectors for the targets were projected onto the semantic space. The template vector for a target is constructed as a 109-dimensional indicator vector containing ones for the target category and all its subordinate categories and zeros for the remaining categories. For instance, the locomotion template has ones for locomotion and for walk, run, crawl, move, ride, and so on. As such, the target template vector indexes the target action as well as actions that are semantically related to the target according to the WordNet hierarchy (see above, Materials and Methods). For each attention task, semantic selectivity of a given voxel for a target category was then quantified as the correlation coefficient between projected 12-dimensional vectors characterizing the voxelwise tuning profile and the idealized template in the semantic space. For each voxel, a tuning shift index ($TSI_{all} \in [-1,1]$) was taken as the difference in semantic selectivity for targets when they were attended versus unattended. A positive TSI_{all} indicates shifts toward the target, a negative TSI_{all} indicates shifts away from the target, and a TSI_{all} of 0 suggests no change in between tasks (see above, Materials and Methods).

We find that voxels across many cortical regions shift their tuning toward the attended category (see Fig. 11*a*, 11-1*a*, 11-2*a*, 11-3*a*, 11-4*a*, 11-5*a*, results in individual brain spaces). The respective tuning shifts are shown in relevant ROIs (see Figure 15*a*). Tuning shifts are significantly greater than zero in many areas across AON including occipitotemporal cortex (pSTS, pMTG), posterior parietal cortex (IPS; AG, SMG), and premotor cortex [Brodmann's areas 44, 45, BA44/45; bootstrap test $q(FDR) < 0.05$; see Fig. 15*a*]. This result suggests that focused attention to specific action categories shifts semantic tuning toward targets in single voxels and that these attentional modulations are present at all levels of the AON hierarchy including occipitotemporal cortex.

Prior evidence suggests that during category-based visual search, semantic tuning shifts grow stronger toward later stages of semantic processing (Çukur et al., 2013). Here, we find that semantic tuning shifts in AG and SMG are significantly stronger than those in occipitotemporal (pSTS, pMTG) and premotor cortices (i.e., averaged over AG and SMG, compared with the average over pSTS and pMTG, and with the average over dPMC

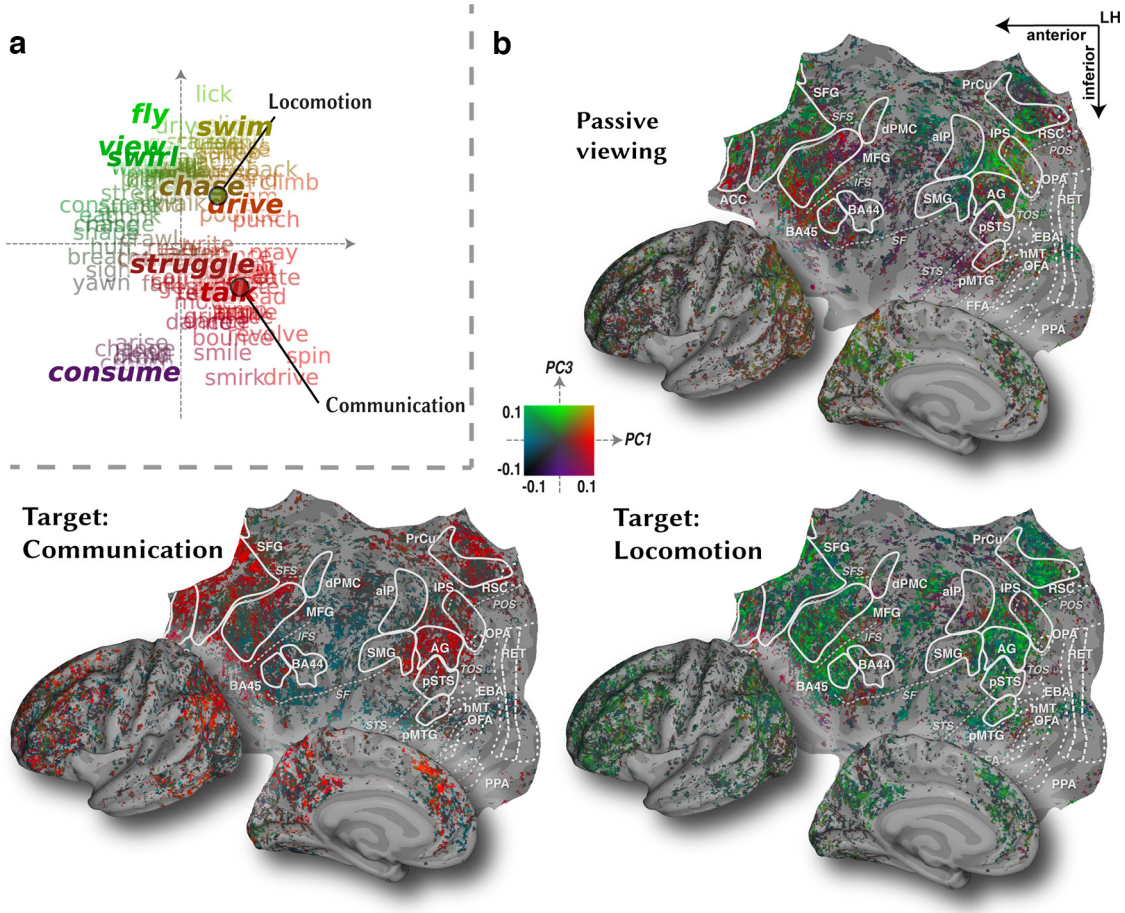

Figure 6. Attention warps semantic representation of action categories. To assess attentional changes, we projected voxelwise tuning profiles onto a continuous semantic space. ***a***, The semantic space was derived from PCA of tuning vectors measured during a separate passive-viewing task and was tested to be consistent across subjects (Fig. 7). To illustrate the semantic information embedded within this space, action categories were projected onto PC1 and PC3 that best delineate the target actions (Fig. 8; words in regular font show projections of individual categories; Fig. 9). To illustrate the semantic content of the PCs, characteristic actions of the movie stimulus were clustered in the semantic space, and cluster centers were projected onto the PCs after getting labeled (bold italic words; see above, Materials and Methods; Fig. 10). Average location of the communication and locomotion actions are indicated with red and green dots. ***b***, Action category responses during passive viewing and during the two search tasks were projected onto the semantic space, and a two-dimensional color map was used to color each voxel based on the projection values along PC1 and PC3 (left, legend). Projections in individual subjects were mapped onto the standard brain template from FreeSurfer, and average projections across subjects are displayed (Extended Data Figs. 6-1–6-5 for data in individual subjects). Figure formatting is identical to that in Figure 3. Many voxels across occipitotemporal, parietal, and prefrontal cortices shift their tuning toward targets, suggesting that attention warps semantic representations of actions. Specifically, voxels in inferior posterior parietal cortex, cingulate cortex, and anterior inferior prefrontal cortex shift their tuning toward communication during search for communication actions. Meanwhile, voxels in superior posterior and medial parietal cortex shift their tuning toward locomotion during search for locomotion actions. Results can be explored via an interactive brain viewer at http://www.icon.bilkent.edu.tr/brainviewer/shahdloo_etal/.

and BA44/45; Cohen's $d = 1.36$, $p < 0.05$). Therefore, the tuning shifts reported here could indicate that AG and SMG are higher nodes in the hierarchy of semantic representation of action categories. In a previous study, we reported that in medial prefrontal cortex, visual search for object categories causes tuning shifts toward targets, whereas it causes tuning shifts away from targets in voxels in PrCu and temporoparietal junction (TPJ; Çukur et al., 2013). Similarly, by qualitative inspection of the flatmaps, here we observe that visual search for action categories causes negative tuning shifts in many voxels across PrCu and TPJ. These results suggest that these areas might be involved in distractor detection and in error monitoring during visual search for actions (Corbetta and Shulman, 2002).

Visual search shifts semantic tuning for nontarget action categories

Natural visual search for object categories was previously suggested to cause changes in representations of not only targets but also nontarget categories (Seidl et al., 2012; Çukur et al., 2013). Thus, it is likely that action-based visual search shifts semantic tuning for nontarget categories. To address this important question, we first examined the separate contributions of tuning changes for target versus nontarget categories to the overall tuning shifts. Specifically, we measured the fraction of overall tuning shifts that can be attributed to the target categories versus nontarget categories (i.e., all categories excluding communication and locomotion actions). We find that both target and nontarget categories significantly contribute to the overall tuning shifts [bootstrap test, $q(FDR) < 0.05$].

However, as would be expected, target categories account for a relatively larger fraction of the overall tuning shifts compared with nontarget categories in all studied ROIs, except in early visual cortex [$q(FDR) < 0.05$; see Fig. 13]. Next, to explicitly quantify tuning shifts for nontarget categories, we calculated a separate tuning shift index exclusively on nontarget categories (TSI_{nt}). To calculate TSI_{nt}, the 109-dimensional action category response vectors were masked to select nontarget categories before projection onto the semantic space (see above Materials and Methods). We observe that tuning shift for nontarget

categories is generally smaller than the overall tuning shift (Fig. 11*b* vs Extended Data Fig. 11*a*, Fig. 12, Extended Data Figs. 11-1*b*, 11-2*b*, 11-3*b*, 11-4*b*, 11-5*b*, results in individual brain spaces). Yet, TSI_{nt} is nonsignificant in all ROIs except AG, SMG, and BA45 [$q(FDR) < 0.05$; see Fig. 15*b*]. Note that an insignificant TSI_{nt} does not necessarily suggest that attention has not altered tuning for nontarget categories, but rather the direction of tuning changes could be merely not aligned toward or away from the target categories in the semantic space. Thus, these results suggest that compared with occipitotemporal areas, attention more diversely warps semantic representations in parietal and premotor AON nodes by shifting tuning for both target and nontarget categories.

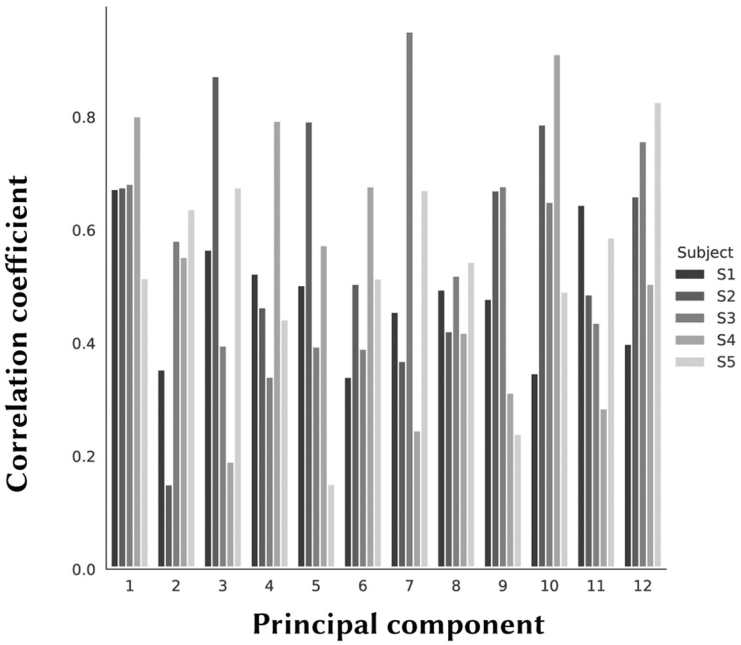

Figure 7. Consistency of the semantic space across subjects. To test whether the estimated semantic space is consistent across subjects, leave-one-out cross-validation was performed. In each cross-validation fold, best-predicted voxels from four subjects were used to derive 12 PCs to construct a semantic space. In the left-out subject, semantic tuning profile for each voxel was obtained by projecting action category responses during passive viewing onto the derived PCs. Next, the product-moment correlation coefficient was calculated between the tuning profiles in the derived space and the tuning profiles in the original semantic space. Results were averaged across semantic voxels in the left-out subject. Correlation coefficients are shown for each PC and each subject. The cross-validated semantic spaces consistently correlate with the original semantic space.

Tuning shifts interact with intrinsic selectivity of cortical voxels for action categories

A study on visual attention has reported that in strongly object-selective regions, voxel tuning for a preferred object might be robust against attention directed to a nonpreferred object (e.g., houses for FFA and faces for PPA; Çukur et al., 2013). This previous result suggests that the degree of response modulations in a brain region might depend on the alignment between the search target and the intrinsically preferred object. It is thus likely that tuning shifts during search for an action category also interact with the intrinsic selectivity of cortical voxels for the target category. Tuning shifts as measured by TSI signal an overall increase in relative selectivity for target versus nontarget categories, aggregated across search tasks. Yet, interaction of tuning shifts with intrinsic selectivity for action categories is task specific by definition. Therefore, to examine potential interactions, we calculated a target preference index ($PI \in [-1,1]$) separately during search for communication actions (PI_{com}) and during search for locomotion actions (PI_{loc}). PI_{com} was taken as the difference in selectivity for communication versus locomotion during search for communication actions. Analogously, PI_{loc} was taken as the difference in selectivity for locomotion versus communication during search for locomotion actions.

Voxelwise PI_{com} and PI_{loc} values were projected onto cortical flatmaps for visual inspection (see Fig. 14; Figs. 11-1*c*, 11-2*c*, 11-3*c*, 11-4*c*, 11-5*c*, results in individual brain spaces) and quantitatively examined in ROIs (see Fig. 15*c*,*d*). We observe that semantic tuning in areas with indiscriminate selectivity for behaviorally relevant action categories (e.g., selective for low-level visual features or static object categories) show insignificant shifts regardless of the search task. Meanwhile, many voxels across anterior parietal, occipital, and cingulate cortices—with intrinsic action category preferences—show differential preference for one of the two target action categories as indicated by a high PI index during either search for communication or search for locomotion actions. Finally, semantic tuning in voxels across posterior parietal and anterior prefrontal cortices with broad selectivity for actions shift toward the attended category regardless of the search target. These specific cases are discussed in detail below.

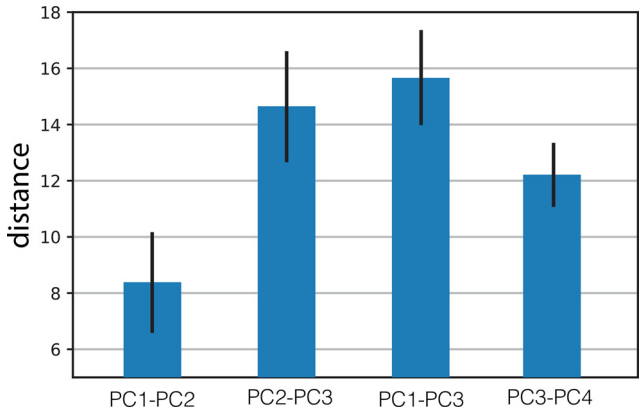

Figure 8. The distance between target actions in subspaces spanned by different pairs of PCs. To visualize attentional modulation of semantic representation in Figure 6, we compared projections of action category responses onto a pair of PCs across the search tasks. To maximize our sensitivity in visualizing the attentional modulations, we chose the pair of dimensions that maximally separates the actions belonging to the two target categories (i.e., communication and locomotion categories). The Mahalanobis distance between communication actions and locomotion actions (mean ± SEM across communication and locomotion actions) in the subspace spanned by each pair of PCs is shown. Target actions are maximally separated across the subspace spanned by the first and third PCs.

Areas where both PI_{com} and PI_{loc} are nonsignificant

We find that PI_{com} and PI_{loc} are nonsignificant in RET [bootstrap test, $q(FDR) > 0.05$] that represent low-level stimulus features, low-level motion-selective area [hMT (human middle temporal); $q(FDR) > 0.05$], and object-selective areas [FFA, OFA, PPA, RSC, and EBA (extrastriate body area); $q(FDR) > 0.05$]. Furthermore, PI_{com} and PI_{loc} are nonsignificant in aIP [$q(FDR) > 0.05$], which is not involved in representing communication or locomotion actions [nonsignificant SI_{com} and SI_{loc}, $q(FDR) > 0.05$; Rizzolatti et al., 1997;

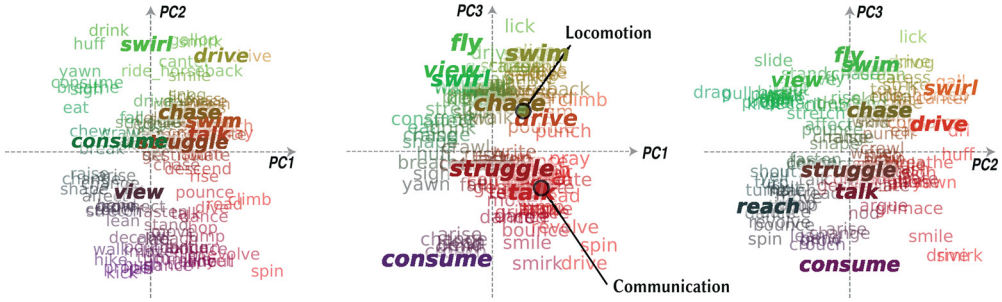

Figure 9. Distribution of action categories across PCs. To illustrate the distribution of action categories embedded within the semantic space, action categories were projected onto the PCs. Projections onto the first three PCs are shown (words in regular font show projections of individual categories). To facilitate illustration, categories were collapsed into 10 clusters, and cluster centers were also projected onto the PCs (bold italic words; see above, Materials and Methods). Average location of the communication and locomotion actions are indicated with red and green dots. The estimated semantic space captures reasonable semantic variance across action categories in natural movies.

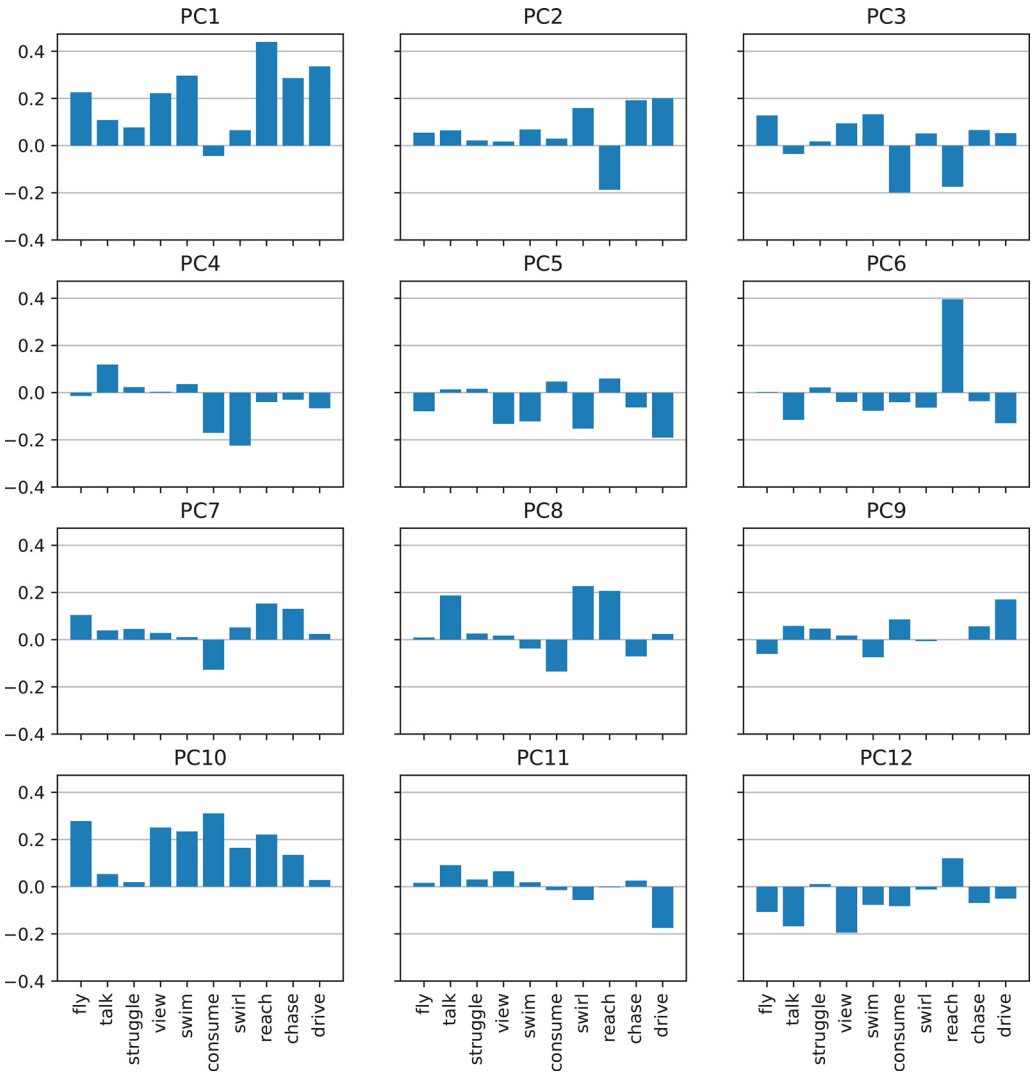

Figure 10. Projections of action category clusters onto PCs. Each of the 109 action categories were projected onto the 12 semantic PCs. The projections were then clustered into 10 groups using *k* means and labeled for interpretation (see above, Materials and Methods). The projections of the cluster centers onto 12 PCs are shown. The first three dimensions were used to visualize the semantic space. The first dimension distinguishes between self-movements (e.g., swirl, consume) and actions that are targeted toward other humans or objects (e.g., reach, talk). The second dimension distinguishes between dynamic (e.g., drive, chase) versus static actions (e.g., consume, struggle). The third dimension distinguishes between actions that involve humans (e.g., talk, reach) and dynamic actions (e.g., fly, swirl).

Noppeney, 2008; Urgen and Orban, 2021]. These results suggest that during an action-based search, semantic tuning does not change substantially in cortical areas that are selective for lower level visual features or for neutral high-level action categories irrelevant to the task.

Areas where either PI_{com} or PI_{loc} are significant

Several previous studies suggest that lateral and medial prefrontal cortices are causally involved in representing communication actions (Van Overwalle, 2009; Wilson-Mendenhall et al., 2013). Here, we find that PI_{loc} is nonsignificant, whereas PI_{com} is

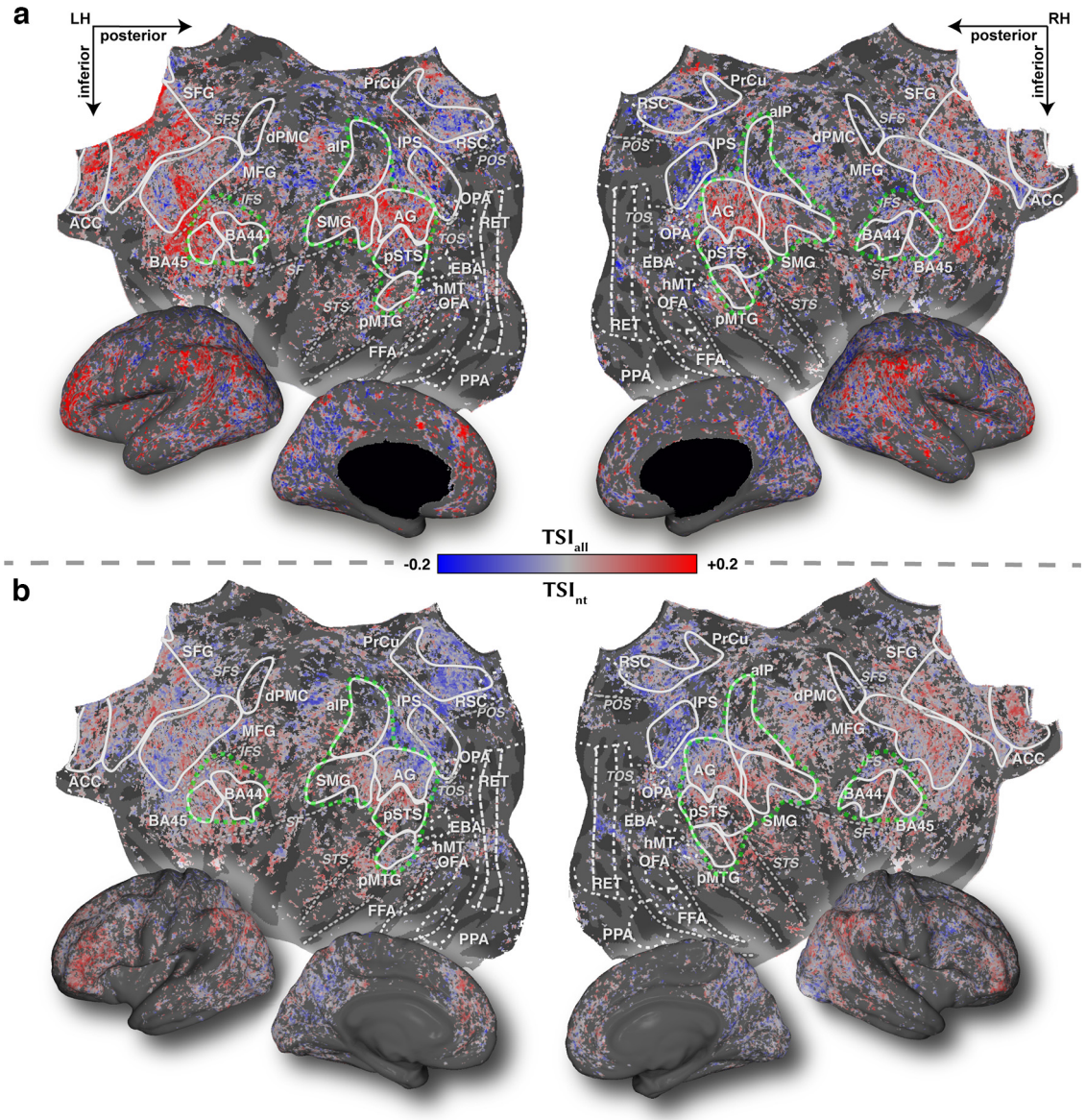

Figure 11. Cortical distribution of tuning shifts. ***a***, To quantify the tuning shifts for the attended versus unattended categories, a tuning shift index ($TSI_{all} \in [-1,1]$) was calculated for each voxel. Tuning shifts toward the attended category would yield positive TSI (red color), whereas negative TSI would indicate shifts away from the attended category (blue color). TSI_{all} values from individual subjects were projected onto the standard brain template and averaged across subjects (Extended Data Figs. 11-1*a*, 11-2*a*, 11-3*a*, 11-4*a*, 11-5*a* for data in individual subjects). Figure formatting is identical to that in Figure 3. AON is outlined by green dashed lines. Voxels across many cortical regions shifted their tuning toward the attended category. These include regions across AON (occipitotemporal cortex, posterior parietal cortex, and premotor cortex), lateral prefrontal cortex, and anterior cingulate cortex. ***b***, To examine how representation of nontarget action categories changes during visual search, we measured a separate tuning shift index specifically for these categories (TSI_{nt}). TSI_{nt} values from individual subjects were projected onto the standard brain template and averaged across subjects (Extended Data Figs. 11-1*b*, 11-2*b*, 11-3*b*, 11-4*b*, 11-5*b* for data in individual subjects). TSI_{nt} shows a similar distribution to TSI_{all} shown in ***a***, albeit with lower magnitude (Fig. 12). Tuning shift for nontarget categories is positive across many voxels within posterior parietal cortex and anterior prefrontal cortex, suggesting a more flexible semantic representation of actions in these cortices, compared with occipitotemporal AON nodes. Results can be explored via an interactive brain viewer at http://www.icon.bilkent.edu.tr/brainviewer/shahdloo_etal/.

significantly greater than zero in anterior inferior frontal gyrus [BA44/45; $d = 1.94$, $q(FDR) < 0.05$; $SI_{com} = 0.12$, $q(FDR) < 0.05$], in SFG [$d = 1.94$, $q(FDR) < 0.05$; $SI_{com} = 0.18$, $q(FDR) < 0.05$)], and in ACC [$d = 0.34$, $q(FDR) < 0.05$; $SI_{com} = 0.18$, $q(FDR) < 0.05$]. On the other hand, previous reports provide evidence for representation of animate locomotion actions in PPC, including IPS [$SI_{loc} = 0.15$, $q(FDR) < 0.05$; Bremmer et al., 2001; Battelli et al., 2003; Ilg et al., 2004; Abdollahi et al., 2013]. In accord, we find that PI_{com} is nonsignificant, whereas PI_{loc} is significantly greater than zero in IPS [$d = 3.95$, $q(FDR) < 0.05$]. Together, our findings suggest that in areas that are strongly selective for specific action categories, visual search for the preferred action shifts tuning more vigorously toward the preferred target category. It is also worth noting that these attentional effects are not limited to the AON but rather extend to higher order cortical areas involved in social cognition. Finally, we find that PI_{loc} is significantly less than zero, whereas PI_{com} is nonsignificant [$d = 0.73$, $q(FDR) > 0.05$; $SI_{loc} = -0.23$, $q(FDR) < 0.05$] in dPMC. This result supports the view that dPMC enhances the representation of distractors during search for locomotion actions (Anticevic et al., 2010; Toepper et al., 2010; Zhou et al., 2012).

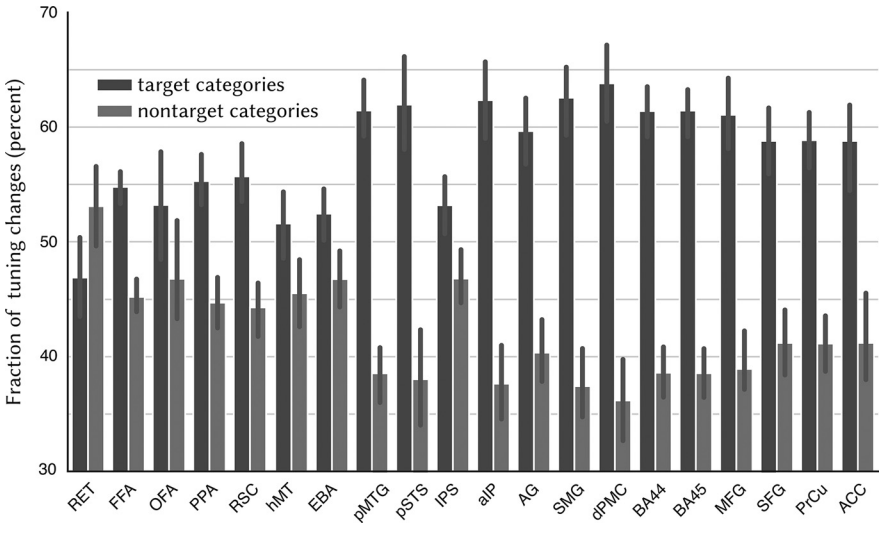

Figure 12. Difference in tuning shift for target, versus nontarget categories. The difference between absolute values of TSI_{all} and TSI_{nt} were calculated in individual ROIs. TSI_{all} is significantly larger than TSI_{nt} in all areas with significant tuning shift.

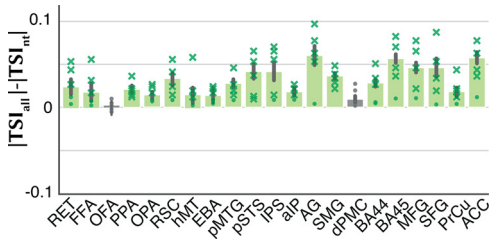

Figure 13. Fraction of the overall tuning shifts. Fraction of the overall tuning shifts explained by shifts in tuning for target categories (mean ± SEM across subjects) and nontarget categories (i.e., excluding the union of communication and locomotion categories) is shown. Target categories explain a greater portion of the overall tuning shifts broadly across ROIs, except for early retinotopic areas. At the same time, nontarget categories significantly contribute to the overall tuning shifts.

Areas where both PI_{com} and PI_{loc} are significant

The pSTS, pMTG, and SMG are considered as AON nodes that maintain representation of actions regardless of their semantic category (Lui et al., 2008; Caspers et al., 2010; Jastorff et al., 2016). We find that both PI_{com} and PI_{loc} are significantly greater than zero in pSTS, pMTG, and SMG, consistent with their generic action selectivity. In addition, several previous studies suggest that MFG, as a node in dorsal attention network, facilitates visual search by maintaining the representation of targets (Corbetta and Shulman, 2002; Mars and Grol, 2007; Paneri and Gregoriou, 2017; Ptak et al., 2017). Accordingly, here we find that PI_{com} and PI_{loc} are significantly greater than zero in MFG [$q(FDR) < 0.05$]. Overall, these results indicate that in areas with generic action selectivity and in high-level cortical areas, attention facilitates an action-based search by shifting representations toward targets regardless of their semantic category.

The results presented here can be explored online via an interactive brain viewer at http://www.icon.bilkent.edu.tr/brainviewer/shahdloo_etal/.

Discussion

Several previous studies have reported response modulations during action-based attention in parietal and prefrontal cortices but not in occipitotemporal areas (Nastase et al., 2017, 2018; Nicholson et al., 2017). Yet we observe significant attentional tuning shifts in occipitotemporal cortex. Unlike previous studies, our analysis approach enables us to measure single-voxel tuning. Our movie stimulus contains a large set of action categories in natural context in contrast to controlled stimuli with a handful of actions on a homogeneous background. Last, we investigate actions that are performed by animate actors, known to elicit robust responses across the occipitotemporal cortex (Thompson and Parasuraman, 2012; Isik et al., 2017; Walbrin et al., 2018; Walbrin and Koldewyn, 2019). These design factors might have enabled us to detect tuning shifts in early stages of AON comprising occipitotemporal areas.

Previous studies emphasize the role of AG and SMG in multimodal semantic representation while observing actions, hearing action sounds, or reading action words (Pizzamiglio et al., 2005; Liljeström et al., 2008; van Dam et al., 2010; Bedny and Caramazza, 2011). Evidence also suggests that during semantic processing these areas act as central connectivity hubs, passing information from low-level perceptual areas onto higher level areas in prefrontal cortex (Hoeren et al., 2013; Farahibozorg et al., 2019). We find significant tuning shifts toward targets in AG and SMG, higher than that in occipitotemporal and premotor AON nodes, regardless of the search target. Our results can be taken to suggest a higher place for AG and SMG in the hierarchy of semantic representations compared with the remaining AON nodes. Another potential account could be that areas with stronger action selectivity might undergo stronger tuning shifts, and future studies are warranted to investigate this issue more directly.

Cortical areas selective for an object category are suggested to retain their preferred tuning even when a nonpreferred category is the search target (Reddy and Kanwisher, 2007; Çukur et al., 2013; Shahdloo et al., 2020). We find that semantic tuning of voxels in locomotion-action-selective superior parietal cortex are shifted toward locomotion actions only during search for this target. Likewise, semantic tuning of voxels in communication-selective anterior prefrontal cortex are shifted toward communication actions only during search for communication. These results suggest that semantic tuning shifts interact with the intrinsic selectivity for target categories.

We used WordNet to label action categories in the stimulus and create a one-hot-encoded stimulus feature matrix. Thus, it is possible to conduct part of the reported analyses by directly examining modulations of category responses. However, assessments in the 922-dimensional category space would treat each category independently ignoring semantic similarities, and they would be inherently noisier, reducing our sensitivity for detecting tuning shifts. To assess TSI for nontarget categories, category responses were masked to zero out responses for communication and locomotion actions. If selectivity measurements had been performed based on one-hot category features, this masking would eliminate all information related to target categories. It would then be impossible to quantify whether tuning for nontarget categories shifts toward/away from the attended category. Therefore, we performed our analyses in a

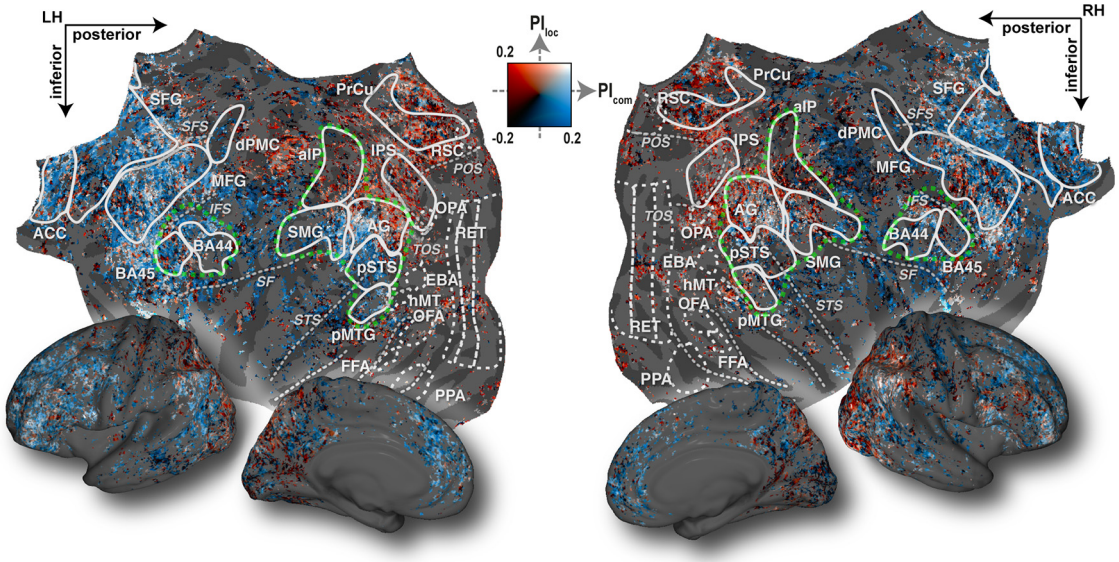

Figure 14. Interaction of tuning shifts with intrinsic selectivity for individual targets. To examine the interaction between tuning shifts and the intrinsic selectivity for individual targets, separate target PIs were calculated during search for communication (PI_{com}), and locomotion (PI_{loc}) categories. PI during search for a specific target action was taken as the difference in selectivity for the target versus distractor during the search for that target. PI_{com} and PI_{loc} values are shown following projection onto the standard brain template (Extended Data Figs. 11-1*c*, 11-2*c*, 11-3*c*, 11-4*c*, 11-5*c* for data in individual subjects). A two-dimensional color map was used to annotate each voxel based on PI_{com} and PI_{loc} values (middle, legend). Figure format is identical to that of Figure 3. AON is outlined by green dashed lines. Semantic tuning in voxels across posterior parietal and anterior prefrontal cortices shift toward the attended category regardless of the search target. However, tuning in many voxels in anterior parietal, occipital, and cingulate cortices shift toward the attended category only during the search for communication or only during the search for locomotion actions.

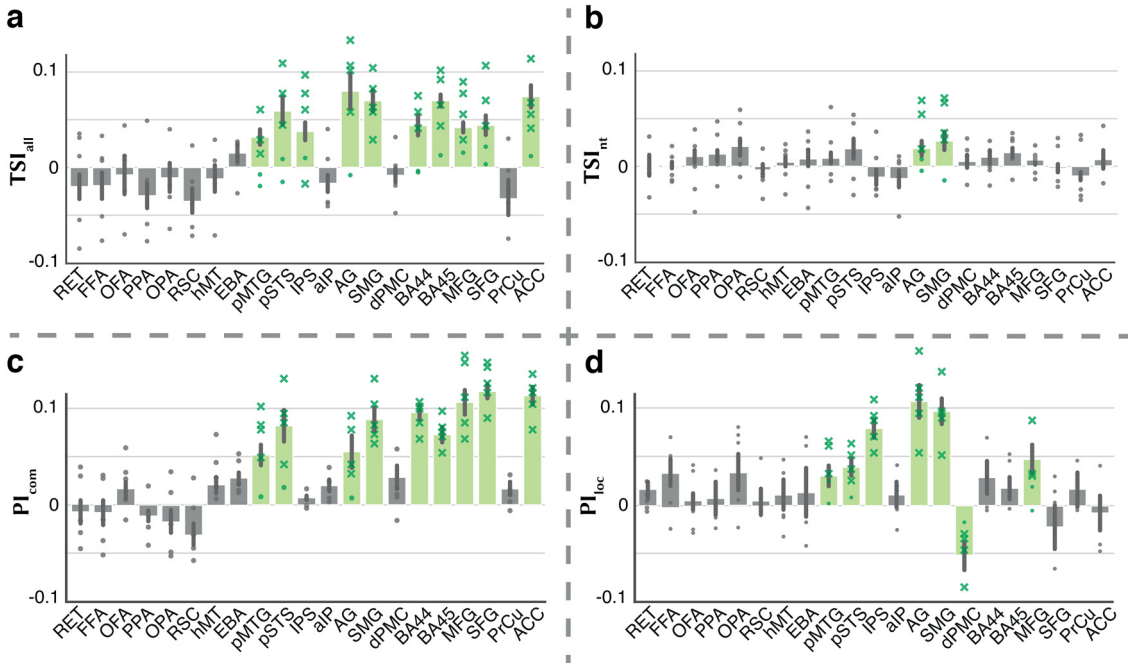

Figure 15. Attentional tuning changes in regions of interest. ***a–d***, Average (***a***) TSI_{all}, (***b***) TSI_{nt}, (***c***) PI_{com}, and (***d***) PI_{loc} values were examined in cortical areas (mean ± SEM across 5 subjects). Significant values are denoted by green bars, and gray bars denote nonsignificant values [bootstrap test, $q(FDR) > 0.05$]. Values for individual subjects are indicated by dots. Gray dots show values in areas with nonsignificant mean, green dots show nonsignificant values in areas with significant mean, and green crosses show significant values in areas with significant mean. Tuning shift is significantly greater than zero in many regions across all levels of the AON including occipitotemporal cortex (pSTS, pMTG), posterior parietal cortex (IPS, AG, SMG), and premotor cortex (BA44, BA45), and in regions across prefrontal and cingulate cortices (SFG, ACC). Compared with occipitotemporal areas, attention more diversely modulates semantic representations in parietal and premotor AON nodes, manifested as significantly positive tuning shift for nontarget categories in posterior parietal cortex (AG, SMG) and anterior inferior frontal cortex (BA45). PI_{com} is significantly greater than zero in BA44/45, SFG, and ACC. In contrast, PI_{loc} is significantly greater than zero in IPS and AG and is significantly less than zero in dPMC. Both PI_{com} and PI_{loc} are significantly greater than zero in pSTS, pMTG, SMG, and MFG. Tuning shifts interact with the attention task and with intrinsic selectivity of cortical areas for target action categories.

dense-encoded semantic space obtained via PCA. An alternative is voxelwise modeling with a dense-encoded stimulus feature matrix derived using embedding models (Mikolov et al., 2013; Devlin et al., 2019). During preliminary experiments in the current study and in prior studies from our lab (Huth et al., 2012; Kiremitçi et al., 2021), we compared the category model against dense embedding models, and estimates of attentional modulations did not vary significantly by choice of model. As such, we do not expect a profound difference between results from these various models, although there could be practical differences in terms of interpretation and feature similarity assessments.

We used communication and locomotion as target categories to maximize our chances for detecting semantic tuning shifts as previous studies suggest that these action categories have broadly distributed and distinctive representations (Urgen and Orban, 2021). Attentional modulations in multivoxel response patterns were reported during search for several other categories related to animal taxonomy or actions (Nastase et al., 2017). We have observed in preliminary experiments that search for many salient categories in natural movies elicits tuning shifts (data not shown). Thus, it is likely that tuning shifts are a ubiquitous mechanism for response modulation during natural visual search for action categories. However, there may be differences in the strength and cortical distribution of tuning shifts depending on the target action, and future studies are warranted to systematically examine whether and how tuning shifts generalize across action categories. Evidence suggests that attending to an object can modulate responses to features correlated with the target (O'Craven et al., 1999). We have previously reported that attending to a target object (e.g., vehicles) enhances the representation of semantically similar actions (e.g., driving; Çukur et al., 2013). It is thus possible that attending to a target action could induce tuning shifts for correlated features such as the object categories pertaining to the actor. Because we restricted the target actions to be performed by the same animate actors, we did not examine tuning changes for objects in this study.

The tuning profile of a voxel refers to its response levels to the examined range of features. Attention can induce different modulations on this profile including baseline changes, gain changes, and tuning shifts. Baseline changes imply an additive offset, gain changes imply a multiplicative offset to responses uniformly across features, with neither changing the shape of the profile. Instead, tuning shifts alter shape by shifting selectivity toward the target, changing responses to both attended and unattended features. Here, we find that the overall tuning shift is attributed to significant tuning changes for both target and nontarget categories. Such broadly distributed changes imply alteration in the shape of the tuning profile. As our measurements are naturally limited by the spatiotemporal resolution of BOLD responses, we cannot make definitive inferences about the neural mechanisms underlying voxel tuning shifts, which could be attributed to baseline, gain, or selectivity changes in single neurons (Connor et al., 1997; Reynolds et al., 2000; David et al., 2008). Further electrophysiological work would be needed to characterize neural tuning shifts during an action-based search.

A common practice in fMRI is to collect a relatively limited dataset from a greater number of subjects to increase reliability of across-subject assessments at the expense of individual subject results. Diverting from this practice, here we collect a larger amount of data per subject to give greater focus to reliability in single subjects. This procedure substantially increased the amount and diversity of fMRI data collected per subject, which enhanced the quality of resulting models and thereby the reliability of individual subject results. However, we acknowledge that future studies are warranted to assess to what degree the results reported in the current study generalize to a broader population of subjects.

The natural movie stimuli used here have greater ecological validity compared with simplified or controlled movie clips used in many action-perception studies. That said, action categories in natural movies might be correlated with low-level features such as global motion-energy (Weiss et al., 2006; Nishimoto et al., 2011) and intermediate-level features such as scene dynamics (Grossman and Blake, 2002). Substantial correlations can confound the estimated category responses and tuning shifts. We used several procedures to control for potential biases. First, to minimize correlations between category responses and global motion-energy, we used a nuisance motion-energy regressor (Nishimoto et al., 2011). Second, we restricted analyses to voxels uniquely predicted by the category model after accounting for motion-energy and STIP features. Voxels in areas such as LOC (lateral occipital complex) might encode multiple levels of features ranging from motion-energy and kinematics to semantics. Thus, controlling for motion-energy and STIP features might reduce sensitivity for attentional modulation of perceptual selectivity in these areas. Our analyses do not consider attentional tuning shifts that might be evident for motion-energy and STIP features or other features such as expected action goals (Hudson et al., 2016a,b), and actors' perceived attitude (Bach and Schenke, 2017). Some level of ambiguity will be naturally evident about what specific aspect of the correlated stimulus features is most relevant for measured cortical representations. Addressing this ambiguity requires complete decorrelation of all possible feature sets, yet conclusions derived using decorrelated stimuli deprived from their natural context might no longer be ecologically relevant. It remains important work to assess the effects of a category-based search on multiple levels of feature representations.

In conclusion, we showed that natural visual search for a specific action category modulates semantic representations, causing tuning shifts toward the target in single voxels within and beyond the AON. Attentional modulations further interact with intrinsic selectivity of neural populations for search targets. This dynamic attentional mechanism can facilitate action perception by efficiently allocating neural resources to accentuate the representation of task-relevant action categories. Overall, these findings offer new insights into the effects of category-based visual search on brain responses (Peelen et al., 2009; Çukur et al., 2013; Harel et al., 2014; Erez and Duncan, 2015) as our results help explain humans' astounding ability to perceive others' actions in dynamic, cluttered daily life experiences.

References

Abdollahi RO, Jastorff J, Orban GA (2013) Common and segregated processing of observed actions in human SPL. Cereb Cortex 23:2734–2753.

Anticevic A, Repovs G, Barch DM (2010) Resisting emotional interference: brain regions facilitating working memory performance during negative distraction. Cogn Affect Behav Neurosci 10:159–173.

Bach P, Schenke KC (2017) Predictive social perception: towards a unifying framework from action observation to person knowledge. Soc Personal Psychol Compass 11:e12312.

Battelli L, Cavanagh P, Thornton IM (2003) Perception of biological motion in parietal patients. Neuropsychologia 41:1808–1816.

Bedny M, Caramazza A (2011) Perception, action, and word meanings in the human brain: the case from action verbs. Ann N Y Acad Sci 1224:81–95.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300.

Bremmer F, Schlack A, Duhamel J-R, Graf W, Fink GR (2001) Space coding in primate posterior parietal cortex. Neuroimage 14:S46–S51.

Caspers S, Zilles K, Laird AR, Eickhoff SB (2010) ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50:1148–1167.

Cavina-Pratesi C, Connolly JD, Monaco S, Figley TD, Milner D, Schenk T, Culham JC (2018) Human neuroimaging reveals the subcomponents of grasping, reaching and pointing actions. Cortex 98:128–148.

Çelik E, Dar SUH, Yılmaz Ö, Keleş Û, Çukur T (2019) Spatially informed voxelwise modeling for naturalistic fMRI experiments. Neuroimage 186:741–757.

Connor CE, Preddie DC, Gallant JL, Van Essen DC (1997) Spatial attention effects in macaque area V4. J Neurosci 17:3201–3214.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Corbo D, Orban GA (2017) Observing others speak or sing activates SPT and neighboring parietal cortex. J Cogn Neurosci 29:1002–1021.

Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–770.

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 1:886–893.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage 9:179–194.

David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59:509–521.

de Lange FP, Spronk M, Willems RM, Toni I, Bekkering H (2008) Complementary systems for understanding action intentions. Curr Biol 18:454–457.

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53:1–15.

Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv. doi:10.48550/arXiv.1810.04805.

Erez Y, Duncan J (2015) Discrimination of visual categories based on behavioral relevance in widespread regions of frontoparietal cortex. J Neurosci 35:12383–12393.

Farahibozorg S-R, Henson RN, Woollams AM, Hauk O (2019) Distinct roles for the anterior temporal lobe and angular gyrus in the spatio-temporal cortical semantic network. Cereb Cortex 30:bhab501.

Ferri S, Rizzolatti G, Orban GA (2015) The organization of the posterior parietal cortex devoted to upper limb actions: an fMRI study. Hum Brain Mapp 36:3845–3866.

Friston KJ, Frith CD, Turner R, Frackowiak RSJ (1995) Characterizing evoked hemodynamics with fMRI. NeuroImage 2:157–165.

Gao JS, Huth AG, Lescroart MD, Gallant JL (2015) PyCortex: an interactive surface visualiser for fMRI. Frontiers in Neuroinformatics 9:162.

Grafton ST, de C Hamilton AF (2007) Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci 26:590–616.

Grossman ED, Blake R (2002) Brain areas active during visual perception of biological motion. Neuron 35:1167–1175.

Handjaras G, Bernardi G, Benuzzi F, Nichelli PF, Pietrini P, Ricciardi E (2015) A topographical organization for action representation in the human brain. Hum Brain Mapp 36:3832–3844.

Harel A, Kravitz DJ, Baker CI (2014) Task context impacts visual object processing differentially across the cortex. Proc Natl Acad Sci U S A 111:E962–71.

Herrington J, Nymberg C, Faja S, Price E, Schultz R (2012) The responsiveness of biological motion processing areas to selective attention towards goals. Neuroimage 63:581–590.

Hoeren M, Kaller CP, Glauche V, Vry M-S, Rijntjes M, Hamzei F, Weiller C (2013) Action semantics and movement characteristics engage distinct processing streams during the observation of tool use. Exp Brain Res 229:243–260.

Holte MB, Moeslund TB, Fihl P (2010) View-invariant gesture recognition using 3D optical flow and harmonic motion context. Comput Vis Image Underst 114:1353–1361.

Hudson M, Nicholson T, Ellis R, Bach P (2016a) I see what you say: prior knowledge of other's goals automatically biases the perception of their actions. Cognition 146:245–250.

Hudson M, Nicholson T, Simpson WA, Ellis R, Bach P (2016b) One step ahead: the perceived kinematics of others' actions are biased toward expected goals. J Exp Psychol Gen 145:1–7.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224.

Ilg UJ, Schumann S, Thier P (2004) Posterior parietal cortex neurons encode target motion in world-centered coordinates. Neuron 43:145–151.

Isik L, Koldewyn K, Beeler D, Kanwisher NG (2017) Perceiving social interactions in the posterior superior temporal sulcus. Proc Natl Acad Sci U S A 114:E9145–E9152.

Jastorff J, Orban GA (2009) Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. J Neurosci 29:7315–7329.

Jastorff J, Begliomini C, Fabbri-Destro M, Rizzolatti G, Orban GA (2010) Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. J Neurophysiol 104:128–140.

Jastorff J, Abdollahi RO, Fasano F, Orban GA (2016) Seeing biological actions in 3D: an fMRI study. Hum Brain Mapp 37:203–219.

Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept Psychophys 14:201–211.

Jurafsky D, Martin J (2021) Speech and language processing. New York: Prentice Hall.

Kiremitçi I, Yilmaz Ö, Çelik E, Shahdloo M, Huth AG, Çukur T (2021) Attentional modulation of hierarchical speech representations in a multitalker environment. Cereb Cortex 31:4986–5005.

Laptev I (2005) On space-time interest points. Int J Comput Vision 64:107–123.

Laptev I, Marszałek M, Schmid C, Rozenfeld B (2008) Learning realistic human actions from movies. Paper presented at 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, June.

Lescroart MD, Gallant JL (2019) Human scene-selective areas represent 3D configurations of surfaces. Neuron 101:178–192.e7.

Liljeström M, Tarkiainen A, Parviainen T, Kujala J, Numminen J, Hiltunen J, Laine M, Salmelin R (2008) Perceiving and naming actions and objects. Neuroimage 41:1132–1141.

Lingnau A, Downing PE (2015) The lateral occipitotemporal cortex in action. Trends Cogn Sci 19:268–277.

Lui F, Buccino G, Duzzi D, Benuzzi F, Crisi G, Baraldi P, Nichelli P, Porro CA, Rizzolatti G (2008) Neural substrates for observing and imagining non-object-directed actions. Soc Neurosci 3:261–275.

Mars RB, Grol MJ (2007) Dorsolateral prefrontal cortex, working memory, and prospective coding for action. J Neurosci 27:1801–1802.

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. ArXiv. doi: 10.48550/arXiv.1301.3781.

Miller GA (1995) WordNet: a lexical database for English. Commun ACM 38:39–41.

Molinari E, Baraldi P, Campanella M, Duzzi D, Nocetti L, Pagnoni G, Porro CA (2013) Human parietofrontal networks related to action observation detected at rest. Cereb Cortex 23:178–186.

Muthukumaraswamy SD, Singh KD (2008) Modulation of the human mirror neuron system during cognitive activity. Psychophysiology 45:896–905.

Muthukumaraswamy SD, Johnson BW, McNair NA (2004) Mu rhythm modulation during observation of an object-directed grasp. Brain Res Cogn Brain Res 19:195–201.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410.

Nastase SA, Connolly AC, Oosterhof NN, Halchenko YO, Guntupalli JS, Di Oleggio Castello MV, Gors J, Gobbini MI, Haxby JV (2017) Attention selectively reshapes the geometry of distributed semantic representation. Cereb Cortex 27:4277–4291.

Nastase SA, Halchenko YO, Connolly AC, Gobbini MI, Haxby JV (2018) Neural responses to naturalistic clips of behaving animals in two different task contexts. Front Neurosci 12:316.

Nicholson T, Roser M, Bach P (2017) Understanding the goals of everyday instrumental actions is primarily linked to object, not motor-kinematic, information: evidence from fMRI. PLoS One 12:e0169700.
Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646.
Noppeney U (2008) The neural systems of tool and action semantics: a perspective from functional imaging. J Physiol Paris 102:40–49.
Nunez-Elizalde AO, Huth AG, Gallant JL (2019) Voxelwise encoding models with non-spherical multivariate normal priors. Neuroimage 197:482–492.
Oberman LM, Pineda JA, Ramachandran VS (2007) The human mirror neuron system: a link between action observation and social skills. Soc Cogn Affect Neurosci 2:62–66.
O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584–587.
Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE (2010) Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. J Neurophysiol 104:1077–1089.
Oosterhof NN, Tipper SP, Downing PE (2012) Viewpoint (in)dependence of action representations: an MVPA study. J Cogn Neurosci 24:975–989.
Oosterhof NN, Tipper SP, Downing PE (2013) Crossmodal and action-specific: neuroimaging the human mirror neuron system. Trends Cogn Sci 17:311–318.
Paneri S, Gregoriou GG (2017) Top-down control of visual attention by the prefrontal cortex. Functional specialization and long-range interactions. Front Neurosci 11:545.
Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature 460:94–97.
Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D'Amico S, Di Russo F (2005) Separate neural systems for processing action- or non-action-related sounds. Neuroimage 24:852–861.
Ptak R, Schnider A, Fellrath J (2017) The dorsal frontoparietal network: a core system for emulated action. Trends Cogn Sci 21:589–599.
Puglisi G, Leonetti A, Landau A, Fornia L, Cerri G, Borroni P (2017) The role of attention in human motor resonance. PLoS One 12:e0177457.
Puglisi G, Leonetti A, Cerri G, Borroni P (2018) Attention and cognitive load modulate motor resonance during action observation. Brain Cogn 128:7–16.
Reddy L, Kanwisher NG (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol 17:2067–2072.
Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. NeuroImage 61:1402–1418.
Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron 26:703–714.
Rizzolatti G, Matelli M (2003) Two different streams form the dorsal visual system: anatomy and functions. Exp Brain Res 153:146–157.
Rizzolatti G, Fogassi L, Gallese V (1997) Parietal cortex: from sight to action. Curr Opin Neurobiol 7:562–567.
Rowe J, Friston K, Frackowiak R, Passingham R (2002) Attention to action: specific modulation of corticocortical interactions in humans. Neuroimage 17:988–998.
Rozzi S, Fogassi L (2017) Neural coding for action execution and action observation in the prefrontal cortex and its role in the organization of socially driven behavior. Front Neurosci 11:492.
Safford AS, Hussey EA, Parasuraman R, Thompson JC (2010) Object-based attentional modulation of biological motion processing: spatiotemporal dynamics using functional magnetic resonance imaging and electroencephalography. J Neurosci 30:9064–9073.
Schuch S, Bayliss AP, Klein C, Tipper SP (2010) Attention modulates motor system activation during action observation: evidence for inhibitory rebound. Exp Brain Res 205:235–249.
Seidl KN, Peelen MV, Kastner S (2012) Neural evidence for distracter suppression during visual search in real-world scenes. J Neurosci 32:11812–11819.
Shahdloo M, Çelik E, Çukur T (2020) Biased competition in semantic representation during natural visual search. Neuroimage 216:116383.
Thompson J, Parasuraman R (2012) Attention, biological motion, and action recognition. Neuroimage 59:4–13.
Thorndike RL (1953) Who belongs in the family? Psychometrika 18:267–276.
Toepper M, Gebhardt H, Beblo T, Thomas C, Driessen M, Bischoff M, Blecker CR, Vaitl D, Sammer G (2010) Functional correlates of distractor suppression during spatial working memory encoding. Neuroscience 165:1244–1253.
Urgen BA, Orban GA (2021) The unique role of parietal cortex in action observation: functional organization for communicative and manipulative actions. Neuroimage 237:118220.
Urgen BA, Pehlivan S, Saygin AP (2019) Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. Neuropsychologia 127:35–47.
van Dam WO, Rueschemeyer S-A, Bekkering H (2010) How specifically are action verbs represented in the neural motor system: an fMRI study. Neuroimage 53:1318–1325.
Van Overwalle F (2009) Social cognition and the brain: a meta-analysis. Hum Brain Mapp 30:829–858.
Verstynen TD, Deshpande V (2011) Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. NeuroImage 55:1633–1644.
Walbrin J, Koldewyn K (2019) Dyadic interaction processing in the posterior temporal cortex. Neuroimage 198:296–302.
Walbrin J, Downing P, Koldewyn K (2018) Neural responses to visually observed social interactions. Neuropsychologia 112:31–39.
Weiss PH, Rahbari NN, Lux S, Pietrzyk U, Noth J, Fink GR (2006) Processing the spatial configuration of complex actions involves right posterior parietal cortex: An fMRI study with clinical implications. Hum Brain Mapp 27:1004–1014.
Wilson-Mendenhall CD, Simmons WK, Martin A, Barsalou LW (2013) Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. J Cogn Neurosci 25:920–935.
Wurm MF, Caramazza A, Lingnau A (2017) Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. J Neurosci 37:562–575.
Zhou X, Katsuki F, Qi X-L, Constantinidis C (2012) Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. J Neurophysiol 108:31–38.