Supplementary Information for:

Attention During Natural Vision Warps Semantic Representation Across the Human Brain

Tolga Çukur[a], Shinji Nishimoto[a,b], Alexander G. Huth[a], Jack L. Gallant[a,c,d]

[a]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA
[b]Present address: Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita, Osaka, Japan
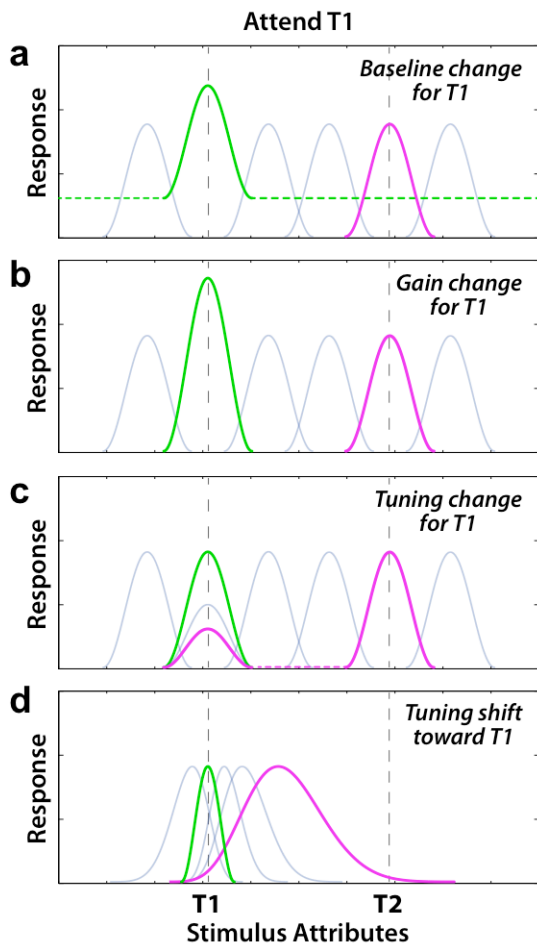[c]Program in Bioengineering, University of California, Berkeley, CA 94720, USA
[d]Department of Psychology, University of California, Berkeley, CA 94720, USA
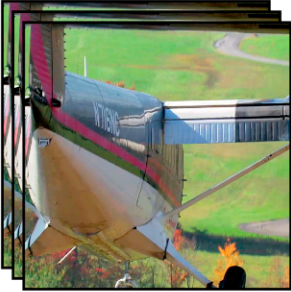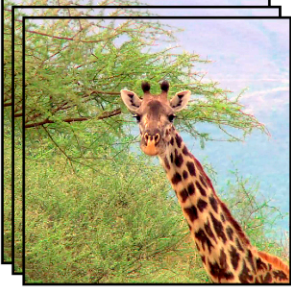

Table of Contents

Supplementary Figures and Tables

Attend T1

Supplementary Figure 1: Potential forms of attentional changes in voxel responses. Different types of attentional response modulations are demonstrated with tuning curves for a population of voxels. Each voxel tuning curve is represented with a Gaussian profile centered on the preferred stimulus attribute in an arbitrary stimulus space. The green curve represents a voxel tuned for attribute T1, whereas the purple curve represents a voxel tuned for attribute T2 during passive viewing. Attentional modulations are shown when attention is directed to attribute T1. **a,** Attention can increase the response baseline for a voxel that prefers the target attribute. **b,** Attention can increase the response gain for a voxel that prefers the target attribute. Any baseline or gain changes can be removed by z-scoring the responses for each voxel. Therefore, neither of these modulations changes the shape of voxel tuning curves, consistent with the labeled-line theory of attention. **c,** Attention can increase the responses to the target attribute, even in voxels that are not tuned for the target during passive viewing. Although this effect changes the shape of the tuning curves, it only affects the responses to the target attribute, while leaving responses to unattended attributes constant. **d,** In contrast, the tuning-shift hypothesis predicts that voxel tuning should shift toward the target attribute, changing responses for both attended and unattended attributes. This implies that tuning changes in single voxels cannot be entirely explained by changes in responses to target attributes.

Movies **Category Labels**

man	woman
adult	adult
person	person
organism	organism
⋮	⋮

airplane	*fly*
aircraft	*travel*
craft	
vehicle	
⋮	

car	man
motor vehicle	⋮
wheeled vehicle	*walk*
vehicle	*travel*
⋮	

giraffe	tree
ungulate	vascular plant
animal	plant
organism	organism
⋮	⋮

Supplementary Figure 2: Labeling of object and action categories. The object and action categories in each one-second natural movie clip were labeled by three raters. Here single frames from four typical movie clips are shown, along with their labels. The labels in black indicate the categories assigned manually, and the labels in gray are the corresponding superordinate categories that were assigned using the hierarchical relationships in WordNet. Action categories are shown in italics.

a

b
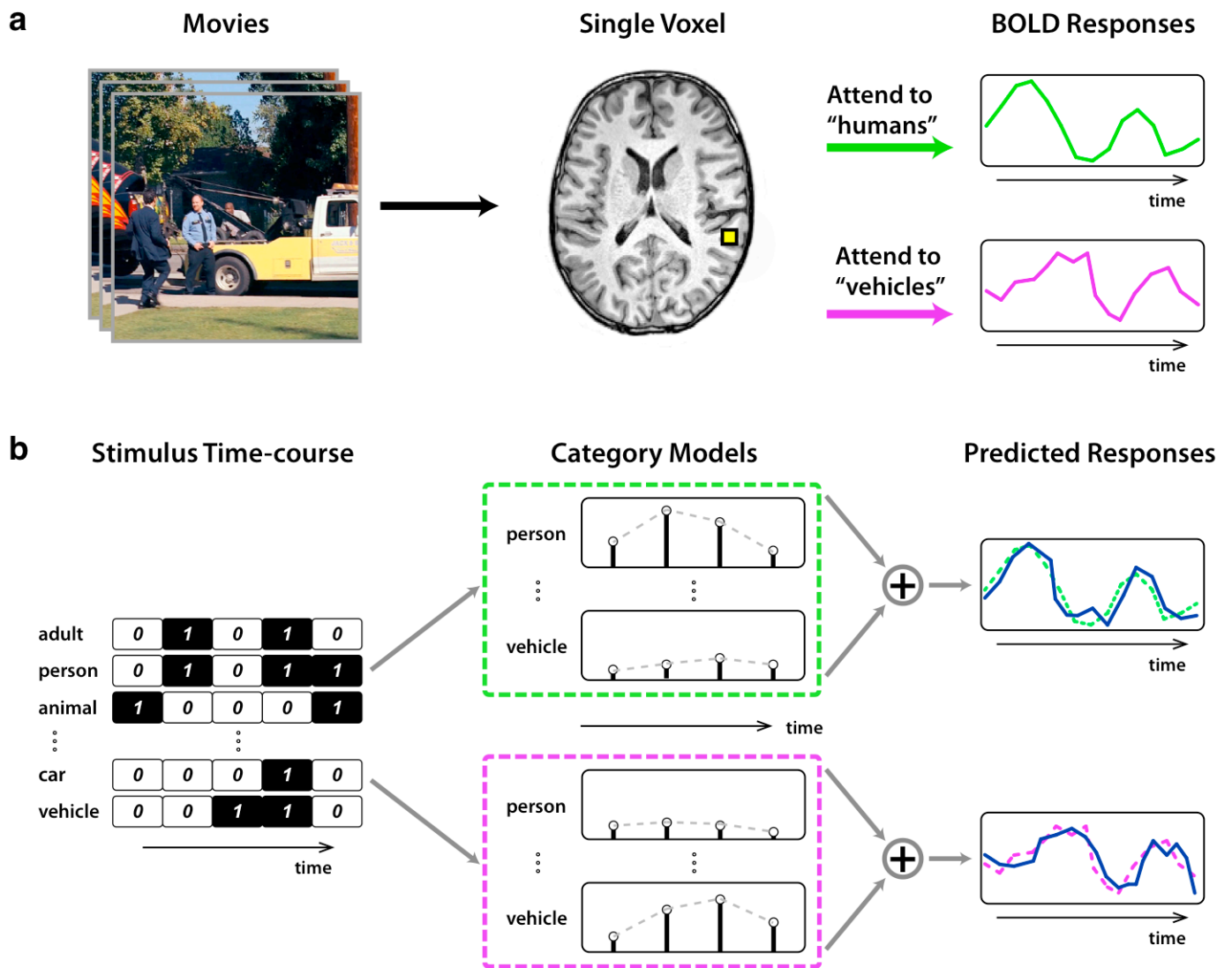
Supplementary Figure 3: Schematic of the model. a, Subjects viewed natural movies under fixation control while covertly searching for one of two object categories, 'humans' or 'vehicles'. BOLD responses were recorded from the whole brain, and each voxel was modeled individually. **b,** A total of 935 distinct object and action categories were labeled in the movies (Supplementary Fig. 2). A separate binary variable was used to indicate the presence or absence of each category in each one-second movie clip. A stimulus time-course was then built from these categorical indicator variables. The time samples were split into training and validation sets. The training set was used to estimate separate voxel-wise models under each attention condition ('humans': green bounding box, 'vehicles': magenta bounding box). The fit models were then evaluated by comparing the predicted responses (blue lines) to the observed responses (dashed lines) in the validation set.
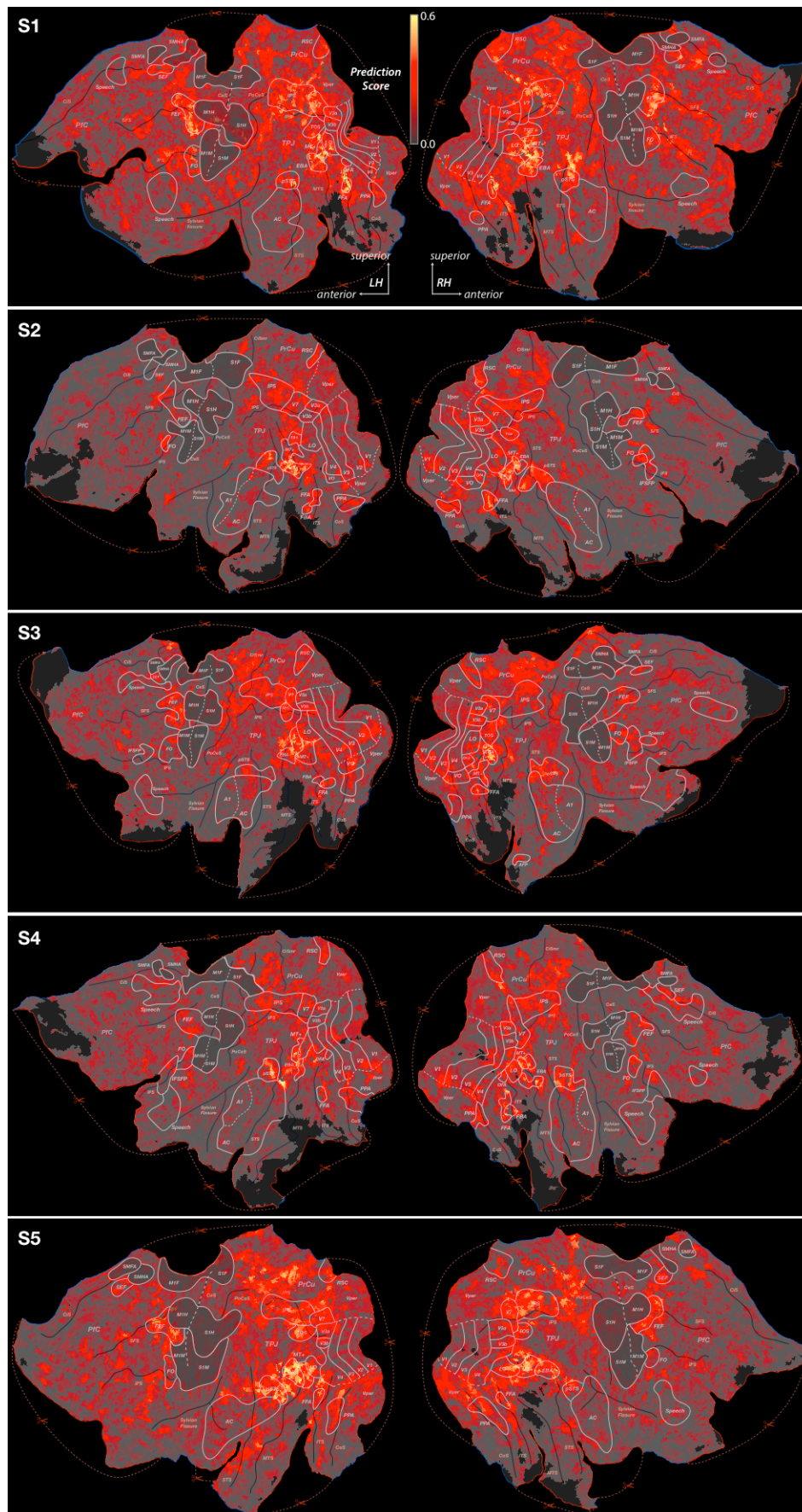
Supplementary Figure 4: Cortical flat maps of prediction scores for subjects S1-S5. The category model characterizes the category tuning of each voxel. To assess the functional importance of category tuning across the cortex, the prediction scores of the category models were visualized on cortical flat maps for subjects S1-S5. The prediction score was quantified as the raw correlation coefficient between the predicted and measured BOLD responses. Yellow voxels have high prediction scores, and gray voxels have low prediction scores (see legend). Insignificant voxels are shown in gray ($p>0.05$). Regions of fMRI signal dropout excluded from all analyses are shown with dark gray patches. The boundaries of cortical areas identified by standard localizers are indicated with solid (functionally-inferred) and dashed (anatomically-inferred) white lines. Major anatomical landmarks (blue font) and sulci (orange font and black lines) are also labeled (see Supplementary Tables 1 and 2 for abbreviations). The locations of flattening surface cuts and the boundary of corpus collosum are shown with red and blue borders. Many voxels in occipitotemporal cortex, posterior superior temporal sulcus, parietal cortex, and frontal cortex have high prediction scores.
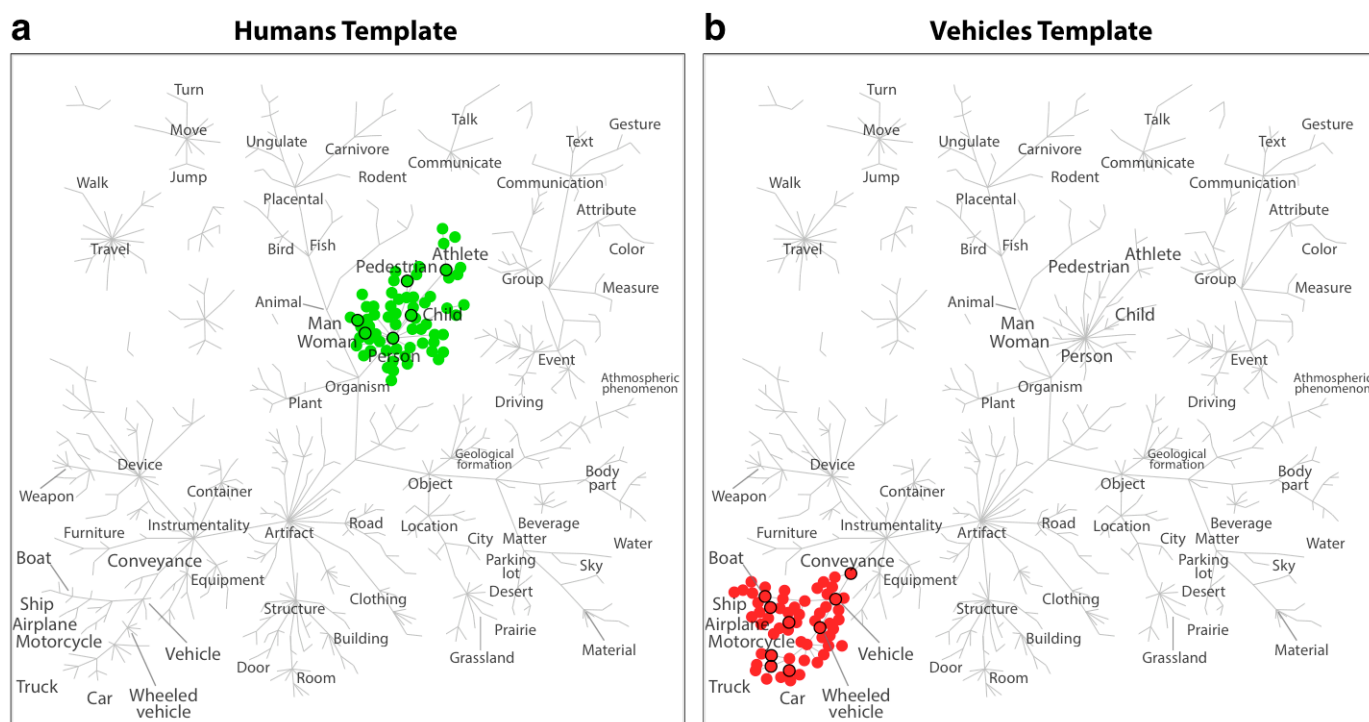
Supplementary Figure 5: Idealized tuning templates. Separate tuning templates were constructed to represent idealized voxels tuned either for **a,** 'humans' or **b,** 'vehicles'. The templates are binary masks that have unity weights for all subordinate categories of the WordNet synsets 'person.n.01' and 'conveyance.n.03', respectively. In this figure, several categories located within each search category template have been labeled and are identified with black circles.
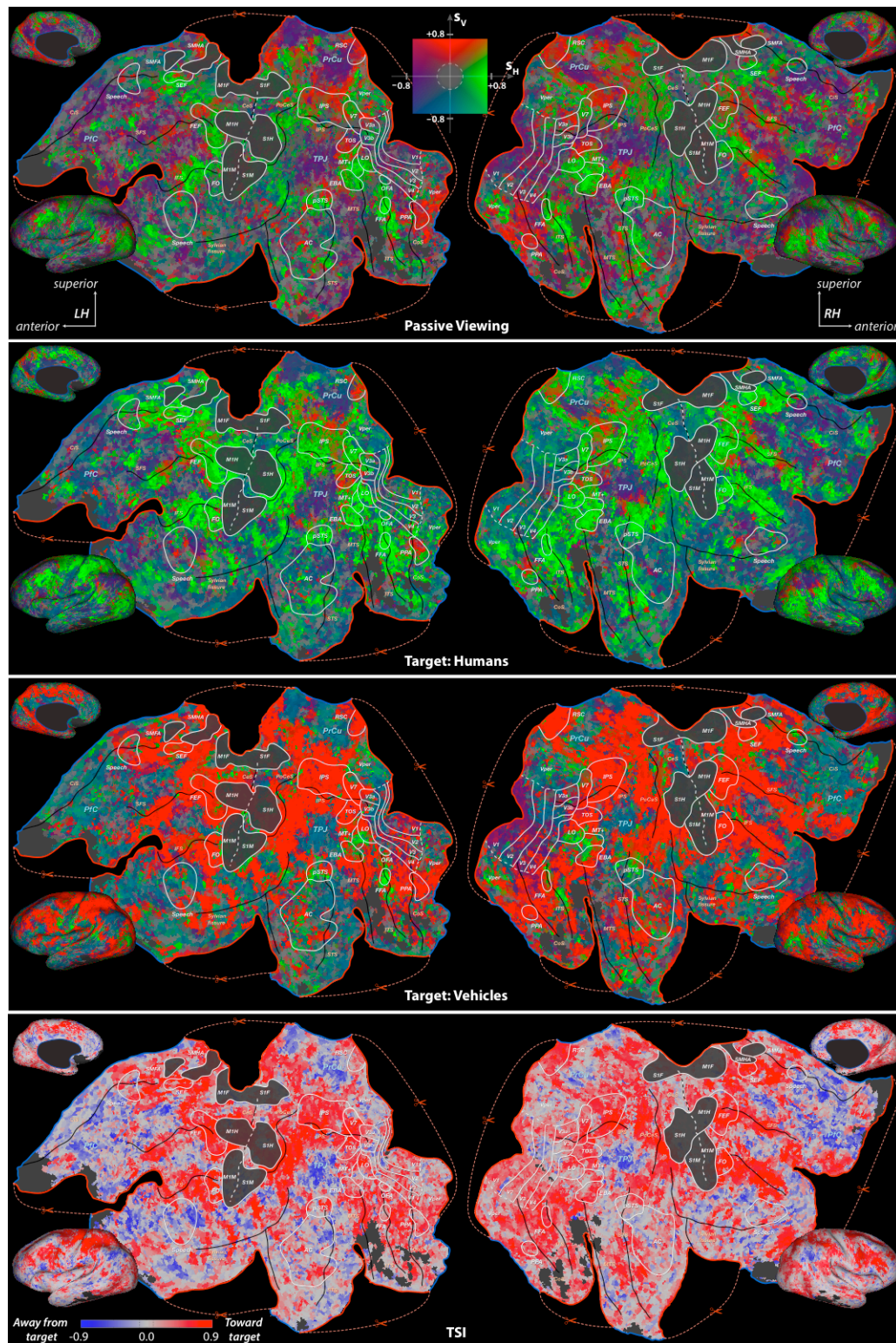
Supplementary Figure 6a: Cortical flat maps of tuning strength and TSI for subject S1. Tuning strength for 'humans' (S_H) and 'vehicles' (S_V) obtained during passive viewing (first row), search for 'humans' (second row), and search for 'vehicles' (third row). See legend in the middle and Online Methods for the color map. Insignificant voxels are shown in gray (p>0.05 within the dashed circle in the color map). Green/red voxels are more selectively tuned for 'humans'/'vehicles', and these predominate during search for 'humans'/'vehicles'. Tuning shift index (TSI, fourth row). The color bar in lower left represents the 95% central range of TSIs. These results indicate that tuning in many regions of cortex shifts toward the attended object category.
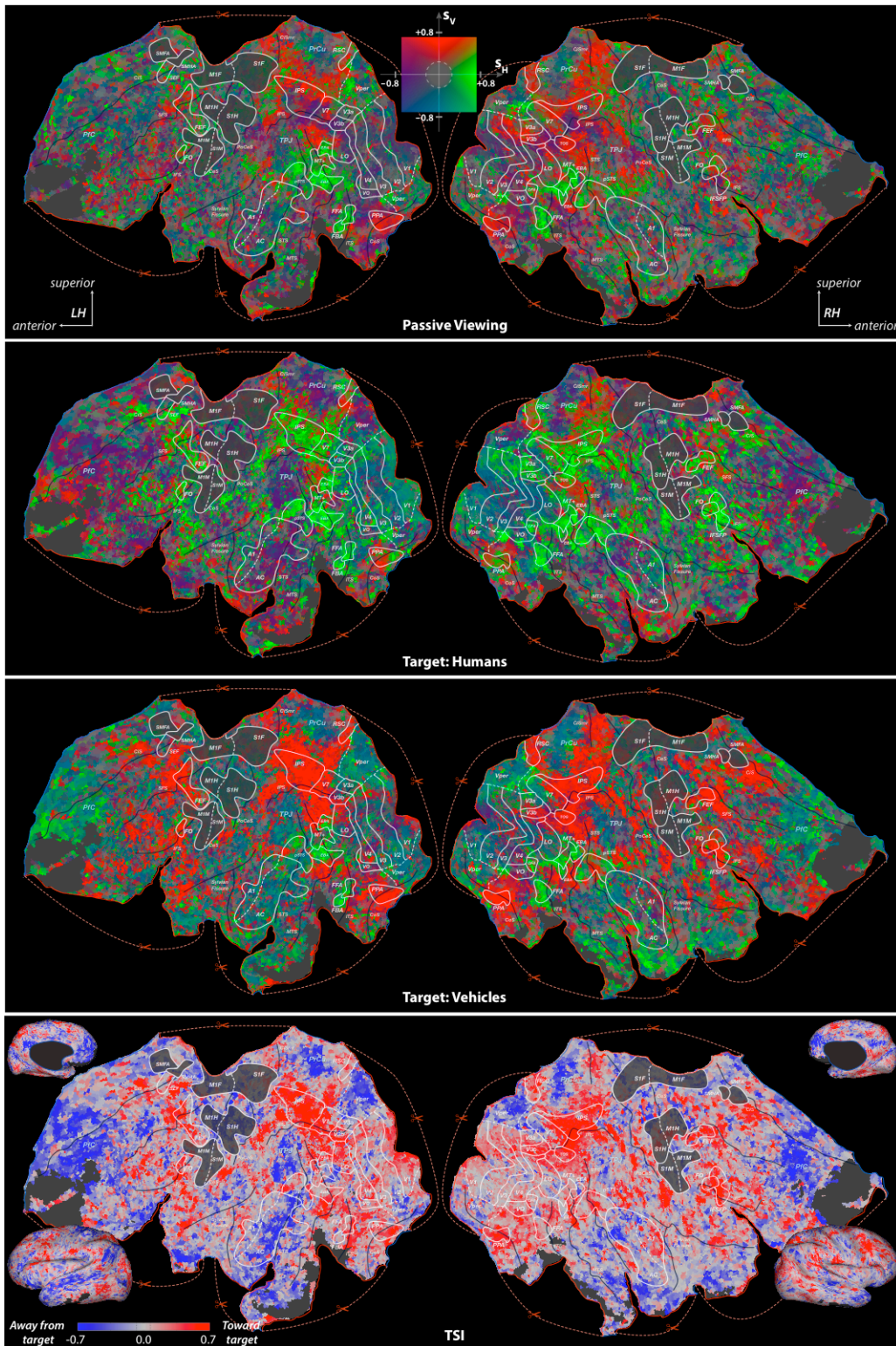
Supplementary Figure 6b: Cortical flat maps of tuning strength and TSI for subject S2. Tuning strength for 'humans' (S_H) and 'vehicles' (S_V) obtained during passive viewing (first row), search for 'humans' (second row), and search for 'vehicles' (third row). See legend in the middle and Online Methods for the color map. Insignificant voxels are shown in gray (p>0.05 within the dashed circle in the color map). Green/red voxels are more selectively tuned for 'humans'/'vehicles', and these predominate during search for 'humans'/'vehicles'. Tuning shift index (TSI, fourth row). The color bar in lower left represents the 95% central range of TSIs. These results indicate that tuning in many regions of cortex shifts toward the attended object category.
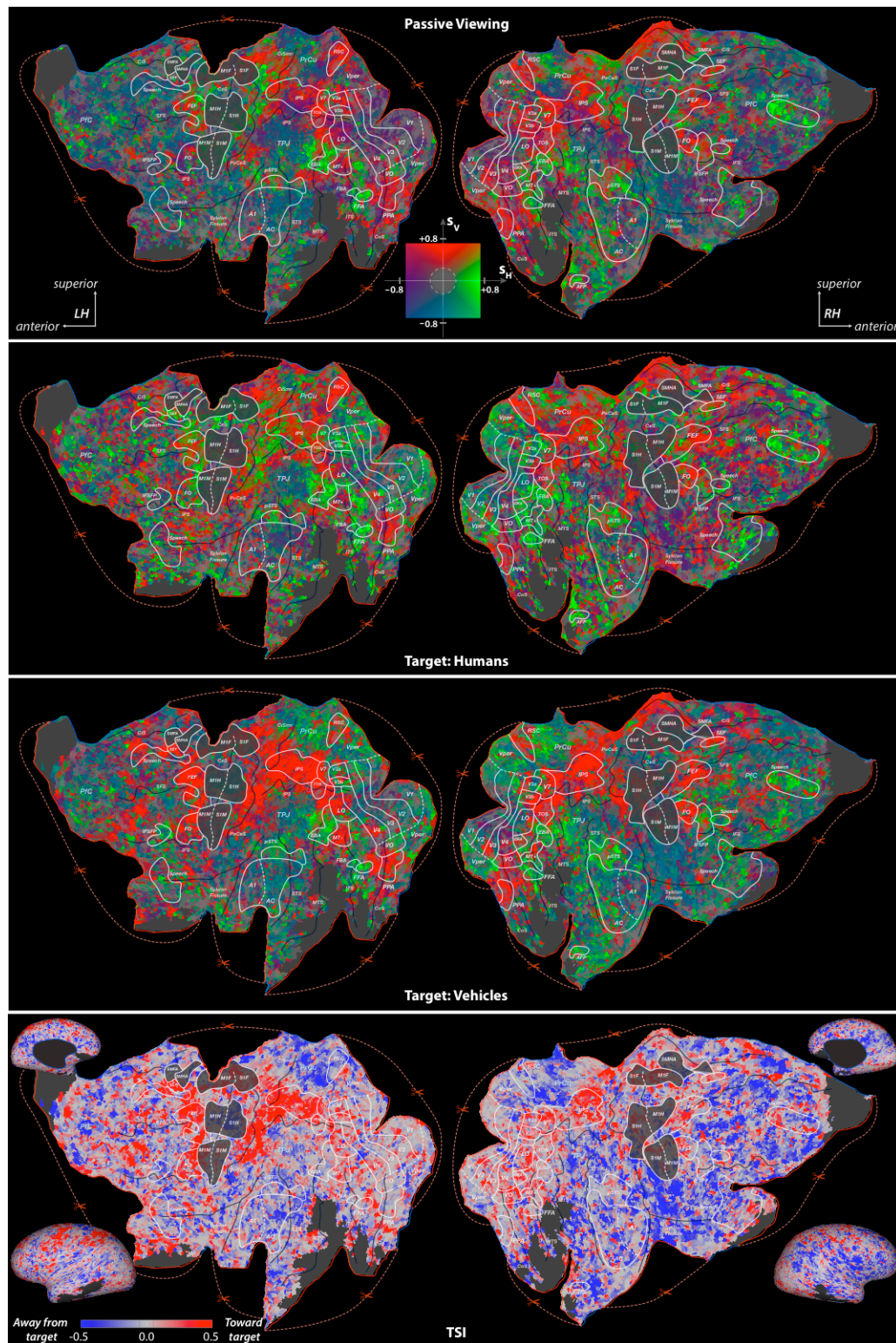
Supplementary Figure 6c: Cortical flat maps of tuning strength and TSI for subject S3. Tuning strength for 'humans' (S_H) and 'vehicles' (S_V) obtained during passive viewing (first row), search for 'humans' (second row), and search for 'vehicles' (third row). See legend in the middle and Online Methods for the color map. Insignificant voxels are shown in gray (p>0.05 within the dashed circle in the color map). Green/red voxels are more selectively tuned for 'humans'/'vehicles', and these predominate during search for 'humans'/'vehicles'. Tuning shift index (TSI, fourth row). The color bar in lower left represents the 95% central range of TSIs. These results indicate that tuning in many regions of cortex shifts toward the attended object category.
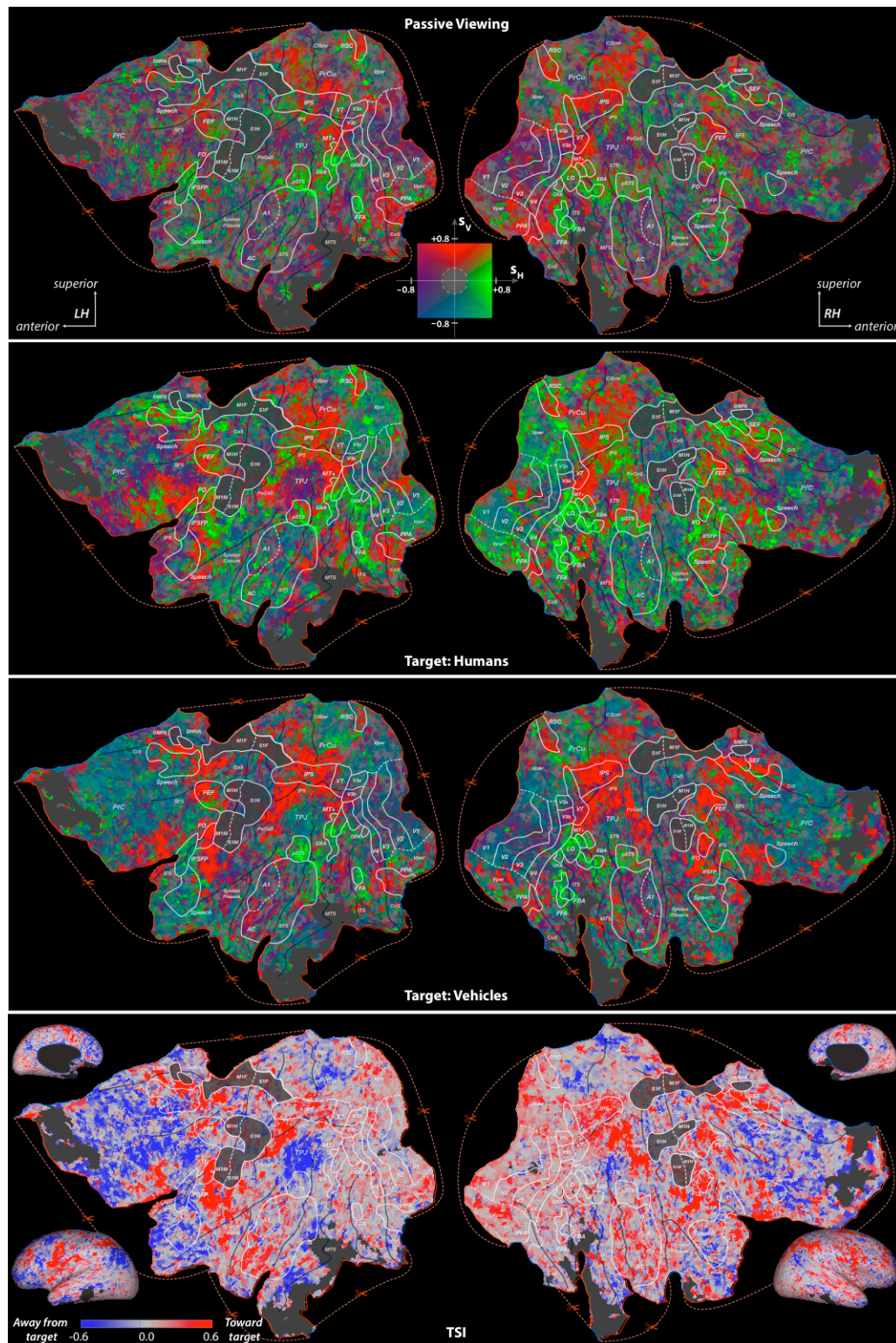
Supplementary Figure 6d: Cortical flat maps of tuning strength and TSI for subject S4. Tuning strength for 'humans' (S_H) and 'vehicles' (S_V) obtained during passive viewing (first row), search for 'humans' (second row), and search for 'vehicles' (third row). See legend in the middle and Online Methods for the color map. Insignificant voxels are shown in gray (p>0.05 within the dashed circle in the color map). Green/red voxels are more selectively tuned for 'humans'/'vehicles', and these predominate during search for 'humans'/'vehicles'. Tuning shift index (TSI, fourth row). The color bar in lower left represents the 95% central range of TSIs. These results indicate that tuning in many regions of cortex shifts toward the attended object category.
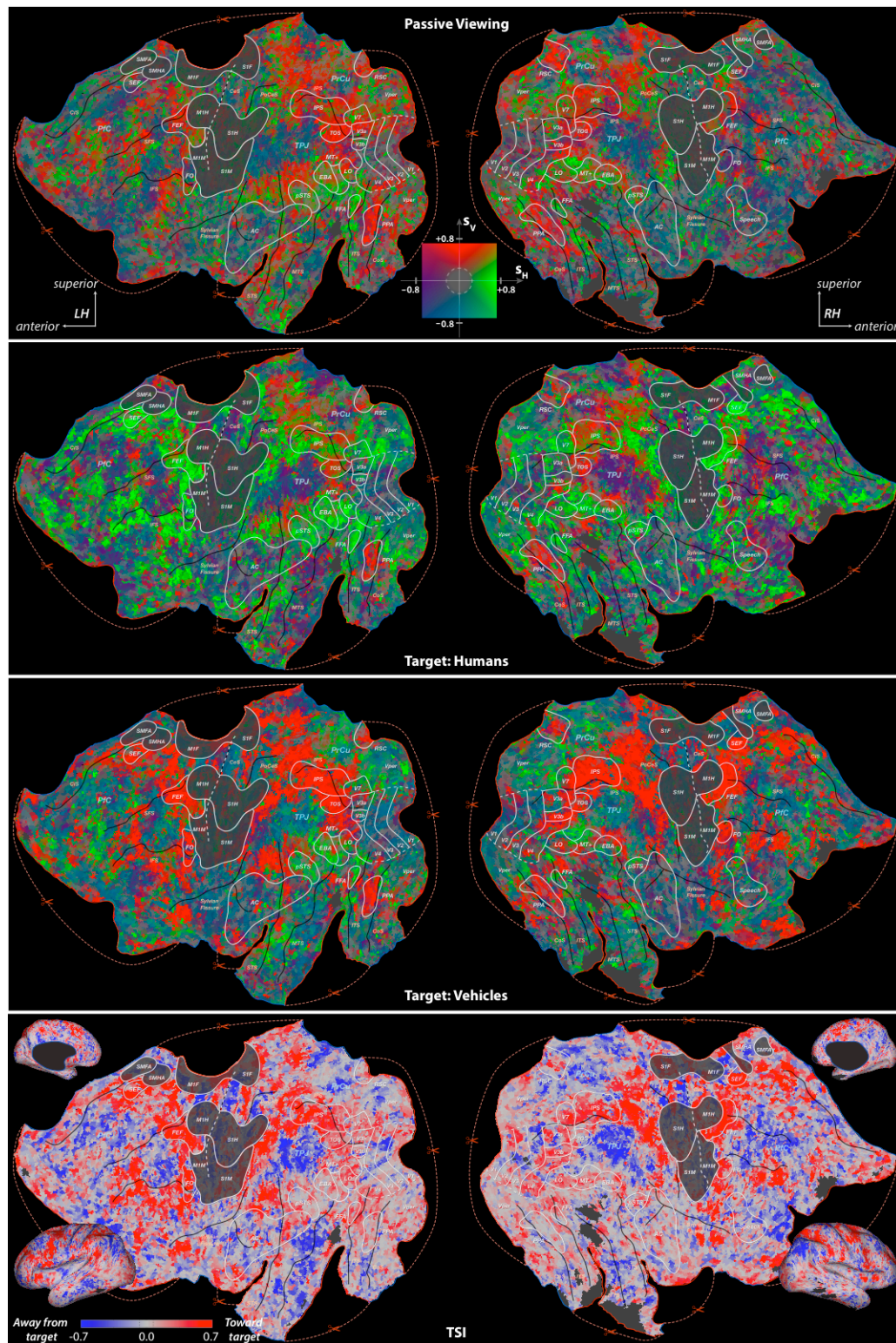
Supplementary Figure 6e: Cortical flat maps of tuning strength and TSI for subject S5. Tuning strength for 'humans' (S_H) and 'vehicles' (S_V) obtained during passive viewing (first row), search for 'humans' (second row), and search for 'vehicles' (third row). See legend in the middle and Online Methods for the color map. Insignificant voxels are shown in gray (p>0.05 within the dashed circle in the color map). Green/red voxels are more selectively tuned for 'humans'/'vehicles', and these predominate during search for 'humans'/'vehicles'. Tuning shift index (TSI, fourth row). The color bar in lower left represents the 95% central range of TSIs. These results indicate that tuning in many regions of cortex shifts toward the attended object category.
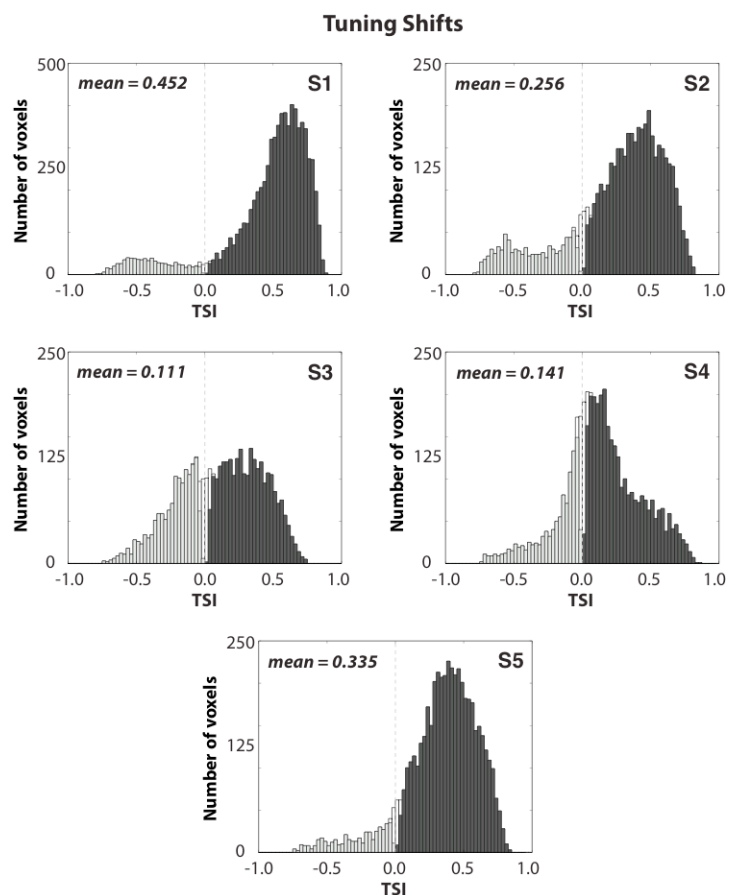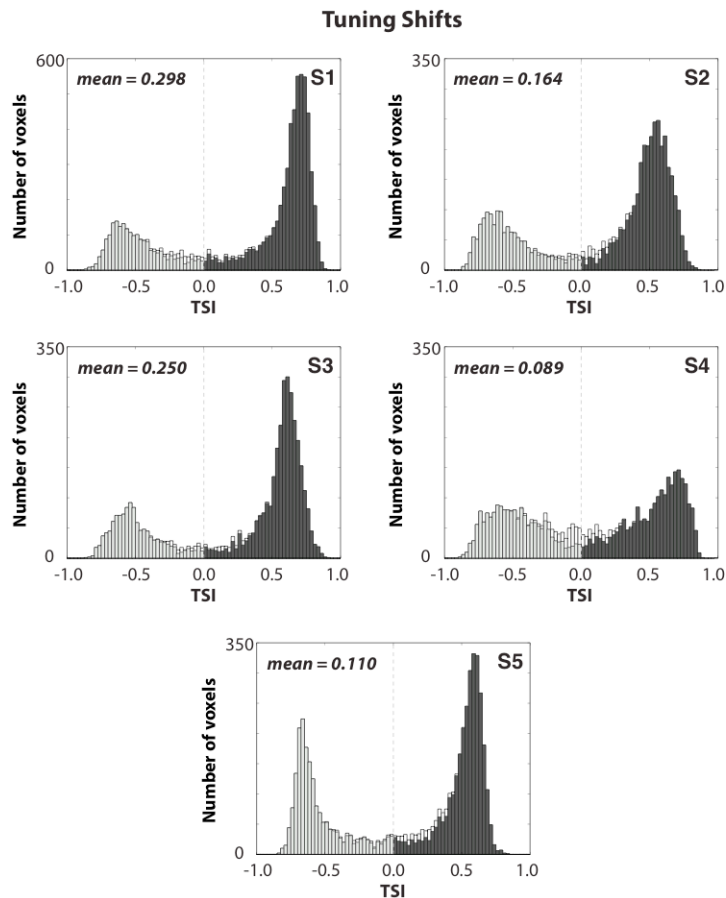
Supplementary Figure 7: Voxel-wise tuning shifts: main analysis. The TSI distributions estimated in the presence of the target stimuli for all cortical voxels of subjects S1-S5. Only voxels that had significant model weights (t-test, $p<0.05$) and high prediction scores (above mean plus 1 standard deviation) were selected for this analysis (4245-7785, depending on subject). Dark gray bars represent TSIs significantly greater than 0 (Wilcoxon signed-rank test, $p<0.05$), and light gray bars indicate TSIs significantly less than 0 ($p<0.05$). Unfilled portions of each bar represent insignificant TSIs. The mean TSI is significantly greater than 0 in all subjects ($p<10^{-6}$). These results indicate that most well-modeled cortical voxels shift their tuning toward the attended category.

Tuning Shifts

S1 mean = 0.298

S2 mean = 0.164

S3 mean = 0.250

S4 mean = 0.089

S5 mean = 0.110

Supplementary Figure 8: Voxel-wise tuning shifts in the absence of targets: control analysis. The TSI distributions for all cortical voxels of subjects S1-S5 in the control analysis where no 'humans' or 'vehicles' were present. Voxel selection and format of each panel are the same as in Supplementary Fig. 7. The mean TSI is significantly greater than 0 in all subjects (Wilcoxon signed-rank test, $p<10^{-6}$). These results clearly show that attention shifts tuning of unattended categories towards the search target even when the target is not present in the movie. Furthermore, the tuning shifts are in consistent directions across the main and control analyses for an average of 65.42±7.73% (mean±s.d.) of cortical voxels across subjects. This result indicates that the tuning shifts found in the main experiment are caused by category-based attention and are not a mere consequence of target detection.
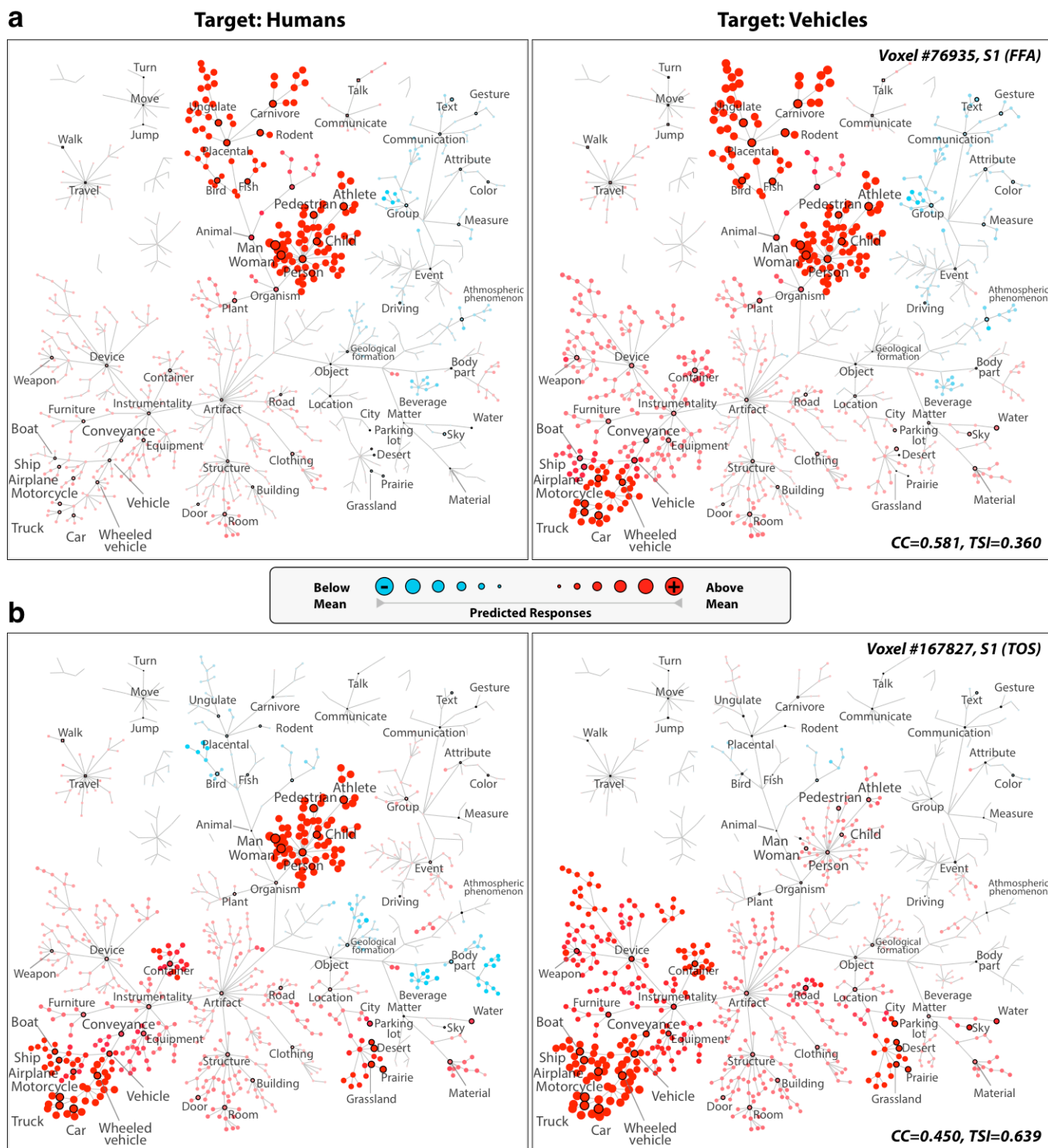
Supplementary Figure 9: Tuning for representative voxels in FFA and TOS. Tuning for 935 object and action categories in two well-modeled voxels (t-test, $p<10^{-6}$) selected from **a,** FFA and **b,** TOS. Tuning is shown separately for two conditions, searching for 'humans' (left) and 'vehicles' (right). CC denotes the prediction score for each voxel. Red versus blue nodes correspond to categories that evoke above- and below-mean responses. In these category-selective ROIs, the tuning for the 'preferred' categories ('humans' for FFA, and 'places' and 'vehicles' for TOS) is maintained regardless of the attended category. Therefore, in the condition where attention is directed to a 'non-preferred' category, these voxels exhibit additional tuning for the attended category.
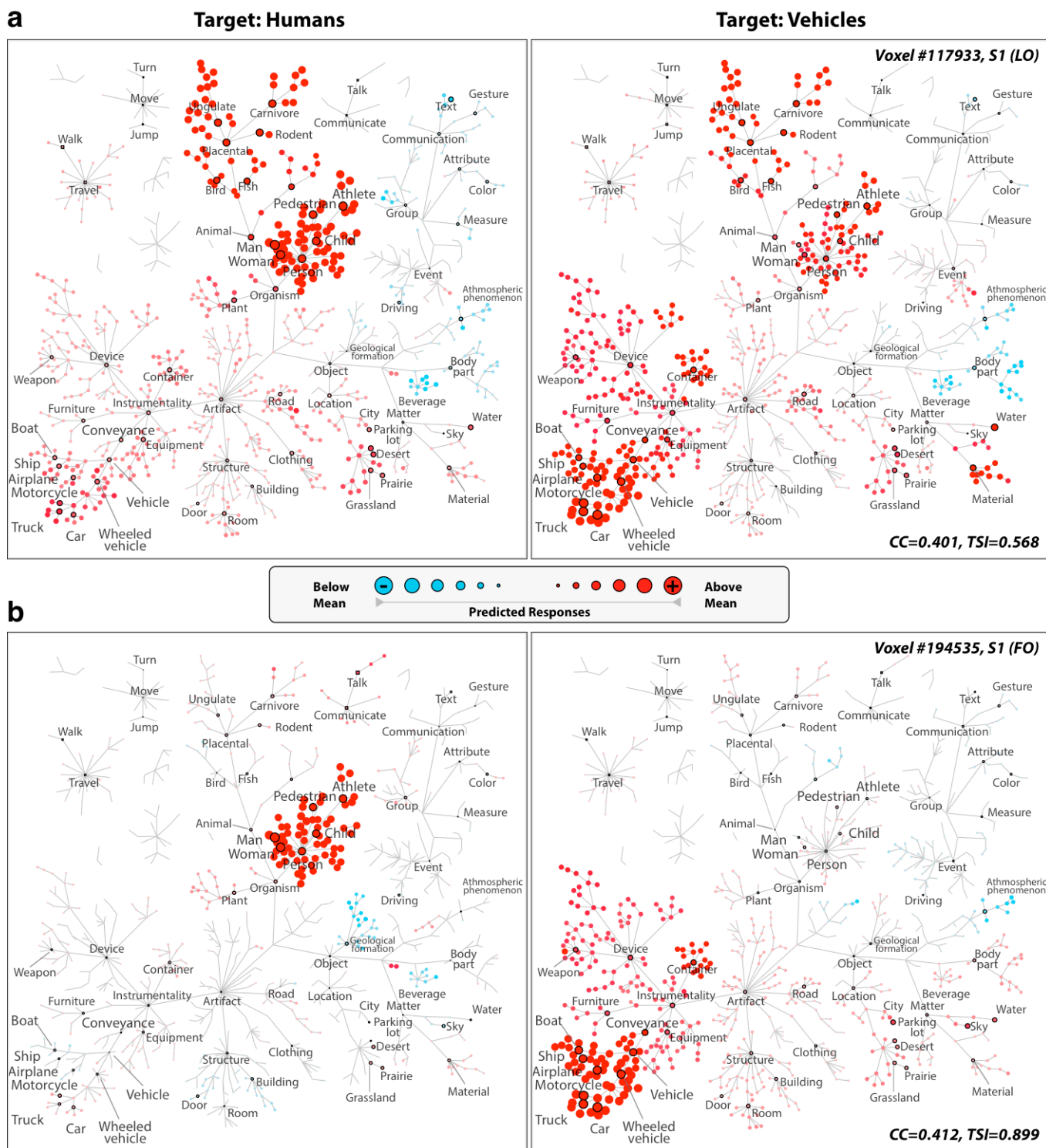
Supplementary Figure 10: Tuning for representative voxels in LO and FO. Tuning for 935 object and action categories in two well-modeled voxels (t-test, $p<10^{-6}$) selected from **a,** LO and **b,** FO. Tuning is shown separately for two conditions, searching for 'humans' (left) and 'vehicles' (right). CC denotes the prediction score for each voxel. Format of each panel is the same as in Supplementary Fig. 9. The LO voxel exhibits strong tuning for the attended category in both conditions. Furthermore, significant albeit weaker tuning is observed for the unattended categories. In contrast, the FO voxel is selectively tuned for the attended category alone.
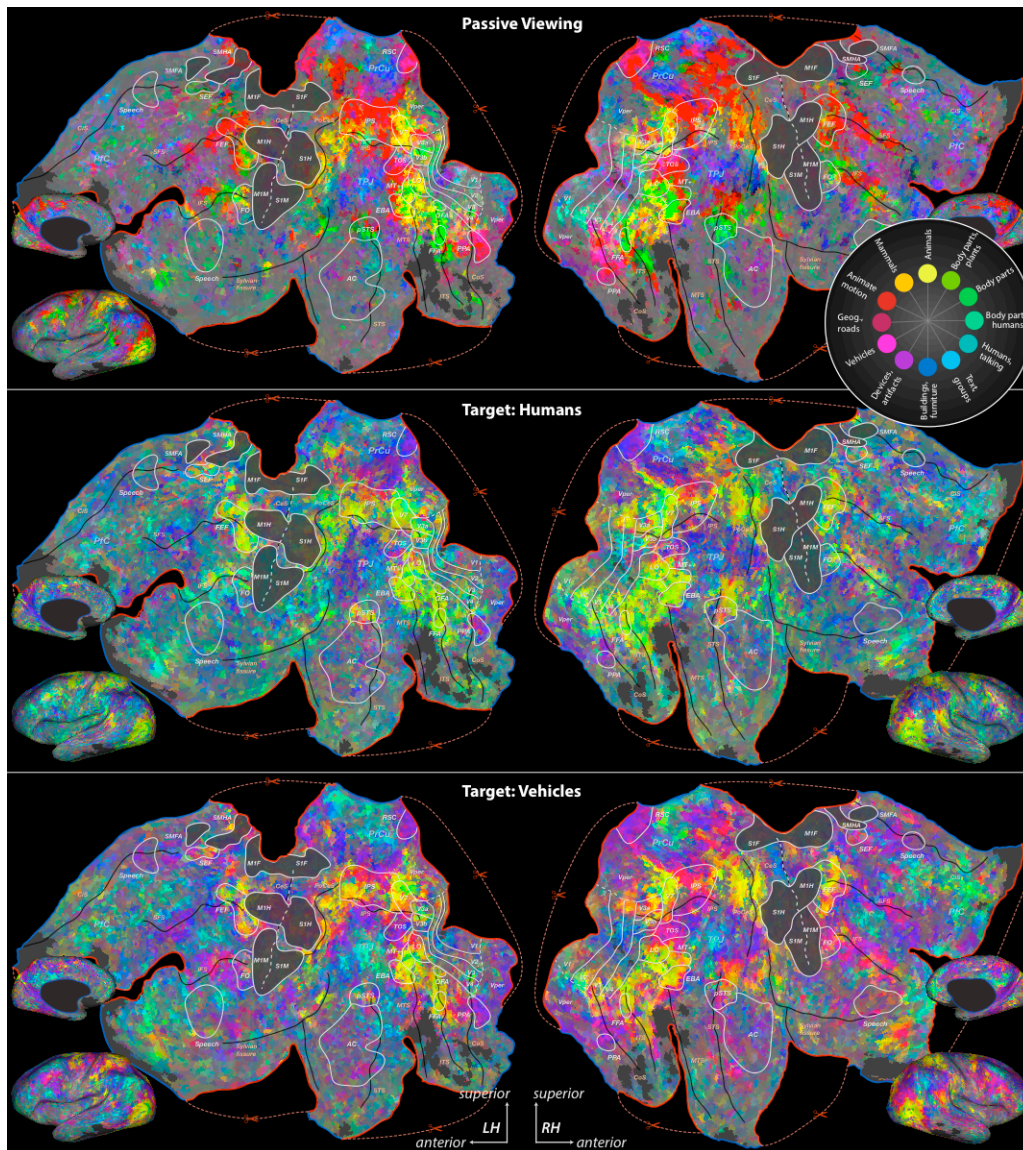
Supplementary Figure 11a: Cortical flat maps of semantic tuning for subject S1. First row,
Distribution of semantic tuning across the cortex during passive viewing. Tuning was estimated from
responses to all available movie clips. Voxels with similar tuning are assigned similar colors (see legend on
the right and Online Methods). Insignificant voxels are shown in gray (p>0.05). Yellow/green voxels are
more selectively tuned for animals and body parts, and purple/red voxels are more selectively tuned for
geographic locations and movement. **Second row,** Distribution of semantic tuning during search for
'humans'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Yellow-green voxels that are tuned for animals and body
parts predominate during search for 'humans'. **Third row,** Distribution of semantic tuning during search
for 'vehicles'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Purple voxels that are tuned for geographic locations and
artifacts predominate during search for 'vehicles'. These results indicate that semantic tuning in many
regions of cortex shifts toward the attended object category, even when the targets are absent.
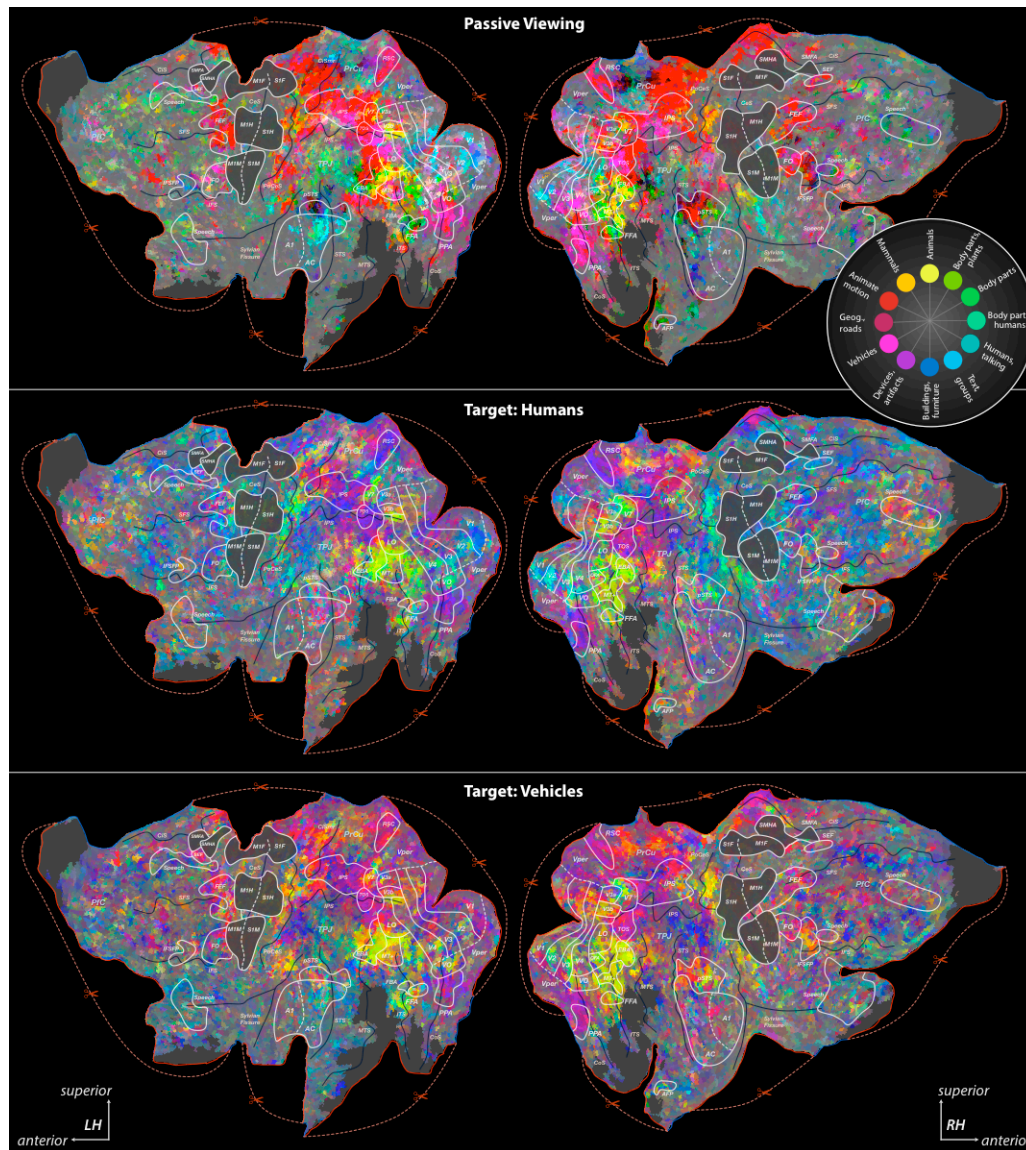
Supplementary Figure 11b: Cortical flat maps of semantic tuning for subject S2. First row,
Distribution of semantic tuning across the cortex during passive viewing. Tuning was estimated from responses to all available movie clips. Voxels with similar tuning are assigned similar colors (see legend on the right and Online Methods). Insignificant voxels are shown in gray (p>0.05). Yellow/green voxels are more selectively tuned for animals and body parts, and purple/red voxels are more selectively tuned for geographic locations and movement. **Second row,** Distribution of semantic tuning during search for 'humans'. Tuning was estimated only from responses evoked by movie clips in which the target did not appear. Color assignment same as in first row. Yellow-green voxels that are tuned for animals and body parts predominate during search for 'humans'. **Third row,** Distribution of semantic tuning during search for 'vehicles'. Tuning was estimated only from responses evoked by movie clips in which the target did not appear. Color assignment same as in first row. Purple voxels that are tuned for geographic locations and artifacts predominate during search for 'vehicles'. These results indicate that semantic tuning in many regions of cortex shifts toward the attended object category, even when the targets are absent.
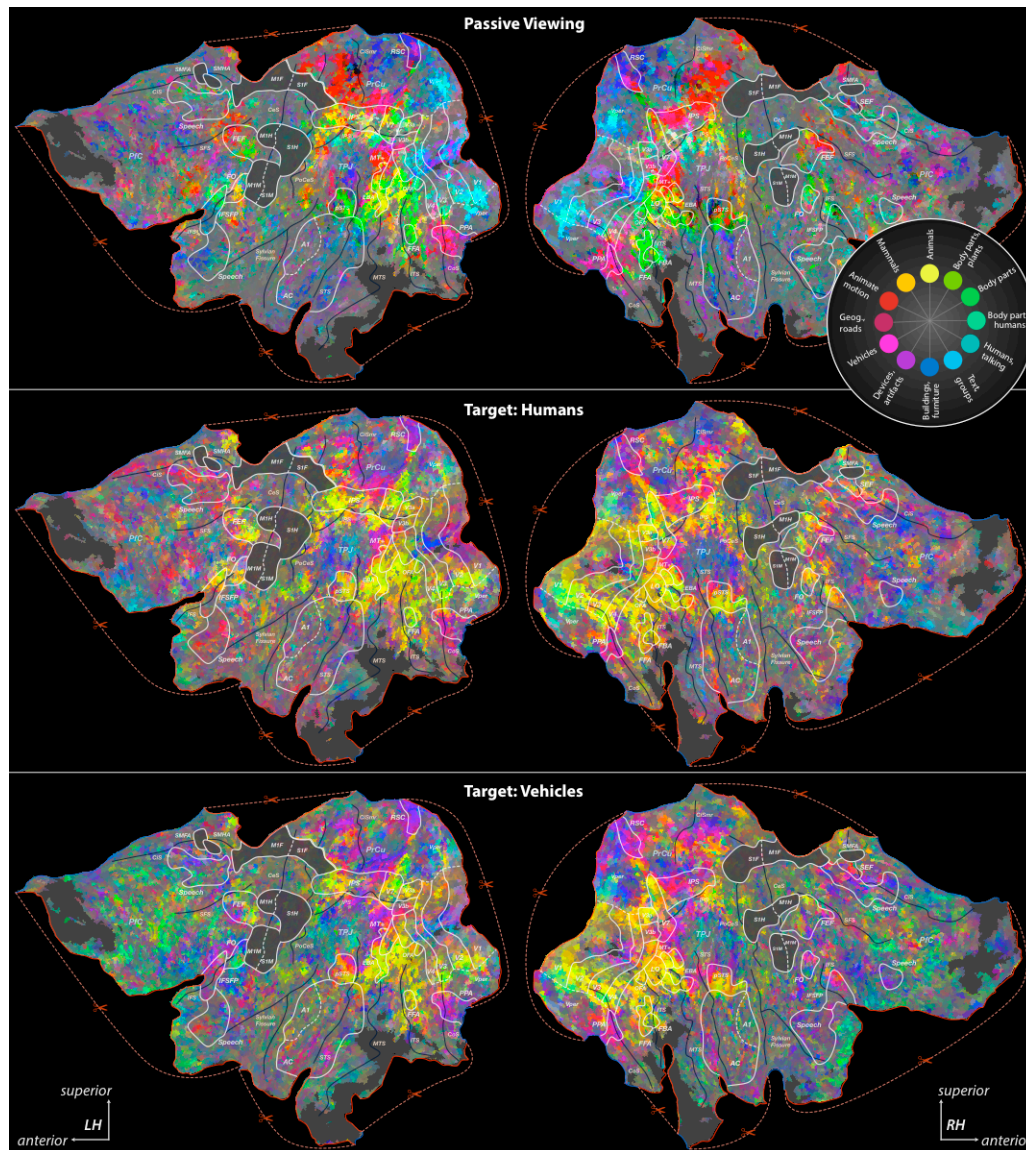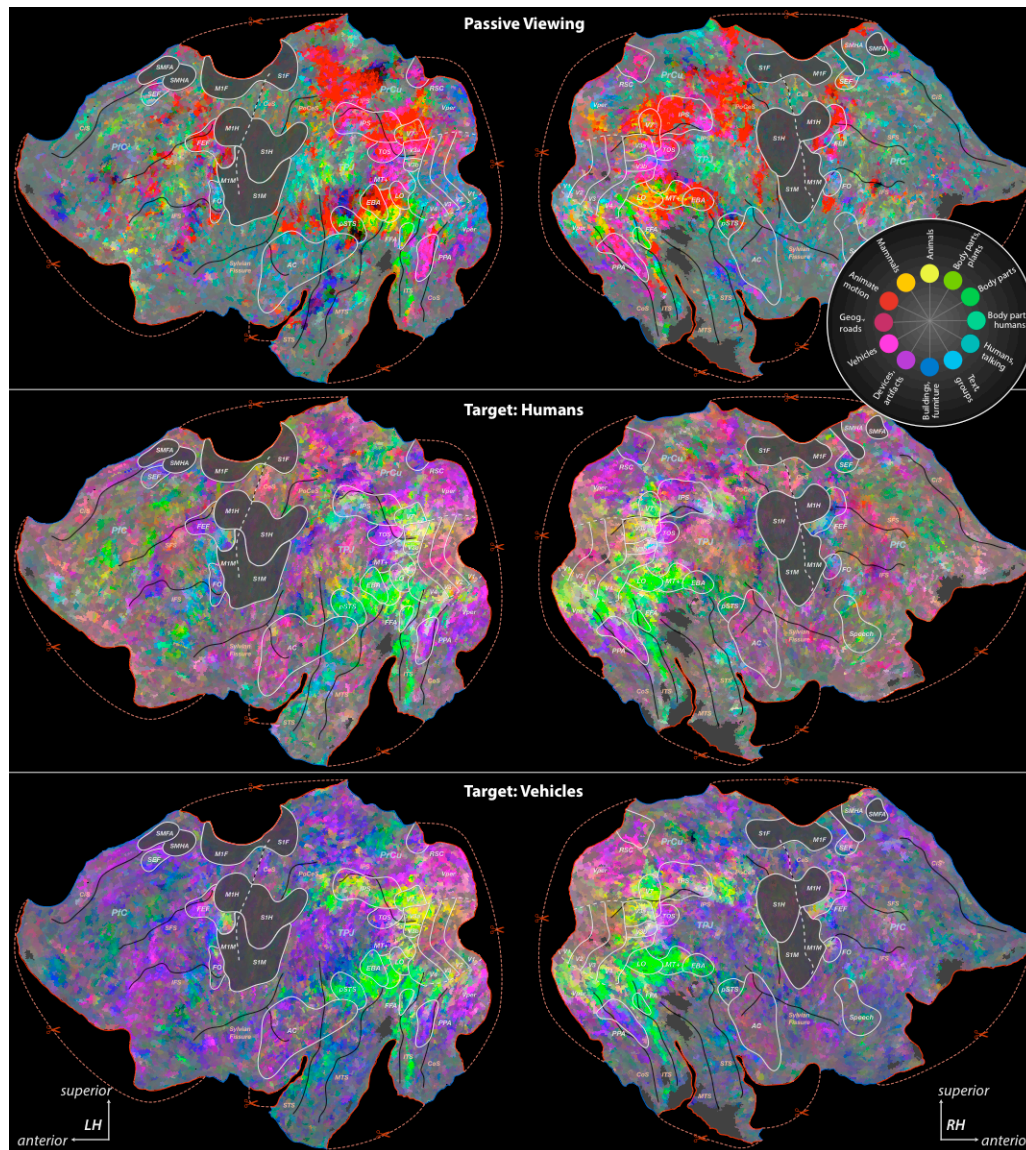
Supplementary Figure 11c: Cortical flat maps of semantic tuning for subject S3. First row,
Distribution of semantic tuning across the cortex during passive viewing. Tuning was estimated from
responses to all available movie clips. Voxels with similar tuning are assigned similar colors (see legend on
the right and Online Methods). Insignificant voxels are shown in gray (p>0.05). Yellow/green voxels are
more selectively tuned for animals and body parts, and purple/red voxels are more selectively tuned for
geographic locations and movement. **Second row,** Distribution of semantic tuning during search for
'humans'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Yellow-green voxels that are tuned for animals and body
parts predominate during search for 'humans'. **Third row,** Distribution of semantic tuning during search
for 'vehicles'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Purple voxels that are tuned for geographic locations and
artifacts predominate during search for 'vehicles'. These results indicate that semantic tuning in many
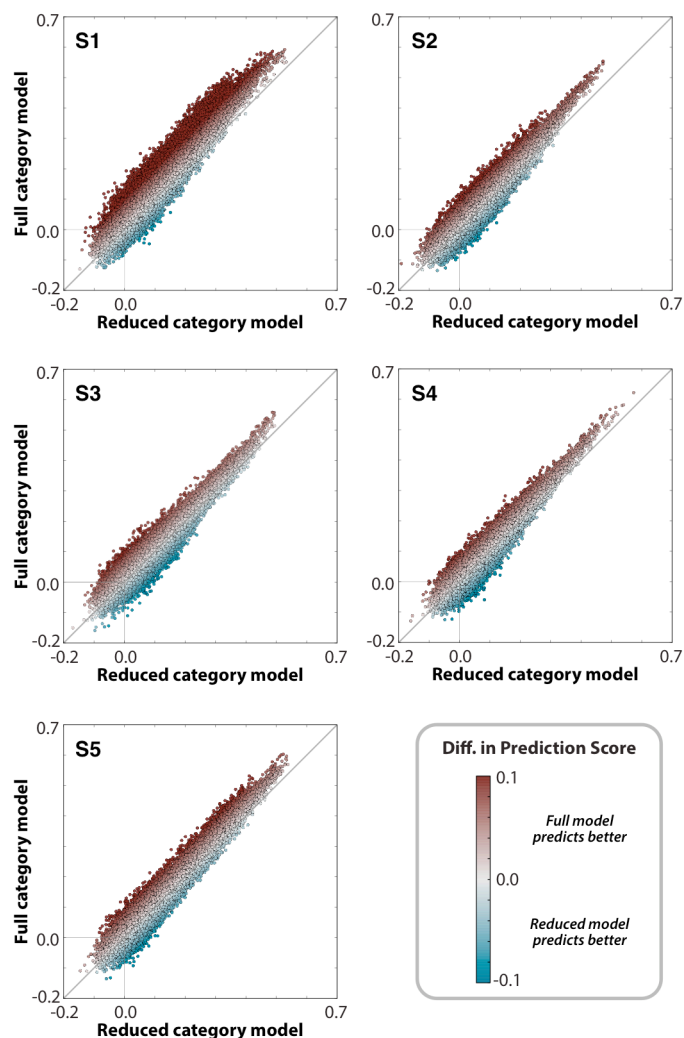regions of cortex shifts toward the attended object category, even when the targets are absent.

Supplementary Figure 11d: Cortical flat maps of semantic tuning for subject S4. First row,
Distribution of semantic tuning across the cortex during passive viewing. Tuning was estimated from
responses to all available movie clips. Voxels with similar tuning are assigned similar colors (see legend on
the right and Online Methods). Insignificant voxels are shown in gray (p>0.05). Yellow/green voxels are
more selectively tuned for animals and body parts, and purple/red voxels are more selectively tuned for
geographic locations and movement. **Second row,** Distribution of semantic tuning during search for
'humans'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Yellow-green voxels that are tuned for animals and body
parts predominate during search for 'humans'. **Third row,** Distribution of semantic tuning during search
for 'vehicles'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Purple voxels that are tuned for geographic locations and
artifacts predominate during search for 'vehicles'. These results indicate that semantic tuning in many
regions of cortex shifts toward the attended object category, even when the targets are absent.

Supplementary Figure 11e: Cortical flat maps of semantic tuning for subject S5. First row,
Distribution of semantic tuning across the cortex during passive viewing. Tuning was estimated from
responses to all available movie clips. Voxels with similar tuning are assigned similar colors (see legend on
the right and Online Methods). Insignificant voxels are shown in gray (p>0.05). Yellow/green voxels are
more selectively tuned for animals and body parts, and purple/red voxels are more selectively tuned for
geographic locations and movement. **Second row,** Distribution of semantic tuning during search for
'humans'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Yellow-green voxels that are tuned for animals and body
parts predominate during search for 'humans'. **Third row,** Distribution of semantic tuning during search
for 'vehicles'. Tuning was estimated only from responses evoked by movie clips in which the target did not
appear. Color assignment same as in first row. Purple voxels that are tuned for geographic locations and
artifacts predominate during search for 'vehicles'. These results indicate that semantic tuning in many
regions of cortex shifts toward the attended object category, even when the targets are absent.

Supplementary Figure 12: Comparison of full and reduced category models. The prediction scores (raw correlation coefficient) of the full (935 regressors) and reduced (604 regressors) category models were measured for all cortical voxels of subjects S1-S5. Each voxel is represented with a dot whose color is scaled according to the legend. Brown versus teal voxels indicate whether the full model or the reduced model yields better prediction scores (t-test, $p < 0.05$). The full category model outperforms the reduced category model for an average of 88.93±7.38% (mean±s.d.; results averaged over all 5 subjects) of well-modeled cortical voxels (with prediction scores above mean plus 1 standard deviation).
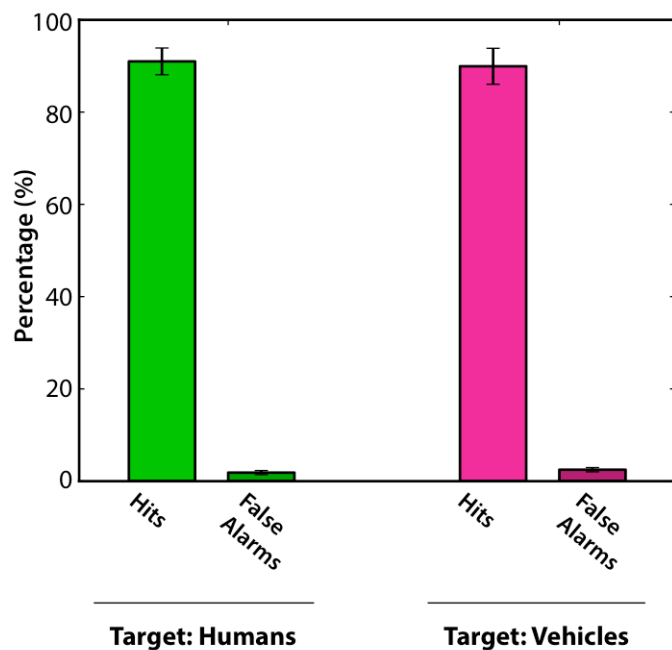
Supplementary Figure 13: Behavioral performance under different attention conditions. The behavioral performance data for the button-press task during the main experiment is displayed for two attention conditions. The bar plots (mean±s.e.m., results averaged over all 5 subjects) show the hit and false alarm rates while subjects searched for 'humans' (green) or 'vehicles' (magenta). There are no significant differences (Wilcoxon rank-sum test) in the hit rates (p>0.84), false alarm rates (p>0.46), or d′ values (p>0.42) for the two search targets.

Supplementary Table 1. Abbreviations for functional regions-of-interest

ROI Name	Anatomical Location	Type of Localizer	References
FFA (fusiform face area)	Posterior fusiform gyrus	Faces – objects	Kanwisher et al., 1997 McCarthy et al., 1997
OFA (occipital face area)	Just anterior to V4v/VO	Faces – objects	Kanwisher et al., 1997 Halgren et al., 1999
IFSFP (inferior frontal sulcus face patch)	IFS anterior to precentral sulcus	Faces – objects	Avidan et al., 2005 Tsao et al., 2008
ATFP (anterior temporal face patch)	Temporal pole	Faces – objects	Rajimehr et al., 2009
EBA (extrastriate body area)	Anterior to MT+ on the medial temporal gyrus	Human bodies – objects	Downing et al., 2001
FBA (fusiform body area)	Fusiform sulcus/gyrus anterior to FFA	Human bodies – objects	Peelen & Downing, 2005 Schwarzlose et al., 2005
PPA (parahippocampal place area)	Collateral fissure	Scenes – objects	Epstein & Kanwisher, 1998
TOS (transverse occipital sulcus)	Just inferior to/overlapping with V7	Scenes – objects	Hasson et al., 2003
RSC (retrosplenial cortex)	Medial wall just superior to PPA	Scenes – objects	Aguirre et al., 1998
FEF (frontal eye fields)	Precentral sulcus and superior frontal sulcus	Self generated saccades – fixation	Paus, 1996
FO (frontal operculum)	Inferior portion of precentral sulcus	Self generated saccades – fixation	Berman et al., 1999 Corbetta et al., 1998
SEF (supplementary eye fields)	Dorsal-medial frontal cortex	Self generated saccades – fixation	Grosbras et al., 1999
V_{per} (visual periphery, including V1-V4)	Surrounding mapped retinotopic visual cortex	Self generated saccades – fixation	
MT+ (middle temporal)	Posterior inferior temporal sulcus	Coherent – scrambled motion	Tootell et al., 1995
pSTS (posterior superior temporal sulcus)	Posterior superior temporal sulcus	High auditory and visual repeatability	
AC (auditory cortex)	Superior temporal gyrus	Auditory repeatability	
S1F/M1F (primary somatosensory and motor cortex, foot)	Superior-medial central sulcus	Foot motion – rest, S1F and M1F split at fundus of central sulcus	Penfield and Boldrey, 1937
S1H/M1H (primary somatosensory and motor cortex, hand)	Central sulcus	Hand motion – rest, S1H and M1H split at fundus of central sulcus	Penfield and Boldrey, 1937

S1M/M1M (primary somatosensory and motor cortex, mouth)	Inferior central sulcus	Mouth motion – rest, S1M and M1M split at fundus of central sulcus	Penfield and Boldrey, 1937
SMHA (supplementary motor hand area)	Middle cingulate gyrus	Hand motion – rest	Fried et al., 1991
SMFA (supplementary motor foot area)	Middle cingulate gyrus/sulcus	Foot motion – rest	Fried et al., 1991
IPS (intraparietal sulcus)	Lateral parietal cortex	Retinotopy	Connolly et al., 2000
V1-V4, V3A, V3B	Occipital cortex	Retinotopy	Hansen et al., 2007
LO (lat. occipital complex)	Anterior to V4	Retinotopy	Hansen et al., 2007
VO	Inferior to V4v	Retinotopy	Brewer et al., 2005
V7	Anterior to V3A/V3B	Retinotopy	Hansen et al., 2007

Supplementary Table 2. Abbreviations for anatomical landmarks

Abbreviation	Full Name
CeS	Central sulcus
CiS	Cingulate sulcus
CiSmr	Marginal ramus of the cingulate sulcus
CoS	Collateral sulcus
IFS	Inferior frontal sulcus
IPS	Intraparietal sulcus
ITS	Inferior temporal sulcus
MTS	Medial temporal sulcus
PfC	Prefrontal cortex
PoCeS	Postcentral sulcus
PrCu	Precuneus
SFS	Superior frontal sulcus
STS	Superior temporal sulcus
TPJ	Temporoparietal junction