



Biased competition in semantic representation during natural visual search

Mohammad Shahdloo^{a,c,*}, Emin Çelik^{b,c}, Tolga Çukur^{a,b,c,**}



^a Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

^b Neuroscience Program, Sabuncu Brain Research Center, Bilkent University, Ankara, Turkey

^c National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

ARTICLE INFO

Keywords:

fMRI
Voxelwise modeling
Attention
Natural movies
Biased-competition
Semantic representation

ABSTRACT

Humans divide their attention among multiple visual targets in daily life, and visual search can get more difficult as the number of targets increases. The biased competition hypothesis (BC) has been put forth as an explanation for this phenomenon. BC suggests that brain responses during divided attention are a weighted linear combination of the responses during search for each target individually. This combination is assumed to be biased by the intrinsic selectivity of cortical regions. Yet, it is unknown whether attentional modulation of semantic representations are consistent with this hypothesis when viewing cluttered, dynamic natural scenes. Here, we investigated whether BC accounts for semantic representation during natural category-based visual search. Subjects viewed natural movies, and their whole-brain BOLD responses were recorded while they attended to “humans”, “vehicles” (i.e. single-target attention tasks), or “both humans and vehicles” (i.e. divided attention) in separate runs. We computed a voxelwise linearity index to assess whether semantic representation during divided attention can be modeled as a weighted combination of representations during the two single-target attention tasks. We then examined the bias in weights of this linear combination across cortical ROIs. We find that semantic representations of both target and nontarget categories during divided attention are linear to a substantial degree, and that they are biased toward the preferred target in category-selective areas across ventral temporal cortex. Taken together, these results suggest that the biased competition hypothesis is a compelling account for attentional modulation of semantic representations.

1. Introduction

In daily life, humans frequently search for a multitude of objects in their visual environment. Yet, apart from search tasks where targets are particularly salient (e.g. searching for black bears in snow), shared features among targets can render the search task more difficult as the number of targets increases. Psychophysical studies showed that reaction time and error rate systematically increase with growing number of items to be attended (Eckstein et al., 2000; Wolfe, 2012; Reynolds and Chelazzi, 2004; Luck et al., 1997). The biased competition hypothesis (BC) has been proposed to account for this performance decline (Duncan, 1984). BC reasons that the brain has limited representational capacity. Thus, simultaneous search for multiple visual objects should result in a competition among their representations. Moreover, attention should bias this competition in favor of the target (Desimone, 1998), irrespective of whether attention is deployed to a spatial location (Keitel et al., 2012; Kastner et al., 1998), to a visual feature (McMains and Kastner, 2011;

Bichot et al., 2005; Boynton, 2005), or to a visual object (Gentile and Jansma, 2010; Reddy et al., 2009).

Several neuroimaging studies provided evidence for competition among cortical representations of multiple objects across visual cortex in the absence of specific task demands (Kastner et al., 1998; MacEvoy and Epstein, 2009; Gentile and Jansma, 2010; Nagy et al., 2011; Baeck et al., 2013; Jeong and Xu, 2017). Gentile and Jansma (2010) measured average BOLD responses in fusiform face area (FFA) while subjects viewed a single or a pair of face images. Response to a pair of faces was lower than the summation of the responses when each of the faces was presented individually. Similarly, Nagy et al. (2011) presented four equispaced isolated images of faces or noise. The number of face images was increased systematically from zero to four. They reported that responses in FFA and lateral occipital complex (LOC) to multiple faces were lower than the summation of the responses to individual faces. MacEvoy and Epstein (2009) further suggested a linear model of representational competition among objects. The authors presented either a single or a

* Corresponding author. National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, TR-06800, Turkey.

** Corresponding author. Department of Electrical and Electronics Engineering, Room 304, Bilkent University, Ankara, TR-06800, Turkey.

E-mail addresses: shahdloo@ee.bilkent.edu.tr (M. Shahdloo), cukur@ee.bilkent.edu.tr (T. Çukur).

<https://doi.org/10.1016/j.neuroimage.2019.116383>

Received 20 June 2019; Received in revised form 31 October 2019; Accepted 18 November 2019

Available online 27 November 2019

1053-8119/© 2019 Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

pair of isolated images of objects from four categories (shoes, chairs, cars, or brushes). Using multivoxel pattern analysis, they showed that the response pattern in object-selective areas in ventral temporal cortex when subjects viewed pairs of objects can be approximated by the mean of response patterns when they viewed each of the objects in isolation.

Recent studies also provided evidence for top-down influences in BC during divided attention to multiple objects (Reddy et al., 2009; Gentile and Jansma, 2010). Reddy et al. (2009) studied BOLD responses while subjects attended to a single or a pair of object categories among four alternatives (faces, houses, shoes, or cars). A multivoxel pattern analysis in category-selective areas in ventral temporal cortex revealed that the response pattern during divided attention to two object categories was a weighted linear combination of response patterns while attending to individual targets. The authors then reported that in PPA (preferentially responsive to houses) and in FFA (preferentially responsive to faces), combination weights were biased toward the preferred object category. Similarly, Gentile and Jansma (2010) reported that during attention to a single face within a pair of face images, FFA responses were biased toward the responses recorded when the target face was presented in isolation.

Previous studies on BC have provided evidence for competition in representation of isolated static objects, and top-down attentional biases in the competition in favor of the target. Yet, natural scenes are intrinsically dynamic and cluttered with many different objects and actions. It has recently been suggested that thousands of visual categories are embedded in a continuous semantic space across cortex (Huth et al., 2012), and that category-based visual search causes broad modulations in these semantic representations (Çukur et al., 2013). It is currently unknown whether BC can account for modulations in semantic representation during natural visual search for object-action categories.

To address this question, we conducted a functional magnetic resonance imaging (fMRI) experiment (Fig. 1). Five human subjects viewed 69 min of natural movies while performing three separate tasks in different runs: attend to “humans”, attend to “vehicles”, and attend to “humans and vehicles” (i.e., divided attention). Whole-brain BOLD responses were recorded and category responses for 831 objects and actions were estimated separately for each task and each individual subject (Nishimoto et al., 2011). Individual subjects’ semantic spaces were estimated via principal component analysis on the category response profiles. Voxelwise semantic tuning was then assessed by projecting estimated category responses onto individual subjects’ semantic spaces. To test whether the semantic tuning during divided attention can be approximated by a weighted linear combination of the tuning during single-target tasks, ordinary least squares was used among semantic tuning profiles for the three tasks. To reveal the interactions between the attentional bias in semantic tuning and the intrinsic selectivity of brain areas, the semantic tuning distribution during divided attention was regressed onto tuning distributions during the two single-target tasks.

2. Materials and methods

2.1. Subjects

Five healthy adult volunteers (four males, one female) with normal or corrected-to-normal vision participated in this study: S1 (age 32), S2 (age 28), S3 (age 28), S4 (age 28), S5 (age 28). Data were collected at the University of California, Berkeley. The main experiment contained three sessions. Functional localizers were collected in two sessions. The protocols for these experiments were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. Written informed consent was obtained from all subjects before scanning.

2.2. MRI protocols

Data were collected using a 3T Siemens Tim Trio MRI scanner (Siemens Medical Solutions) using 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 msec, water-excitation pulse with flip angle = 70°, voxel size = 2.24 mm × 2.24 mm × 4.13 mm, field of view = 224 mm × 224 mm, 32 axial slices. To construct cortical surfaces, anatomical 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 msec, flip angle = 10°, voxel size = 1 mm × 1 mm × 1 mm, field of view = 256 mm × 212 mm × 256 mm.

2.3. Stimuli

Continuous natural movies were used as the stimulus in the main experiment. Three 7 min 40 s movie segments were compiled from 10 to 20 s movie clips without sound. Movie clips were selected from a wide variety of sources as detailed in Nishimoto et al. (2011). High-definition movie frames were cropped into a square frame and downsampled to 512 × 512 pixels, covering a 24° × 24° field of view. Subjects were directed to fixate on a color square of 0.16° × 0.16° at the center. The color of the fixation dot was changing at 1 Hz to ensure visibility. The stimulus was presented at a rate of 15 Hz using an MRI-compatible projector (Avotec) and a custom-built mirror arrangement. Subjects’ eye positions were monitored at 60 Hz using a custom-built camera system equipped with an infrared source and the ViewPoint EyeTracker software suite (Arrington Research; see *Supplementary Methods*).

2.4. Experiment design

The main experiment was performed in nine runs of 8 min 30 s duration each to maintain relatively low inter-run scan overhead. Three repetitions of 3 distinct movie segments were presented across runs, where each segment lasted 7 min 40 s. The movie segments were

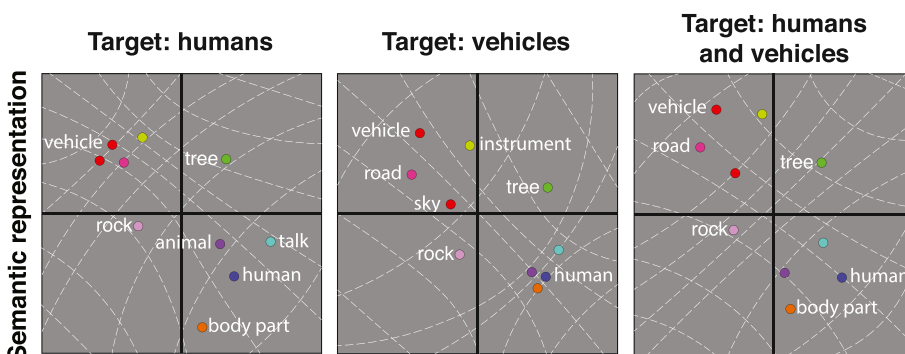


Fig. 1. Hypothesized changes in semantic representation during divided attention. Recent studies have proposed that the human brain represents thousands of object and action categories by embedding them in a low-dimensional space based on their semantic similarity (Huth et al., 2012). It has further been shown that attention warps semantic representation in favor of the target category, and nontarget categories that are semantically similar to the target (Çukur et al., 2013). If the biased competition hypothesis is to mediate semantic representations, the representation during divided attention should be a weighted linear combination of representations during attention to individual targets.

compiled by randomly concatenating 10–20 s short movie clips without sound. Across the entire movie stimulus of 23 min, only humans were present in 303 s, only vehicles were present in 365 s, both categories were present in 280 s, and neither of the two categories were present in 432 s. Subjects were instructed to covertly search for the target categories in the movies. A cue word was displayed before each run to indicate the attention task: “humans”, “vehicles”, or “humans and vehicles”. In the “attend to humans” task, subjects searched for human categories (e.g. woman, man, boy). For the “attend to vehicles” task, subjects searched for vehicle categories (e.g. car, truck, bus). Subjects searched for targets from either of the human or vehicle categories in the divided attention task. Note that while two of the tasks here (i.e. attend to “humans” and attend to “vehicles”) are similar to the tasks in a previous study from our lab (Çukur et al., 2013), the data collected for the attention tasks are independent from that previous study.

Two procedures were performed during the experiments to alleviate potential biases due to subject expectation. First, the clips within each movie segment were ordered so as to counterbalance –within and across runs– the presence of scenes containing only humans, only vehicles, both categories, and neither of the two categories. Second, the order of the attention tasks were counterbalanced across runs (Table 1). Note that this procedure gives more systematic control over a fully randomized ordering, and it maximizes the time between repeated presentation of a given movie clip across different attention tasks. To maintain vigilance, subjects were asked to press a button when they detected a target on the screen. BOLD responses were recorded from the whole brain. To minimize the effect of transient confounds, data from the first 20 s and the last 30 s of each run were discarded. These procedures resulted in 690 data samples for each attention task.

2.5. Data preprocessing

Functional images collected in the main experiment were motion corrected. Using the SPM12 software package (Friston et al., 1995), the functional images were aligned to the first image from the first session of the main experiment. Three subjects that participated in this study were common with a previous study Çukur et al. (2013). Functional images of these subjects were aligned to the reference images from that previous study. Non-brain tissues were removed using the brain extraction tool (BET) from the FSL software package (Smith, 2002). Within each run, low-frequency drifts were removed from BOLD responses in each voxel using a second order Savitzky-Golay filter over a 240 s temporal window. The resulting voxelwise time series were z-scored to attain zero mean and unit variance.

2.6. Model fitting

For each voxel, a linearized model was estimated to relate the movie stimulus to the BOLD responses as detailed in *Supplementary Methods* and in a previous study from our lab (Çukur et al., 2013). First, a feature representation of the movie stimulus was generated by labeling the presence of hundreds of objects and actions in the movies. Then, category models were fit to estimate category response vectors that represented the contribution of each category to single-voxel BOLD responses.

Table 1

Presentation order of the movie segments and order of the search tasks during the main experiment. The three 7 min 40 s movie segments were repeatedly displayed three times and to minimize subject expectation bias, the order of search tasks was interleaved across runs. “H” denotes attend to “humans”, “V” denotes attend to “vehicles”, and “B” denotes attend to “both humans and vehicles” tasks.

Run Number	1	2	3	4	5	6	7	8	9
Attention Task	H	V	B	V	B	H	B	H	V
Movie Segment Number	1	2	3	1	2	3	1	2	3

2.7. Semantic representation of objects and actions

In this study, we question whether semantic tuning during divided attention can be well described as a weighted linear combination of semantic tuning during single-target attention tasks. Semantic tuning profiles for each attention condition were obtained by projecting the respective category responses onto semantic spaces in individual subjects (Huth et al., 2012). Category response vectors for the divided attention task were not included to prevent bias in the estimated semantic space, and subsequently in performance of the weighted linear combination model. The collection of principal components (PCs) that described at least 90% of the variance in response profiles was selected. This resulted in $L \in [36, 47]$ PCs for the five subjects that are each 831-dimensional. Semantic tuning for each voxel was obtained by projecting the category response vector onto these PCs. This projection outputs an L -dimensional vector.

2.8. Linearity of semantic tuning during divided attention

To test whether semantic tuning during divided attention could be predicted using a weighted linear combination of semantic tuning during the two single-target tasks, we compared voxelwise semantic tuning across search tasks. A mask vector, $\bar{m} \in \mathbb{R}^{1 \times 831}$, was used to select the categories of interest among 831 categories (i.e. human categories, vehicle categories, target categories, nontarget categories that are semantically similar to targets, nontarget categories that are semantically dissimilar to targets). Elements of \bar{m} were one for categories of interest and zero elsewhere. Masked response profiles, \bar{w}_i , were obtained by element-wise multiplication

$$\bar{w}_i = \bar{w}_i \odot \bar{m} \quad (1)$$

where \bar{w}_i is the response profile for voxel i and task $t \in \{H, V, B\}$ denoting attend to “humans”, attend to “vehicles”, and attend to “both humans and vehicles”, and \odot represents element-wise multiplication. Masked response profiles were then projected onto the PCs to assess semantic tuning profiles, \bar{S}_i

$$\bar{S}_i = \bar{w}_i T \quad (2)$$

where $T \in \mathbb{R}^{831 \times L}$ is the matrix of L PCs. Semantic tuning profile during divided attention was predicted as a weighted linear combination of semantic tuning during the two single-target tasks using ordinary least-squares. A voxelwise linearity index (LI) was then quantified as the Pearson’s correlation coefficient between measured and predicted semantic tuning during divided attention (\hat{S}_{Bi} ; see Fig. 3a)

$$LI_i = \text{corr}(\bar{S}_{Bi}, \hat{S}_{Bi}) \quad (3)$$

A positive LI indicates that semantic tuning during divided attention can be partly described as a weighted linear combination of the tuning profiles during the two single-target tasks, whereas a negative LI indicates that no significant variance in tuning during divided attention can be explained via tuning during single-target attention tasks. Note that from a model fitting perspective, there are no meaningful differences among negative LI values in $[-1, 0]$. To study the linearity of the semantic representation during divided attention in an ROI, LIs were averaged across voxels with significant prediction scores within the ROI (q (FDR) < 0.05). The assessed LI averages demonstrate the degree to which semantic representation in individual ROIs during divided attention can be explained via representations during attention to each target individually.

2.9. Bias in semantic representation during divided attention

We questioned whether semantic representation during the divided

attention task was biased toward any of the single-target attention tasks. To address this issue, we studied the distribution of semantic representation within an ROI for each individual task. Semantic tuning profiles of significantly predicted voxels within each ROI ($q(\text{FDR}) < 0.05$) were pooled to obtain the distribution of tuning profiles

$$S_t = [\bar{S}_{t1} | \bar{S}_{t2} | \dots | \bar{S}_{tm}], \quad t \in \{H, V, B\} \quad (4)$$

where S_H , S_V , S_B represent distribution of tuning profiles for attend to “humans”, attend to “vehicles”, and attend to “both humans and vehicles” tasks, and n is the number of significantly predicted voxels within the ROI. Note that S_t can also be expressed as

$$S_t = [\bar{P}_{t1}^T | \bar{P}_{t2}^T | \dots | \bar{P}_{tL}^T]^T, \quad t \in \{H, V, B\} \quad (5)$$

where $\bar{P}_{ij} \in \mathbb{R}^{1 \times n}$ is a row vector that represents the projections of the response profiles for task $t \in \{H, V, B\}$ on the j th PC across ROI voxels. To emphasize semantic axes that explain higher variance, projections (\bar{P}_{ij}) were weighted by the proportion of the explained variance of the corresponding PCs. This yielded the semantic tuning distributions, S'_t . The tuning distribution during divided attention was then regressed onto the distributions during the two single-target tasks

$$S'_B = b_H S'_H + b_V S'_V \quad (6)$$

The bias index (BI) was quantified (Fig. 4a) as

$$BI = \frac{b_H - b_V}{|b_H| + |b_V|} \quad (7)$$

Bias toward the “attend to humans” task would yield $BI \in (0, 1]$. A BI of 0 means that the tuning distribution during divided attention is not biased toward any of the single-target attention tasks. Whereas, a BI of 1 means that tuning distribution during divided attention is completely biased toward “attend to humans” task. Similarly, bias toward the “attend to vehicles” task would yield $BI \in [-1, 0)$, where a BI of -1 means that tuning distribution during divided attention is completely biased toward the “attend to vehicles” task. Note that since the response profiles for the three tasks were projected onto the same PCs the calculated BI is immune to changes in the direction of PCs. The assessed BIs in individual ROIs demonstrate the degree and the direction of bias in semantic representation during divided attention toward representations during attention to each target individually.

3. Results

3.1. Representation of categories during visual search

To investigate attentional modulation of semantic representations during attention to multiple targets, we estimated voxelwise category models to measure tuning for hundreds of object and action categories during the three search tasks. We find that the category model accurately predicts responses in many voxels across ventral-temporal, parietal and prefrontal cortices (Fig. 2, see *Supplementary Methods*).

We compared prediction scores of the category model against a null model to assess significance of attentional modulations between the three search tasks: attend to “humans”, attend to “vehicles”, and attend to “both humans and vehicles”. Prediction scores of the null model were obtained by randomly shuffling 24-sec blocks of predicted BOLD responses across the three tasks (see *Materials and Methods*). Cortical voxels with significant differences in prediction score between the original and null models were identified. We find that the category model significantly outperforms the null model in $77.03 \pm 2.49\%$ of cortical voxels (mean \pm sem across five subjects; bootstrap test, $p < 0.05$). Moreover, control analyses revealed that these attentional modulations could not be attributed to nuisance factors including head motion, physiological

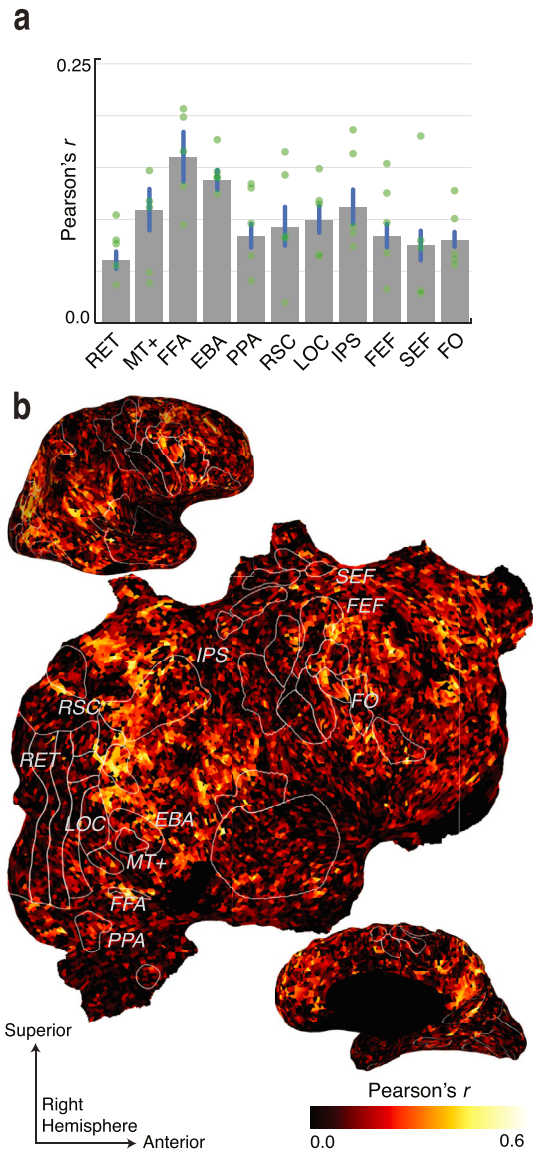


Fig. 2. Prediction performance of the category model. To assess the performance of the category model, BOLD responses were predicted using the estimated category responses in each voxel. Pearson's correlation coefficient between the predicted and measured BOLD responses was taken as the prediction score. (a) Prediction score in functional cortical areas (mean \pm sem across five subjects). Green dots indicate prediction scores for individual subjects. RET, early visual areas V1–3; MT+, human MT; FFA, fusiform face area; EBA, extrastriate body area; PPA, parahippocampal place area; RSC, retrosplenial cortex; LOC, lateral occipital complex; IPS, intraparietal sulcus; FEF, frontal eye fields; SEF, supplementary eye fields; FO, frontal operculum. (b) Cortical flat map of the prediction score for a representative subject. Prediction scores are shown in the right hemisphere. Voxels with high prediction scores appear in yellow color and voxels that have low prediction scores appear in dark gray color (see colorbar). Most voxels in the occipitotemporal cortex, parietal cortex, and prefrontal cortex are well modeled.

noise, and eye movements (see *Supplementary Methods*).

3.2. Linearity of semantic tuning during divided attention

We hypothesized that if the attentional modulations during search for multiple targets are mediated by BC, semantic tuning profile during divided attention should be a weighted linear combination of tuning profiles during isolated attention to individual targets. To test this prediction, we projected estimated category response vectors onto

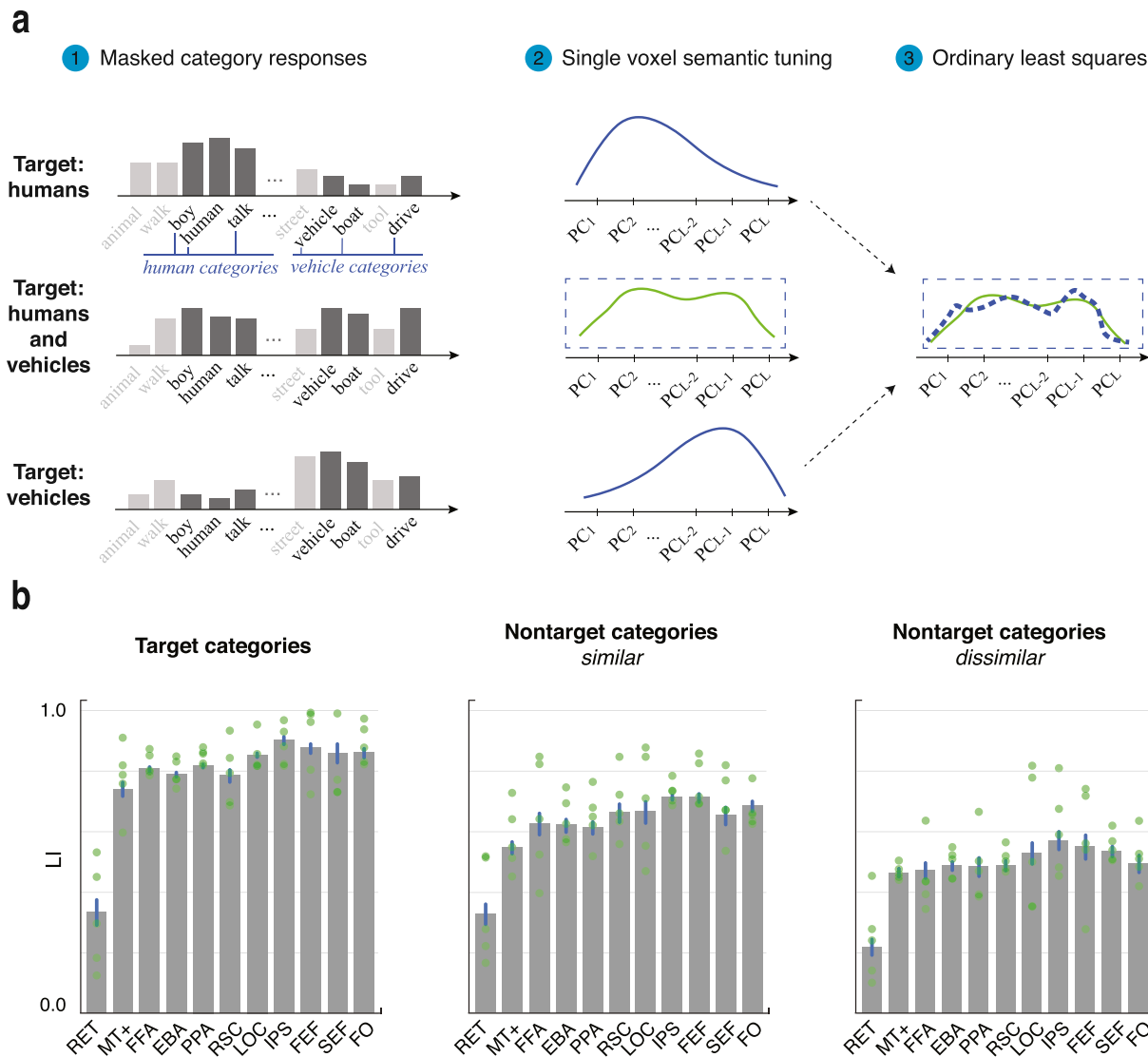


Fig. 3. Linearity of semantic tuning during divided attention. (a) Masked category responses for the three attention tasks were projected onto individual subjects' semantic spaces to estimate the voxelwise semantic tuning profiles. The semantic space in each subject was estimated by performing principal component analysis (PCA) on response profiles pooled during the two single-target attention tasks. The collection of principal components (PCs) explaining at least 90% of the variance in the data was selected. Semantic tuning profile during divided attention was predicted via a weighted linear combination of tuning profiles during the two single-target tasks (dashed lines indicate predicted tuning; solid lines indicate measured tuning). Pearson's correlation coefficient between the predicted and measured semantic tuning during divided attention was taken as the linearity index (LI). (b) LI in functional cortical areas (mean \pm sem across five subjects) for target categories (left), nontarget categories that are semantically similar to targets (middle), and nontarget categories that are dissimilar to targets (right). LIs for individual subjects are indicated by green dots. A substantial portion of semantic tuning during divided attention is described as a weighted linear combination even in the absence of target categories. For all cases, LI is significantly higher in attentional-control areas (IPS, FEF, SEF, and FO; $p = 0.006$ for target categories, $p = 0.012$ for similar nontarget categories, $p = 0.055$ for dissimilar nontarget categories) and in LOC ($p = 0.023$ for target categories, $p = 0.048$ for similar nontarget categories, $p = 0.001$ for dissimilar nontarget categories) than in category-selective areas (FFA, EBA, PPA, and RSC).

individual subjects' semantic spaces. To calculate LI for specific subsets of categories among all 831 categories, prior to projection onto semantic spaces, we masked estimated category responses to include only the categories of interest. Ordinary least-squares was then used to predict voxelwise semantic tuning profile during divided attention as a weighted linear combination of semantic tuning profiles during the two single-target attention tasks. Linearity index (LI) was taken as Pearson's correlation coefficient between the predicted and measured semantic tuning during divided attention (Fig. 3a, see *Materials and Methods*).

We find that LI for target categories (union of human and vehicle categories) is 0.34 ± 0.08 in early visual cortex (RET), 0.76 ± 0.04 in MT+, 0.81 ± 0.01 in category-selective areas (FFA, PPA, EBA, and RSC), 0.86 ± 0.02 in general object-selective area LOC, and 0.86 ± 0.01 in attentional-control areas (IPS, FEF, SEF, and FO; mean \pm sem; Fig. 3b). LI

is significantly greater than zero in all of the studied functional areas (bootstrap test, $p < 10^{-4}$). These results imply that a substantial portion of semantic tuning during divided attention is linearly described as a weighted linear combination of tuning during attention to individual targets. LI is lower in RET ($p < 10^{-4}$), and MT+ ($p = 0.009$) compared to category-selective areas in inferior temporal cortex. This suggests that representations in RET and MT+ that are selective for lower-level visual features are less linear compared to higher-level visual areas in inferior temporal cortex (Kastner and Ungerleider, 2001). In contrast, in LOC, LI is significantly higher than that of category-selective areas ($p = 0.023$). Moreover, LI in attentional-control areas is significantly higher than that of category-selective areas ($p = 0.006$). This result suggests that semantic tuning for target categories better conform to the weighted linear combination model in the general object-selective area and in later stages of

visual processing compared to visual areas that have a strong category preference.

A recent study from our laboratory showed that during category-based attention voxelwise tuning for nontarget categories that are semantically similar to targets shifts toward target categories (Çukur et al., 2013). We thus asked whether BC accounts for semantic representation of nontarget categories. We also asked whether linearity of semantic representation for nontarget categories depends on their semantic similarity to targets. Note that the voxelwise modeling framework allows us to estimate category responses for 831 distinct categories of objects and actions. Thus, by masking 831-dimensional category response vectors to select a desired subset of categories, an LI can be calculated specifically for the given subset. We used separate masks to calculate LI independently for nontarget categories that are semantically similar to targets (i.e. animals and social places for humans; devices and buildings for vehicles), and for nontarget categories that are dissimilar to targets (all categories except the target categories and the semantically similar nontarget categories). Akin to the results that we reported for representation of target categories, LIs for nontarget categories that are semantically similar to targets and for nontarget categories that are dissimilar to targets are significantly greater than zero in all of the studied functional areas (bootstrap test, $p < 10^{-4}$). This result implies that semantic tuning for non-target categories during divided attention is also well explained by a weighted linear combination tuning during single-target tasks. LI for similar categories is 0.64 ± 0.01 in category-selective areas, 0.67 ± 0.05 in LOC, and 0.68 ± 0.02 in attentional-control areas. LI for dissimilar categories is 0.50 ± 0.01 in category-selective areas, 0.54 ± 0.06 in LOC, and 0.55 ± 0.02 in attentional-control areas. Overall, LI for similar categories is higher than that for dissimilar categories in all functional ROIs ($p < 0.002$). This finding suggests that during divided attention to multiple targets, the competition in representation of nontarget visual objects is carried over to objects that are similar to targets (McMains and Kastner, 2010; Beck and Kastner, 2007). Meanwhile, LI is significantly lower in category-selective areas than in LOC ($p = 0.048$ for similar categories, $p = 0.001$ for dissimilar categories), and in attentional-control areas ($p = 0.012$ for similar categories, $p = 0.055$ for dissimilar categories). This result implies that, as for target categories, semantic tuning for nontarget categories better conforms to the weighted linear combination model in LOC and in later stages of visual processing compared to visual areas that have a strong category preference.

3.3. Bias in semantic representation during divided attention

BC observes that inherent selectivity of cortical areas during passive viewing can bias the competition in favor of the preferred target during visual search (Desimone and Duncan, 1995). To study the interactions between the attentional bias in semantic representation and the intrinsic category-selectivity of cortical areas, we expressed the semantic representation during divided attention task as a weighted linear combination of the semantic representations during the two single-target tasks in individual cortical areas. We then investigated whether weights in the weighted linear combination were biased toward any of the single-target attention tasks. Masked response profiles across voxels within the ROI were projected onto individual subjects' semantic spaces to assess the semantic tuning distribution. To calculate BI for specific subsets of categories among all 831 categories, masks were used to select subsets of interest in estimated category response vectors prior to projection onto semantic spaces. We regressed the semantic tuning distribution during divided attention onto distributions during the two single-target tasks. We then quantified a bias index (BI) using the regression weights. According to this index, bias in semantic representation during divided attention was taken to be in the range $[-1, 1]$, where positive and negative values indicate bias toward the "attend to humans" and "attend to vehicles" tasks, respectively (Fig. 4a, see *Materials and Methods*).

We find that BI for target categories is 0.32 ± 0.06 in human-selective

areas (FFA and EBA), and -0.29 ± 0.09 in scene-selective areas (PPA and RSC; mean \pm sem; bootstrap test, $p < 10^{-4}$; Fig. 4b). BI in MT+, which is responsive for animate motion, is 0.28 ± 0.08 ($p = 0.001$). BI is non-significant in attentional-control areas (IPS, FEF, SEF, and FO), in early visual areas (RET), and the general object-selective area LOC ($p > 0.05$). These results suggest that the competition in representation of target categories during divided attention is biased in favor of the preferred target category in cortical areas that are selective for targets. On the contrary, semantic representation is not biased in areas without any specific category preference.

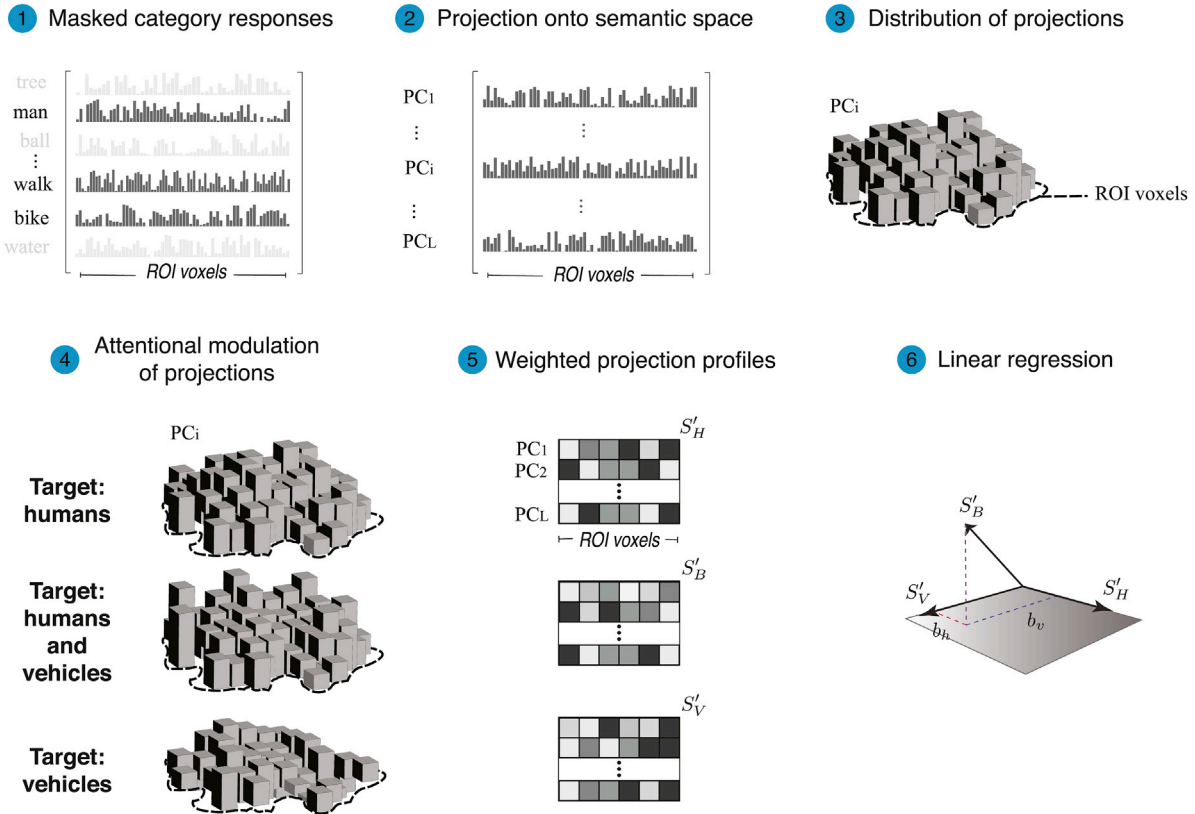
In a previous study we showed that attention shifts semantic tuning for both target and nontarget categories (Çukur et al., 2013). Thus, we asked if there is any bias in representation of nontarget categories during divided attention. To answer this question, we separately calculated BI for nontarget categories that are similar to targets and nontarget categories that are dissimilar to targets. BI for similar categories is 0.38 ± 0.18 in human-selective areas, -0.11 ± 0.27 in scene-selective areas (mean \pm sem; bootstrap test, $p < 10^{-4}$; non-significant in RSC ($p = 0.288$)), and 0.23 ± 0.09 in MT+. BI is non-significant in attentional-control areas, in RET, and in LOC ($p > 0.05$). Meanwhile, BI for dissimilar categories is 0.32 ± 0.26 in human-selective areas, and -0.11 ± 0.19 in scene-selective areas (mean \pm sem; $p < 0.05$; non-significant in EBA ($p = 0.610$) and in RSC ($p = 0.754$)). BI is non-significant in attentional-control areas, in RET, in LOC, and in MT+ ($p > 0.05$). These results indicate that representation of nontarget categories that are similar to targets is biased in favor of the preferred target in areas that are selective for targets. Recent reports on semantic representation of categories in category-selective areas suggest that FFA is strongly selective for human categories, whereas EBA has broader selectivity for categories that are semantically similar to humans (e.g. animals, mammals, animate movement). Similarly, PPA is shown to be strongly selective for vehicles, whereas RSC has broader selectivity for semantically similar categories (e.g. structures, devices, artifacts; Huth et al., 2012; Çukur et al., 2013). Our findings here indicate that representation of nontarget categories that are dissimilar to targets is only biased in areas that are strongly selective for the targets and not in areas that are more broadly tuned for categories that are semantically similar to targets.

The target categories used here (i.e. humans versus vehicles) show high semantic dissimilarity. This raises the possibility that the biases in semantic representation differ between human categories and vehicle categories. To examine this issue, we compared BI for human and vehicle categories separately. We find that BI for human categories is 0.59 ± 0.22 in human-selective areas, and -0.02 ± 0.06 in scene-selective areas (mean \pm sem; bootstrap test, $p < 10^{-4}$ in FFA, EBA; non-significant in PPA ($p = 0.881$), and in RSC ($p = 0.945$); Fig. 5).

Meanwhile, BI for vehicle categories is 0.38 ± 0.21 in human-selective areas, and -0.50 ± 0.01 in scene-selective areas (mean \pm sem; $p < 10^{-4}$; non-significant in EBA ($p = 0.622$)). BI in human-selective areas is significantly positive for both human categories and vehicle categories. Yet, it is stronger for human categories compared to vehicle categories. Whereas in scene-selective areas, bias is significant only for vehicle categories. These results imply that scene-selective areas show greater task dependence in representation of their nonpreferred object categories compared to human-selective areas during divided attention.

In MT+, BI for human categories is 0.21 ± 0.08 , and BI for vehicle categories is non-significant (bootstrap test, $p = 0.341$). A previous study showed that MT+ represents categories that are semantically similar to humans (e.g. body parts and animals) in addition to animate movement (Huth et al., 2012). Thus, this result can be associated with semantic selectivity of MT+ for categories that are similar to humans. In RET, BI for human categories is non-significant ($p = 0.440$), and BI for vehicle categories is -0.19 ± 0.13 . Previous studies suggest that early visual areas are selective for coherent movement and organized contours

a



b

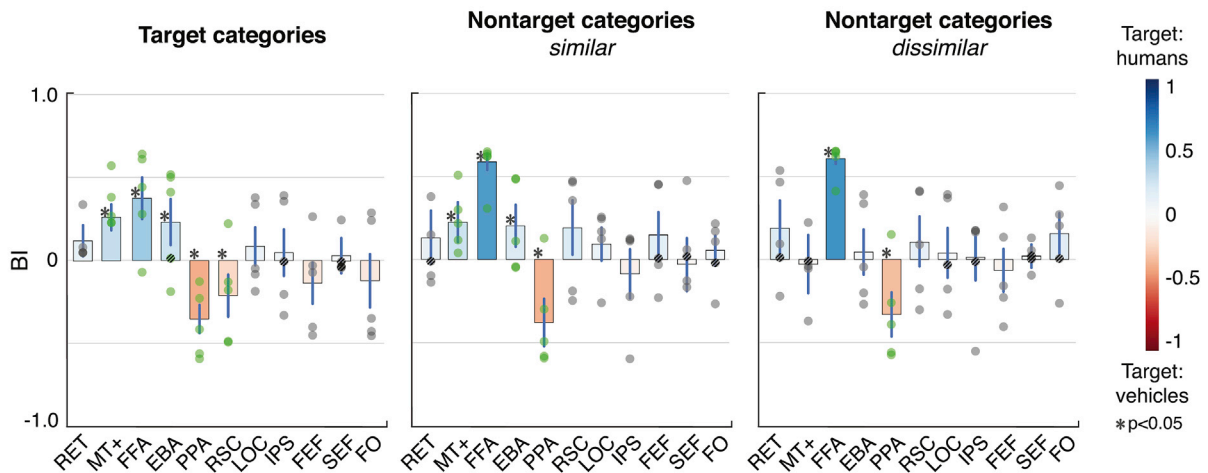


Fig. 4. Bias in semantic representation during divided attention.(a) To assess the bias in semantic representation during divided attention, masked category responses for the three attention tasks were projected onto individual subjects’ semantic spaces. The axes of the semantic space were derived as PCs of category response profiles in the two single-target attention tasks. Projections onto the PCs were weighted by their proportion of explained variance to yield the semantic tuning distributions. Semantic tuning distribution during divided attention was then regressed onto tuning distributions during the two single-target tasks. Regression weights were used to calculate a bias index (BI). (b) BI in functional cortical areas (mean \pm sem across five subjects) for target categories (left), nontarget categories that are similar to targets (middle), and nontarget categories that are dissimilar to targets (right). BIs for individual subjects are indicated by green dots in areas with significant mean values, and by gray dots in areas with non-significant mean values. Hatched dots indicate non-significant BIs (bootstrap test, $p > 0.05$). Blue versus red bars indicate bias toward “attend to humans” versus “attend to vehicles” tasks. Semantic representation of target categories in category-selective areas (FFA, EBA, PPA, and RSC) is biased toward the task in which the preferred object-category is the sole search target. Representation in MT + that encodes animate motion is biased toward the “attend to humans” task. Bias is non-significant in early visual areas (RET), general object-selective area LOC, and attentional-control areas (IPS, FEF, SEF, and FO; $p > 0.05$). Representation of nontarget categories is also biased toward the preferred target in FFA and PPA.

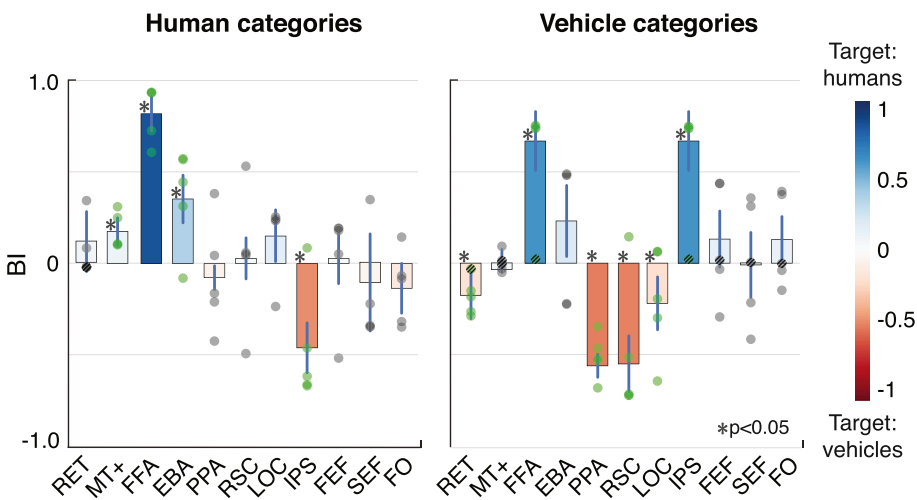


Fig. 5. Bias in representation of human and vehicle categories. BI for human categories (left), and vehicle categories (right) in functional cortical areas (mean \pm sem across five subjects). BIs for individual subjects are indicated by green dots in areas with significant mean values, and by gray dots in areas with non-significant mean values. Hatched dots indicate non-significant BIs (bootstrap test, $p > 0.05$). Representations of both human and vehicle categories in FFA are biased toward the “attend to humans” task. Meanwhile, representation of vehicles but not humans is biased toward the “attend to vehicles” task in PPA ($p = 0.881$). Representations of both of the target categories are not significantly biased in attentional-control areas in prefrontal cortex ($p > 0.05$). Representation in IPS is biased in favor of the distractor category.

(Kourtzi et al., 2003; Altmann et al., 2003). A potential explanation for the observed difference in BI for humans versus vehicles might be related to systematic differences in low-level visual features between the two categories. Vehicles generally possess rigid structure that leads to well-defined contours and an affine motion model (Koller et al., 2005). Thus, negative BI for vehicles in early visual areas might be associated with strong selectivity for coherent motion and organized contours in these areas.

Among the attentional-control areas, BI is significant for both human and vehicle categories only in IPS (bootstrap test, $p < 10^{-4}$). In IPS, BI is -0.48 ± 0.13 for human categories, and 0.60 ± 0.14 for vehicle categories. This result indicates that in IPS, representation of human categories is biased toward “attend to vehicles” task, and representation of vehicle categories is biased toward “attend to humans” task. Previous studies suggest that IPS is part of a cortical network that enhances detection of distractors (Greenberg et al., 2010; Mevorach et al., 2010; Sakai et al., 2002; Bledowski et al., 2004). In line with these studies, our finding here suggests that IPS enhances the representation of distractors during category based visual search.

4. Discussion

In this study, we tested whether the biased competition hypothesis can account for modulation of semantic representations during divided attention. We fit a category model to characterize category responses of single voxels during search for “humans”, “vehicles”, and “both humans and vehicles”. We found that the category model explains significant response variance in many voxels across ventral-temporal, parietal, and prefrontal cortices. We estimated the semantic space underlying category models, and then assessed semantic representations by projecting the estimated category responses for the three search tasks onto the semantic space.

4.1. Linearity of the semantic representation during divided attention

We find that a large portion of the variance in semantic tuning during divided attention can be explained using a weighted linear combination of tuning during isolated attention to individual targets. We find that semantic tuning for target categories is more accurately predicted via the weighted linear combination model compared to semantic tuning for nontarget categories. In a recent study, we reported that attention shifts semantic tuning for target categories to a higher degree compared to that for nontarget categories (Çukur et al., 2013). Thus, our results can be attributed to the higher degree of attentional tuning shift for target categories compared to that for nontarget categories.

Several previous studies have investigated differences in the level of

competition between strongly category-selective areas and areas without a specific category preference. Reddy and Kanwisher (2007) and MacEvoy and Epstein (2009) showed that response patterns to a pair of objects can be better predicted by a linear combination of responses to constituent objects in LOC compared to in FFA or PPA. In line with these studies, here we find that semantic representation better conforms to the weighted linear combination model in LOC compared to that in category-selective areas. These results raise the possibility that semantic representation may also be more linearly additive in attentional-control areas which are not selective for any specific categories compared to strongly category-selective areas (Çukur et al., 2013). Here we find that semantic representations of target and nontarget categories during divided attention are better explained using the weighted linear combination of representations during search for individual targets in attentional-control areas compared to category-selective areas. Our results suggest that higher order areas that are not tuned for specific categories have a more flexible representation of natural scenes during divided attention. Previous studies suggest that the level of representational competition increases from early visual cortex toward higher-level visual areas in inferior temporal cortex (Kastner and Ungerleider, 2001). Consistent with these reports, here we find that semantic representations are less linear in early visual cortex and area MT + compared to higher-level visual areas in occipitotemporal and prefrontal cortices.

4.2. Bias in semantic representation during divided attention

4.2.1. Category-selective areas

The bias in the competition among representation of multiple objects have been investigated in several previous studies. Reddy et al. (2009) reported that BOLD responses to multiple static images of faces, houses, shoes, or cars in category-selective areas, FFA and PPA, are biased toward the preferred target. In line with this finding, here we find that semantic representation of target categories during divided attention in category selective areas FFA, EBA, RSC, and PPA is biased toward the preferred category.

Moreover, we find that semantic representations of nontarget categories that are similar to targets in FFA and PPA are also biased toward the preferred target. Furthermore, representation of nontarget categories that are dissimilar to targets is biased in FFA and PPA, albeit to a lower degree. In a previous study (Çukur et al., 2013), we reported in single voxels within FFA and PPA that attention increases selectivity for nontarget categories that are semantically similar to targets, whereas it decreases selectivity for semantically-dissimilar nontarget categories. For object-selective ROIs, this finding implies that attention modulates representation of a category depending on its semantic similarity to the target. Yet, Çukur et al. (2013) did not examine distributed semantic

representations across these ROIs, and it did not question whether attentional modulations depend on the interaction between the target category and the preferred object category of an ROI (e.g., “humans” in FFA and “vehicles” in PPA). Here we find that attention biases distributed representation of nontarget categories across FFA and PPA toward the preferred target (i.e., when the target matches the preferred category). We also find that these attentional biases are stronger for semantically-similar versus dissimilar nontarget categories. Taken together, these results suggest that semantic similarity to preferred targets enhances the level of bias in semantic representation across category-selective areas.

In human-selective areas, we find that the semantic representations of target categories (both humans and vehicles) during divided attention is biased toward the representation during the “attend to humans” task. Meanwhile, in vehicle-selective areas, the representation of vehicles but not humans is biased toward the representation during the “attend to vehicles” task. This suggests a differential role for human- and vehicle-selective areas in representing nonpreferred targets during divided attention. A potential explanation of this result is that vehicle-selective areas show greater task dependence in representation of nonpreferred targets compared to human-selective areas (Grill-Spector et al., 2004).

4.2.2. Attentional-control areas

We did not observe bias in semantic representation of targets toward any of the target categories in the prefrontal areas that are considered to be part of the attentional-control network. This is expected considering the lack of tuning to specific categories in these areas (Huth et al., 2012). In IPS, we find that semantic representation of humans during divided attention is biased toward the representation during the “attend to vehicles” task and that the representation of vehicles during divided attention is biased toward the representation during the “attend to humans” task. Several previous studies suggest that areas in parietal cortex including IPS enhance visual search by maintaining the representation of distractors (Mevorach et al., 2010; Bledowski et al., 2004), in addition to spatial guidance of attention toward targets (Ptak, 2012; Preston et al., 2013). Consistent with this hypothesis, our results can be interpreted to imply that IPS facilitates natural visual search by maintaining representations of distractor categories.

4.3. Limitations and future work

Category-based visual search in natural scenes is a complex, high-level task that likely involves multiple cognitive processes. Most previous studies on this topic have examined search on static natural images, where the primary component of target detection is spatial visual search. These studies reported that attention biases multi-voxel response patterns in ventral-temporal cortex toward the response patterns for target objects viewed in isolation (Martin, 2007; Peelen et al., 2009; Peelen and Kastner, 2014). Our results here provide further evidence for similar top-down biases in semantic representation during category-based visual search in dynamic natural scenes. Note, however, that visual search on natural movies involves search through both space and time due to the dynamic nature of the stimulus. This dynamic search task might additionally involve preliminary processes associated with top-down modulation of cortical representations (Peelen and Kastner, 2011), as well as subsequent processes associated with working memory maintenance of the search template (Downing, 2000; Moores et al., 2003; Downing and Dodds, 2004). Further work is needed to functionally dissociate potential contributions of these processes to representational biases during natural visual search.

Natural stimuli contain correlations among various levels of features (Hamilton and Huth, 2018). In particular, there might be correlations among low-level visual features of natural scenes and object categories within these scenes (Lescroart et al., 2015). Such correlations can then bias the category responses that we estimated here, which can lead to a biased assessment of semantic

representations. To minimize correlations between category features and global motion-energy of the movie clips, we used a motion-energy regressor in our modeling procedure. However, we do not rule out the possibility that there might be residual correlations between category features and intermediate features of the movie stimuli, such as object shape characteristics (Op de Beeck et al., 2008) and scene layout (Mullally and Maguire, 2011). Note that it is perhaps impossible to create natural stimuli completely free from these correlations. However, future studies may mitigate this problem by compiling more controlled natural stimuli that minimize stimulus correlations while maintaining high variance in individual categories.

Here, we employed category-based search tasks with relatively distinct target object categories: “humans” and “vehicles”. We found that semantic tuning profile during divided attention to two targets can be well explained via a weighted linear combination of semantic tuning profiles during isolated attention to individual targets. Thus, the biased-competition hypothesis gives a compelling account for the results outlined here. Yet, it is possible that nonlinear response modulations are manifested with growing competition among search targets (e.g., increased feature similarity among target categories, higher number of targets). In such cases, alternative competition models can be evaluated. A well-known theoretical account for representational competition in vision is the normalization model of attention (Reynolds and Heeger, 2009; Heeger, 1992). The normalization model observes that single neurons have limited representational capacity for visual stimuli. It proposes that in conditions where competing stimuli can not drive neurons to their saturation point (e.g., gratings with low contrast levels), response to a pair of stimuli is a weighted average of responses to individual stimuli. In saturation conditions, however, response to a pair of stimuli is subject to a nonlinear normalization, where the normalization factor depends on the pooled response across neighboring neurons. An alternative model advocated for object representation is the hierarchical HMAX model (Riesenhuber and Poggio, 1999; Lee et al., 1999; Roos et al., 2014). The HMAX model does not explicitly condition on saturation conditions or lateral interactions with neighboring neurons. Instead, MAX pooling operators are used across afferent responses to determine postsynaptic responses at each layer. This pooling procedure can cause downstream neurons to be selectively responsive to a specific object category, suggesting a winner-takes-all framework for representational competition.

5. Conclusion

In summary, we present evidence for biased competition during visual search, that affects not only individual object representations but also high-level semantic representations. This competition is evident among target categories as well as nontarget categories within natural scenes. Yet, the linearity and bias in representation of nontargets depend on their semantic similarity to targets. Overall, these results help explain the human ability to perform concurrent search for multiple targets in complex visual environments.

Declaration of competing interest

The authors declare no competing financial interests.

Acknowledgement

The authors gratefully acknowledge Salman U.H. Dar for his assistance in writing the manuscript, and Kübra Keskin for her assistance in the analyses. This work was supported in part by a Marie Curie Actions Career Integration Grant (PCIG13-GA-2013-618101), by a European Molecular Biology Organization Installation Grant (IG 3028), by a TUBA GEBIP 2015 fellowship, and by a BAGEP 2017 award.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116383>.

References

- Altmann, C.F., Bühlhoff, H.H., Kourtzi, Z., 2003. Perceptual organization of local elements into global shapes in the human visual cortex. *Curr. Biol.* 13, 342–349.
- Baeck, A., Wagemans, J., Op de Beeck, H.P., 2013. The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *Neuroimage* 70, 37–47.
- Beck, D.M., Kastner, S., 2007. Stimulus similarity modulates competitive interactions in human visual cortex. *J. Vis.* 7, 1–12, 19.
- Bichot, N.P., Rossi, A.F., Desimone, R., 2005. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308, 529–534.
- Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F.E., Linden, D.E.J., 2004. Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage* 22, 530–540.
- Boynton, G.M., 2005. Attention and visual perception. *Curr. Opin. Neurobiol.* 15, 465–469.
- Çukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* 16, 763–770.
- Desimone, R., 1998. Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Biol. Sci.* 353, 1245–1255.
- Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.
- Downing, P.E., 2000. Interactions between visual working memory and selective attention. *Psychol. Sci.* 11, 467–473.
- Downing, P.E., Dodds, C.M., 2004. Competition in visual working memory for control of search. *Vis. Cogn.* 11, 689–703.
- Duncan, J., 1984. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* 113, 501–517.
- Eckstein, M.P., Thomas, J.P., Palmer, J., Shimozaki, S.S., 2000. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept. Psychophys.* 62, 425–451.
- Friston, K.J., Frith, C.D., Turner, R., Frackowiak, R.S.J., 1995. Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2, 157–165.
- Gentile, F., Jansma, B.M., 2010. Neural competition through visual similarity in face selection. *Brain Res.* 1351, 172–184.
- Greenberg, A.S., Esterman, M., Wilson, D., Serences, J.T., Yantis, S., 2010. Control of spatial and feature-based attention in frontoparietal cortex. *J. Neurosci.* 30, 14330–14339.
- Grill-Spector, K., Knouf, N., Kanwisher, N.G., 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* 7, 555–562.
- Hamilton, L.S., Huth, A.G., 2018. The revolution will not be controlled: natural stimuli in speech neuroscience. *Lang. Cognit. Neurosci.* 27, 1–10.
- Heeger, D.J., 1992. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197.
- Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210–1224.
- Jeong, S.K., Xu, Y., 2017. Task-context-dependent linear representation of multiple visual objects in human parietal cortex. *J. Cogn. Neurosci.* 29, 1778–1789.
- Kastner, S., De Weerd, P., Desimone, R., Ungerleider, L.G., 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282, 108–111.
- Kastner, S., Ungerleider, L.G., 2001. The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39, 1263–1276.
- Keitel, C., Andersen, S.K., Quigley, C., Müller, M.M., 2012. Independent effects of attentional gain control and competitive interactions on visual stimulus processing. *Cerebr. Cortex* 23, 940–946.
- Koller, D., Weber, J., Malik, J., 2005. Robust Multiple Car Tracking with Occlusion Reasoning in *Lecture Notes In Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 189–196.
- Kourtzi, Z., Tolias, A.S., Altmann, C.F., Augath, M., Logothetis, N.K., 2003. Integration of local features into global shapes: monkey and human fMRI studies. *Neuron* 37, 333–346.
- Lee, D.K., Itti, L., Koch, C., Braun, J., 1999. Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* 2, 375–381.
- Lescroart, M.D., Stansbury, D.E., Gallant, J.L., 2015. Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Front. Comput. Neurosci.* 9, 1–20.
- Luck, S.J., Chelazzi, L., Hillyard, S.A., Desimone, R., 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42.
- MacEvoy, S.P., Epstein, R.A., 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* 19, 943–947.
- Martin, A., 2007. The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.
- McMains, S., Kastner, S., 2011. Interactions of top-down and bottom-up mechanisms in human visual cortex. *J. Neurosci.* 31, 587–597.
- McMains, S.A., Kastner, S., 2010. Defining the units of competition: influences of perceptual organization on competitive interactions in human visual cortex. *J. Cogn. Neurosci.* 22, 2417–2426.
- Mevorach, C., Hodsoll, J., Allen, H., Shalev, L., Humphreys, G., 2010. Ignoring the elephant in the room: a neural circuit to downregulate salience. *J. Neurosci.* 30, 6072–6079.
- Moore, E., Laiti, L., Chelazzi, L., 2003. Associative knowledge controls deployment of visual selective attention. *Nat. Neurosci.* 6, 182–189.
- Mullally, S.L., Maguire, E.A., 2011. A new role for the parahippocampal cortex in representing space. *J. Neurosci.* 31, 7441–7449.
- Nagy, K., Greenlee, M.W., Kovács, G., 2011. Sensory competition in the face processing areas of the human brain. *PLoS One* 6 e24450.
- Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646.
- Op de Beeck, H.P., Haushofer, J., Kanwisher, N.G., 2008. Interpreting fMRI data: maps, modules and dimensions. *Nat. Rev. Neurosci.* 9, 123–135.
- Peelen, M.V., Fei-Fei, L., Kastner, S., 2009. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460, 94–97.
- Peelen, M.V., Kastner, S., 2011. A neural basis for real-world visual search in human occipitotemporal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12125–12130.
- Peelen, M.V., Kastner, S., 2014. Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250.
- Preston, T.J., Guo, F., Das, K., Giesbrecht, B., Eckstein, M.P., 2013. Neural representations of contextual guidance in visual search of real-world scenes. *J. Neurosci.* 33, 7846–7855.
- Ptak, R., 2012. The frontoparietal attention network of the human brain. *The Neuroscientist* 18, 502–515.
- Reddy, L., Kanwisher, N.G., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr. Biol.* 17, 2067–2072.
- Reddy, L., Kanwisher, N.G., VanRullen, R., 2009. Attention and biased competition in multi-voxel object representations. *Proc. Natl. Acad. Sci.* 106, 21447–21452.
- Reynolds, J.H., Chelazzi, L., 2004. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27, 611–647.
- Reynolds, J.H., Heeger, D.J., 2009. The normalization model of attention. *Neuron* 61, 168–185.
- Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Roos, M.J., Wolmetz, M., Chevillet, M.A., 2014. A hierarchical model of vision (HMAX) can also recognize speech. *BMC Neurosci.* 15, 1019.
- Sakai, K., Rowe, J.B., Passingham, R.E., 2002. Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nat. Neurosci.* 5, 479–484.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Wolfe, J.M., 2012. Saved by a log: how do humans perform hybrid visual and memory search? *Psychol. Sci.* 23, 698–703.