BIASED COMPETITION IN SEMANTIC REPRESENTATIONS ACROSS THE HUMAN BRAIN DURING CATEGORY-BASED VISUAL SEARCH

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Mohammad Shahdloo

January 2017

Biased competition in semantic representations across the human brain during category-based visual search

By Mohammad Shahdloo

January 2017

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Tolga Çukur(Advisor)

HacıHulusi Kafalıgönül

Erol Özçelik

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

BIASED COMPETITION IN SEMANTIC REPRESENTATIONS ACROSS THE HUMAN BRAIN DURING CATEGORY-BASED VISUAL SEARCH

Mohammad Shahdloo
M.S. in Electrical and Electronics Engineering
Advisor: Tolga Çukur
January 2017

Humans can perceive thousands of distinct object and action categories in the visual scene and successfully divide their attention among multiple target categories. It has been shown that object and action categories are represented in a continuous semantic map across the cortical surface and attending to a specific category causes broad shifts in voxel-wise semantic tuning profiles to enhance the representation of the target category. However, the effects of divided attention to multiple categories on semantic representation remain unclear. In line with predictions of the biased-competition model, recent evidence suggests that brain response to two objects presented simultaneously can be described as a weighted average of the responses to individual objects presented in isolation, and that attention biases these weights in favor of the target object. We question whether this biased-competition hypothesis can also account for attentional modulation of semantic representations. To address this question, we recorded participants' BOLD responses while they performed category-based search in natural movies that contained 831 distinct objects and actions. Three different tasks were used: search for "humans", search for "vehicles", and search for "both humans and vehicles" (i.e. divided attention). Voxel-wise category models were fit separately under each task, and voxel-wise semantic tuning profiles were then obtained using a principal components analysis on the model weights. Semantic tuning profiles were compared across the single-target tasks and the divided-attention task. We find that in higher visual cortex a substantial portion of semantic tuning during divided attention can be expressed as a weighted average of the tuning profiles during attention to single targets. We also find that semantic tuning in category-selective regions is biased towards the preferred object category. Overall, these results suggest that the biased-competition theory accounts for attentional modulation of semantic representations during natural visual search.

Keywords: fMRI, visual perception, attention, biased-competition, semantic representation.

ÖZET

KATEGORİ TEMELLİ GÖRSEL TARAMA ESNASINDA BEYİNDEKİ ANLAM TEMSİLLERİNDE OLUŞAN TARAFLI REKABET ETKİLERİ

Mohammad Shahdloo
Elektrik ve Elektronik Mühendisliği, Yüksek Lisans
Tez Danışmanı: Tolga Çukur
Ocak 2017

İnsanlar, görsel alanlarındaki binlerce farklı nesne ve eylem kategorisini algılayabilir ve dikkatlerini birden fazla hedef kategoriye yönlendirebilirler. Önceki çalışmalar, nesne ve eylem kategorilerinin korteks boyunca kesintisiz yerleşmiş anlam haritaları ile beyinde temsil edildiğini ve dikkatin belirli bir kategoriye atanmasının voksellerin seçicilik profillerinde hedef kategorinin temsillini arttıracak şekilde ciddi kaymalara neden olduğunu göstermiştir. Ancak, birden fazla kategoriye aynı anda bölünmüş dikkatin beyindeki anlam temsilleri üzerine etkileri henüz belirlenmemiştir. Yakın zamanlı bulgular, taraflı-rekabet modelinin öngörüleri ile uyumlu şekilde, aynı anda gösterilen iki nesnenin oluşturduğu beyin tepkilerinin iki nesnenin ayrı ayrı gösterilmesinin oluşturduğu tepkilerin ağırlıklı ortalaması olarak ifade edilebileceğini ve dikkatin ortalamadaki ağırlıkları hedef nesne lehine saptırdığını göstermiştir. Bu çalışmada, taraflı-rekabet hipotezinin anlam temsillerindeki dikkat kaynaklı değişimleri de açıklayıp açıklayamayacağı sorgulanmıştır. Bu amaçla katılımcıların beynindeki kan akışına dayanan fonksiyonel manyetik rezonans görüntüleme (fMRG) sinyalleri, içerisinde 831 farklı kategorinin geçtiği doğal filmleri kategori hedefli tarama görevi yaparak izledikleri esnada kaydedilmiştir. Üç farklı tarama görev belirlenmiştir: hedef kategori "insanlar", hedef kategori "araçlar", hedef kategori "insanlar ve araçlar" (bölünmüş-dikkat). İncelenen her bir voksel için çeşitli tarama görevlerine karşılık gelen farklı kategori modelleri geliştirilmiş ve voksel-temelli anlam seçicilik profilleri, model ağırlıklarına temel bileşenler analizi yapılarak elde edilmiştir. Tek hedef ve bölünmüş-dikkat görevlerine ait anlam seçicilik profilleri karşılaştırılmıştır. Bu incelemeler sonucunda yüksek görsel kortekste bölünmüş-dikkat durumunda ortaya çıkan anlam seçiciliğinin ciddi bir kısmının tekli dikkat durumundaki anlam seçiciliklerinin ağırlıklı ortalaması olarak ifade edilebileceği bulunmuştur. Ayrıca,

kategori seçici bölgelerdeki anlam seçiciliklerinin tercih edilen nesne kategorisi lehine saptığı bulunmuştur. Sonuç olarak, elde edilen bulgular taraflı-rekabet hipotezinin doğal görsel tarama durumunda anlam temsillerindeki dikkat kaynaklı değişimlere açıklama getirebileceğini göstermektedir.

Anahtar sözcükler: fMRG, görsel algılama, dikkat, taraflı-rekabet, anlam temsilleri.

To Hasti,
for her endless love and support during my difficult times...

Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Prof. Tolga Çukur for the continuous support of my study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my MSc study.

My sincere thanks also goes to Prof. Ergin Atalar, who provided the highly professional laboratory and research facilities in UMRAM, and to TÜBİTAK for their extensive financial support through TÜBİTAK 3501 program, project No. 114E546. Without their precious support it would not be possible to conduct this research.

Last but not the least, I would like to thank my family: my parents and to my brother for supporting me spiritually throughout my life, and I would like to specially thank the love of my life, Hasti. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past eight years of my life have been built.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Humans are very proficient in extracting most relevant information from cluttered visual scenes. This is due to the human ability to focus attention on specific targets in a scene and to switch it between different targets rapidly. Suppose you are seeing a common daily scene like the one in Figure 1.1. Initially, you might highlight the more salient elements in the scene, like the red car and ignore the people in the sidewalk or advertisement banners. Such "bottom-up" factors are caused by stimulus saliency and influence the perception of the scene by reflecting sensory inputs. On the other hand, if you are supposed to find your friend in the crowd, you might very well ignore the cars and bicycles and search for the people in the sidewalk in a finer detail. During this process, ones cognitive state enhances the targets percept by modulating the neural responses in various areas of the brain. These "top-down" factors, refine neural responses to serve the requirements of the current and future goals to optimize the processing in order to do the planned task [12]. In everyday life, visual perception is mediated by a mixture of bottom-up and top-down factors and understanding the governing dynamics of this interaction plays an important role in our understanding of the visual perception process. In the following sections, the process of visual perception and its attentional modulations will be introduced. We will focus specifically on attentional modulation of semantic representations across the human brain during search for object categories, the main topic of study in this thesis.

Figure 1.1: **Visual perception of a cluttered scene.** Visual perception is mediated by both "bottom-up" factors, reflecting the salient stimulus features, and "top-down" factors, addressing the requirements of processing the cognitive state. (photo courtesy: Aaron Bialick)

1.1 Visual perception

> *"The eye is like a mirror, and the visible object is like the thing reflected in the mirror."* – Avicenna, early 11th century

Since ancient times, eye was believed to be the organ enabling us to "see" the world. But it was only after Kepler at seventeenth century that image reflection on the retina became widely accepted as the major mean of vision. Our understanding of the vision process has grown from then on. One of the major means of interpreting surrounding environment is processing the information in visible light. At the first step of the process of visual perception, illumination and color of the image is captured by photoreceptors. Photoreceptors are specialized cells on the retina that are sensitive to various properties of light. This information is projected through the optic nerve to the most posterior area of the cortex known as the lower visual cortex. Lower visual cortex is divided into multiple functional and anatomically segregated areas [34]. First area involved in early visual processing is the primary visual cortex (also known as V1). Neurons in V1 are sensitive to simple image features. Single-neuron experiments have revealed that different types of neurons in V1 respond to bars with various angles and to bars moving in various directions [34]. There is also a well-defined spatial mapping of the visual field in V1 [56].

Second major visual processing area in the cortex is the prestriate cortex (also known as secondary visual cortex or V2). Neurons in V2 respond to more complex features such as orientation and illusionary contours which are particularly important in visual association [34]. An important function of V2 is that it helps identify whether a piece of visual stimuli is part of an object or not, and to segregate objects from background [24]. V2 neurons receive strong connections from V1 and send strong connections to V3 and V4. V3 (also known as the third visual complex) is known to be an important cortical area enabling one to have a perception of depth [60]. V4 (also known as the fourth visual complex) is widely known for its role in processing orientation, spatial frequency and color, similar to V2 but in a much more complex manner. In particular, high selectivity to color is a major functional difference between V4 and V2. V4 also plays role in perception of geometric shapes [46]. There are multiple studies suggesting that visual processing is divided into two main paths beyond V4 [26, 23].

Ventral pathway (the "what" pathway) consisting of V4 and extends to inferior temporal cortex (IT). Cortical areas in the "what" pathway play a major role in perception of objects within visual scenes. Inferio-temporal cortex includes various functional sub-regions, among which are FFA, PPA, EBA and LO. Fusiform face area (FFA) located on the ventral surface of the temporal lobe, responds strongly to faces. The idea of a specific cortical region responsive to faces was first introduced by Kanwisher et al. [35]. Generally, FFA is localized as voxels[1] in lateral side of fusiform gyrus which show significant increased response while subjects view images of faces versus while they see images of general objects. Parahippocampal place area (PPA) is a part of the parahippocampal cortex that resides in medial inferior temporal cortex. PPA was first introduced by Epstein and Kanwisher [21] as a cortical area selective for scenery. PPA is localized to voxels in the parahippocampal cortex, significantly responding to images of scenery versus images of general objects. Extrastriate body area (EBA) is an area selective for human body parts. EBA is located in the lateral occipitotemporal cortex and is localized as voxels which show significant response to human body parts versus images of general objects [17]. There is also an area in the lateral occipital

[1]The term "voxel" used throughout this thesis is the short hand form of "volumetric pixel", the smallest volume unit in the acquired functional data.

cortex (LO) which strongly responds to general objects with a clear 3D shape [37]. The object-selective area LO, is commonly localized to voxels in lateral occipital cortex that show significant response to images of objects versus scrambled texture images. A major observation regarding the selectivity of the cortical areas along ventral stream is that the complexity of the information processed by the neurons increases as we go from retinotopic areas toward object-selective regions.

The dorsal stream (the "where" pathway) appears to have the role of processing visual motion and visual control of attention. The main dorsal cortical area known to be responsive to moving stimuli is MT area, located in the middle temporal lobe. It sends strong connections to intraparietal sulcus (IPS) and to areas in prefrontal cortex, such as frontal eye field (FEF) and frontal operculum (FO). These areas are believed to function in controlling saccadic eye movements and producing top-down attentional modulating signals [40, 41], thus they are known as part of an attention-control network.

1.2 Attentional modulations of visual perception

Attention plays a crucial role in dedicating the limited resources of the brain in a way to process the cognitive demands and the external triggers. According to the goal of attention, it may be categorized into "spatial attention", "feature-based attention" and "object-based attention".
Various single-cell studies have revealed different aspects of neuronal activity enhancements due to attention. Prinzmetal et al. [49] have reported that observers detect faint stimuli better if the stimuli appear at an attended location. This is the consequence of both lowered neural firing threshold (i.e. enhanced sensitivity) and increased firing rate as a result of spatial attention. These observations are reported in various single cell studies [53, 52] and are consistent throughout the visual system [61, 55].

Cognitive enhancements of covert attention (i.e. directing spatial attention without changing the direction of gaze) are characterized as enhancing visual sensitivity and reducing reaction times. A recent study by Pestilli and Carrasco [48] showed that covert cued attention to the left or right of a fixation point enhances sensitivity to a flashing object presented for a small period of time. That is, subjects were more successful at correctly identifying presence of a flashing target in either side of the fixation point, if it was preceded by a directional cue shown at the fixation point. In another cognitive experiment, Prinzmetal et al. [49] had presented flashing stimuli in two sides of the fixation point and recorded reaction time in detecting a target, with and without presence of a spatial cue in the fixation point, employing the recording from a button pressed by subject if he detects the target. They have shown that reaction times are lowered if the flashing stimuli is preceded by a spatial cue. Top-down modulations also rise from object-based attention. For example, attending to humans in a scene triggers neural activations in attention-control cortical areas which in turn modulate the activation in cortical areas selective for humans in favor of processing human features [55].

1.2.1 Semantic representation of scenes and its attentional modulation

Humans are able to perceive thousands of distinct object and action categories in visual scenes. So it is highly unlikely that each category be represented in an individual area in the cortex. A recent study has proposed that a wide range of actions and object categories are represented in a low-dimensional semantic space across the cortex [31]. Using an experiment in which subjects passively viewed natural movies while their brain activity was recorded using fMRI, Huth et al. [31] have showed that object categories are embedded in a four-dimensional semantic space, shared across subjects. They state that semantically similar categories are projected close to each other and semantically different categories are projected to distinct locations on the cortex. It is also shown that attention to an object category, warps semantic representation across the cortex in favor of the

attended target by expanding its representation across cortex [9]. According to these studies, representation of object categories that are semantically related to the target are also expanded. This suggests that attention dynamically modulates cortical representation in favor of the processing of behaviorally relevant objects during natural vision.

1.2.2 Different accounts for divided attention

Behavioral effects of attending to multiple items has been a subject of debate in recent years. A large body of studies from both neuro-imaging and single cell perspectives have reported that task performance is decreased when the number of items to be attended are increased [44, 45, 20]. In other words, when the observer is about to pick the target among distractors, it takes more time, and the selection error rate across trials is increased as the number of items present on the screen increases [44]. There are mainly two theories presented to account for this lowering in performance.

According to "decision integration model", quality of representation for each item is not affected as the number of attended items increase, but rather the divided attention results in an increase in the error rate in target detection [20]. This account assumes that the observer has noisy representations of all of the targets and distractors and considers the attention mechanism as a target detection task [58]. In this view, each item in the field of view elicits an internal response in the observer plus a random internal noise due to uncertainty in the neuronal activity. Because of this internal noise, the observer might misidentify a distractor as the target, resulting in a decrease in detection performance and an increase in reaction time [58]. By running experiments in which human subjects searched for an ellipse with multiple different features (i.e. color, orientation, size) among multiple other ellipses, Palmer [44] developed a model that succeeds in showing the predictions of the decision integration model using signal detection theory.

On the other hand, the perceptual coding (also known as "biased competition") hypothesis assumes that since the brain has limited computational power,

dividing attention between multiple objects limits perceptual resources and their representation quality is lowered compared to when each of them are the sole search target [19]. The biased competition theory is based on three main principles. First, representation of visual information in different visual systems (sensory and motor, cortical and subcortical) in the brain is "competitive". Second, this competition is "controlled" meaning that it can be biased in favor of one the multiple competing stimuli. Third, the competition is retained across systems. If some portion of the stimulus gains dominance in one system (e.g. early visual cortex) it will retain its dominance in other systems (e.g. higher order visual areas) as well [4].

Several neuro-imaging studies have based their findings on the ground of biased-competition paradigm [36, 57, 51, 14, 38, 3, 2]. Beck and Kastner [2] have studied the effect of attention to Gabor patches with specific orientations among multiple patches with different orientations. Their findings suggest that neurons in early visual areas evoke less response when multiple items appear simultaneously, as a sign for the divided attention being "competitive". They propose that this effect is more prominent as we go from lower to higher stages of visual processing (i.e. from V1 to V4). There is also strong evidence stating that origin of the top-down signals altering the bias in the biased competition are areas outside visual cortex, including IPS, FEF, SEF and FO [7, 30, 12].

One recent study by Reddy et al. [51] has assessed the biased-competition theory using BOLD responses recorded while subjects viewed static images of isolated objects and attended to either a single object category among four possible choices (faces, houses, shoes, cars), or two categories simultaneously. A multi-voxel pattern analysis on BOLD responses recorded in category-selective areas in ventral-temporal cortex suggested that response patterns elicited during divided attention to a pair of categories can be expressed as a weighted combination of the response patterns elicited during attention to each constituent category alone. In PPA and FFA, the weights for this combination showed a significant bias towards the preferred object category [51]. These results can be taken to imply that category-selective regions in ventral-temporal cortex retain highly modular representations of preferred categories under divided attention. In another study, Reddy and Kanwisher [50] were able to decode object categories present in the

visual stimuli, from the cortical BOLD responses in FFA and PPA in various clutter levels and in the presence of divided attention. This shows that representation of the preferred object categories in object-selective areas FFA and PPA, is robust in presence of clutter and diverted attention.

In this thesis we investigated the effects of divided attention on semantic representation across the human brain. We hypothesized that divided attention among multiple categories modulates semantic representation in accordance with the biased-competition theory. We recorded whole-brain BOLD responses while subjects viewed a diverse selection of natural movies and performed category-based search tasks. In the divided-attention task subjects searched for "humans" and "vehicles", and in the two isolated-attention tasks they searched for "humans" and for "vehicles" individually. We used a voxel-wise modeling framework to measure category-tuning profiles of single voxels in each individual subject and under each attention condition [42]. We then evaluated the biased-competition theory by testing two key predictions. First, we asked whether semantic-tuning profile of a voxel during divided attention can be expressed as a weighted average of semantic-tuning profiles during the two isolated-attention conditions. We then asked whether the weights for this linear weighted average show significant bias towards a preferred category in regions across ventral-temporal cortex.
In chapter 2 we introduce some basic concepts needed for the general reader to have a grasp on the materials provided throughout the thesis, including the experimental methods, analytical basis and practical considerations taken to reach the results. Chapter 3 includes the results derived from the study and a discussion following the conclusion of the results is provided in chapter 4.

Chapter 2

Materials and Methods

2.1 Behavioral task

There have been many studies of visual attention using still images or simple stimuli like moving bars. Although these simple stimuli provide valuable insight into basic properties of visual perception, they lack the complexity of the visual stimuli human beings encounter in daily life. Another problem in experiments with simple stimuli is that the position of stimuli is usually known beforehand. Although some studies have used random positions for placing the stimuli in the receptive field they still lack the generality of the cluttered natural scenes in which multiple objects appear in various locations in the visual field. In this thesis we have used natural movie clips including a diverse set of objects under clutter and in various contexts to overcome the limitations of the above mentioned simplified experimental paradigms.

Subjects viewed a continuous, color natural movie compiled from 10-20 second clips extracted from Google, YouTube videos. A broad range of indoor and outdoor scenes including 831 object and action categories were present in the clips. The stimulus was separated into three 8-minute-long blocks, and each block was shown in a separate run. Subjects performed three separate visual-search tasks while viewing the same movie stimulus: visual search for “humans”, visual

search for "vehicles", and visual search for both "humans" and "vehicles". The experiment was performed in a total of 9 runs and the search task was interleaved across runs. To ensure subject vigilance, subjects were asked to respond with a button press whenever a target was present in the stimulus. Button-press responses were analyzed to ensure that the subjects were performing the search task and to assess any major difference in task difficulties.

2.2 Data collection

2.2.1 Subjects

Five human subjects (four males and one female) have participated in the experiment. The experimental data analyzed in this thesis was collected at the University of California, Berkeley.

2.2.2 Recording the brain activity

Functional magnetic resonance imaging (fMRI) is a noninvasive modality to assess neural activity in the brain, indirectly through changes in blood flow and blood oxygenation using magnetic resonance imaging (MRI). When a neural population in the brain gets activated, blood circulation and the amount of oxygen in the blood in the activated area increases. The protein responsible for transporting oxygen in the blood is Hemoglobin. Magnetic properties of Hemoglobin change when it gets oxygenated. It is diamagnetic when oxygenated but gets paramagnetic when deoxygenated. Thus, fluctuations in the oxygenation level in a brain area is related to minor changes in its magnetic properties. These fluctuations in magnetic properties lead to small changes (typically not more than 4 percent [54]) in the MR signal which is measured as a proxy for neuronal activity. This signal is generally called the blood oxygenation level dependent (BOLD) signal.

2.2.2.1 Hemodynamic Response

The BOLD signal is sensitive to the local cerebral blood flow, the rate of oxygen consumption by the activated neurons (also known as "cerebral metabolic rate of oxygen"), and the local cerebral blood volume [8]. The overall effect of these factors that relate the neuronal activity with BOLD signal is often called the hemodynamic response. The BOLD signal lags neural activation by 1-2 seconds and have a temporal width of 5-6 seconds. It commonly ends with an undershoot [8].

Buxton et al. [8] have proposed a simple computational model which explains dynamics of the BOLD signal change, reflecting the undergoing physiological processes. A simplified description of the model which relates BOLD signal change to local blood volume ($\frac{V_{act}}{V_0}$) and oxygenation($\frac{E_{act}}{E_0}$) changes can be expressed as

$$\frac{\Delta S}{S_0} \approx C.\left[1 - \frac{V_{act}}{V_0}(\frac{E_{act}}{E_0})^{\beta}\right] \tag{2.1}$$

in which $\frac{\Delta S}{S_0}$ is the MR signal modulation, C is a locally constant scaling term and β lumps the effects of the MR scanner magnetic fields. The aforementioned dynamics of the BOLD signal is due to the varying dynamics of local blood volume and oxygenation during the aftermath of the neural activation.

Figure 2.1 illustrates the shape of a typical hemodynamic response function (HRF). The HRF resulting from two separate brief stimuli results in the plot in Figure 2.2 in which two consequent peaks are the result of convolving the HRF with the sequence of stimuli. In fact, HRF functions as a low-pass filter, smoothing the underlying neuronal firings, thus, it limits the temporal resolution of the recorded BOLD signal. This effect is illustrated in Figure 2.3. To avoid unnecessary complexities, we have modeled the hemodynamic response using an FIR filter, which will be discussed in later sections.

In this study, we used functional MRI to measure brain activity while subjects performed the requested tasks. MRI data were assessed on a 3T Siemens scanner located at the University of California, Berkeley using a 32-channel receiver coil. Functional data were acquired using a T2*-weighted gradient-echo EPI pulse sequence using the following parameters: repetition time(TR) = 2s, echo time(TE)

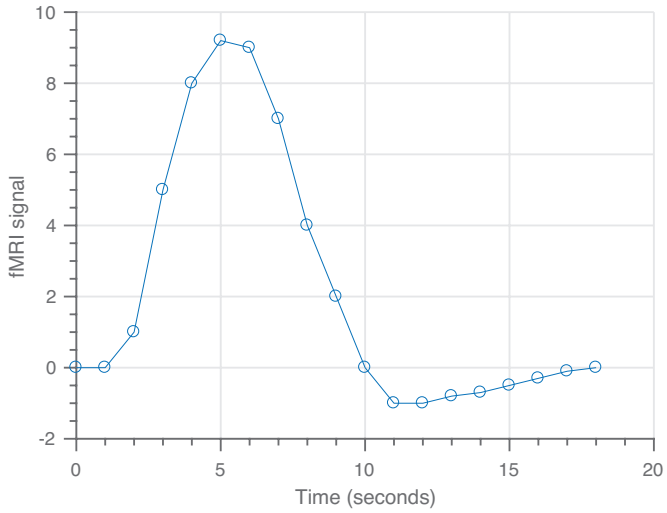

Figure 2.1: **Typical shape of the hemodynamic response filter.** Hemodynamic response incorporates a time lag in the output, beside the smoothing it produces due to its limited frequency width.

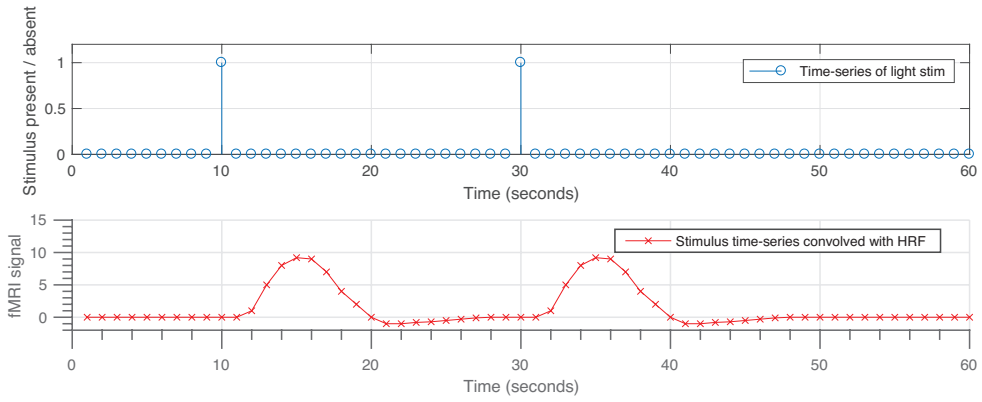

Figure 2.2: **Hemodynamic response resulting from two consecutive brief stimuli.** Hemodynamic response filter smooths out the high-frequency content of the input and incorporates delays. This results in two slow varying peaks recorded using fMRI.

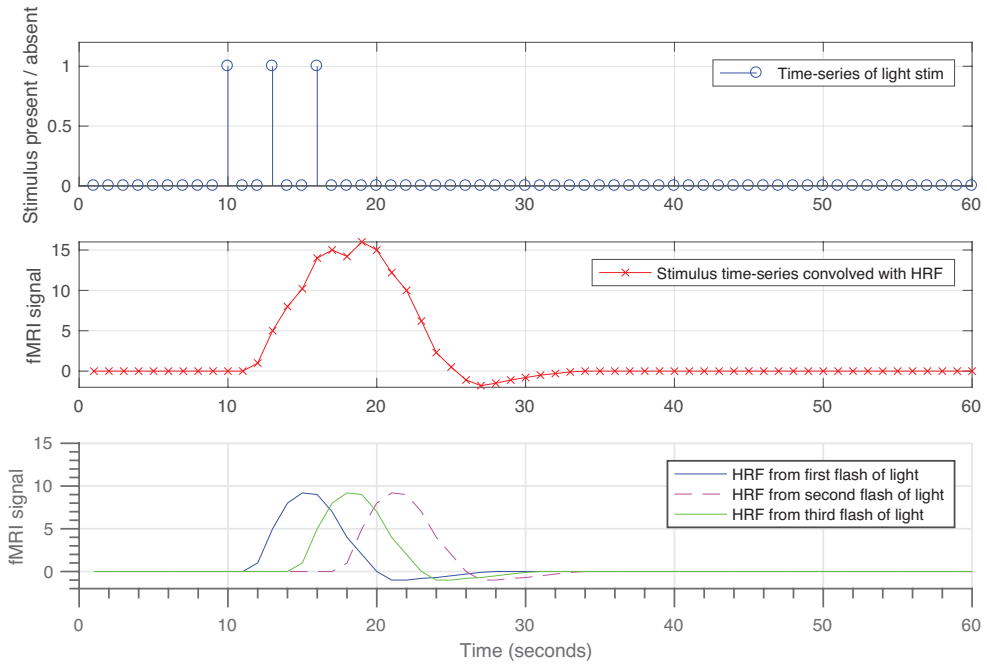

Figure 2.3: **Hemodynamic response leads to limited temporal resolution in BOLD response.** Due to low-pass filtering effect of the HRF, neural firings resulting from closely presented stimuli are represented as a single peak in the BOLD signal, this leads to limited temporal resolution of the recorded BOLD signal.

= 34ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm^3, field of view = 224 × 224 mm^2. 32 axial slices were acquired to cover the entire cortex. To create cortical flat maps we used a three-dimensional T1-weighted MP-RAGE sequence to collect anatomical data with 1 × 1 × 1 mm^3 voxel size and 256 × 212 × 256 mm^3 field of view.

2.3 Data preprocessing

A collection of steps need to be performed in order to make an fMRI BOLD response ready for analysis. The preprocessing steps employed in this study are motion correction, compensation for low frequency drifts and brain tissue extraction.

2.3.1 Motion correction

Motion correction (also known as "realignment") is a process to correct the head motion effects on the acquired functional data. Even slight head movements cause voxel displacements and changes the signal received from a voxel, reducing the quality of the data. Motion correction tries to correct the movements by aligning different image frames across the fMRI series to a reference volume image, using a rigid body affine transformation. The reference image could be any single image in the series, or the mean image of the series. Head movements are characterized by six variables, three translation parameters, corresponding to linear translation in three Cartesian axes, and three rotation parameters, corresponding to amount of rotation around the three axes. Motion correction in this study was performed using the tools in SPM12 package (http://www.fil.ion.ucl.ac.uk/spm/), and using the first image frame as the reference in each session. Motion correction might lead to flips or sudden rotations of the brain tissue in some cases. Thus, to ensure the validity of realignment, motion-corrected time series was inspected manually for any flaws in each session separately.

2.3.2 De-trending

There are low frequency drifts present in the BOLD signal due to physiological activity (e.g. respiration and cardiac activities). There are also slow drifts due to the scanner warming and eddy currents which incorporate modulations in MRI scanner magnetic fields during the experiment. To compensate for these effects, we filtered the fMRI time series using a second order Savitzky-Golay filter with 120 second window. De-trended responses for each voxel were then z-scored to attain zero mean and unit variance.

2.3.3 Brain extraction

fMRI provides us a 3D image of the brain activation in each time instance, but it includes other tissues such as skull as well. So the brain tissue should be segmented out before moving further. There are various tools used in neuroscience studies for segmentation and extraction of the brain tissue. We used "brain extraction tool (Bet)" from FSL 5.0 package for this purpose [32]. As the result of the brain extraction, we ended up with roughly 50000 to 80000 brain voxels for the subjects of this study.

2.4 Functional regions of interest

Functional regions of interest (ROIs) are generally identified using "localizer" scans to identify voxels in specific anatomical areas showing a particular response (as described in section 1.1). The ROIs of study are determined as the cortical voxels which show significance in the localizer scans plus a neighborhood of voxels in their 2mm vicinity in depth of the brain. Broad ROI boundaries were drawn on the cortical flat maps. Because of the differences in signal-to-noise-ratio values of recorded fMRI data among individual subjects, the broad regions of interest were shrunk to include only voxels having contrast response above half of maximum

response across cortex in each subject individually. Lastly, voxels in the 2mm depth vicinity of shrunk cortical ROI voxels were included. Table 2.4 summarizes the localizer experiments used to identify each ROI.

ROI name	**Anatomical location**	**Localizer experiment**
FFA	Posterior fusiform gyrus	Faces - Objects [35]
PPA	Collateral fissure	Places - Objects [21]
EBA	Anterior to MT+ on the medial temporal gyrus	Body parts - Objects [17]
RSC	Medial wall just superior to PPA	Places - Objects [1]
LO	Anterior to V4	Objects - Scrambled Texture [27]
OPA	Anterior to V4v/VO	Places - Objects [28]
IPS	Lateral parietal cortex	Self-generated Saccades [10]
FEF	Superior frontal sulcus	Self-generated Saccades [47]
FO	Precentral sulcus	Self-generated Saccades [13]

Table 2.1: **Functional ROI locations and localizers.** Functional ROIs studied in this thesis were identified using localizer scans performed independently of the main experiment.

Localizer experiments for category-selective areas were performed in six 4.5min runs of 16 blocks. In each block, subjects passively viewed 20 static images randomly selected from one of the "objects", "scenes", "faces", "body parts" and "spatially scrambled objects" for a total of 16 seconds [27]. Sample images used in localizer scans is shown in Figure 2.4. Each image was shown for 300ms following a 500ms blank. To localize attention-control areas, one 10min run of 30 blocks was used. Either a self-generated saccade, in which the subject was asked to follow a target pattern, or a blank period was prescribed in each 20sec block [10].

Figure 2.4: **Sample images used in localizer scans.** To localize category-selective areas, 20 static images randomly selected from one of the "objects", "scenes", "faces", "body parts" and "spatially scrambled objects" were shown for a total of 16 seconds in each block.

2.5 Voxel-wise modeling

2.5.1 Category model

A voxel-wise model was fit to the data to describes the contribution of different object and action categories to the recorded BOLD activity [9]. Each one second of the clips were labeled for the presence of objects (i.e. nouns) and actions (i.e. verbs) in them, using a collection of terms present in the WordNet lexicon [39]. WordNet is a set of hierarchical directed graphs assigning semantic relationships among words. For example, "entity $\rightarrow$ person $\rightarrow$ child" is a branch in the WordNet graph, indicating that a "child" is a "person" and thereby an "entity". Presence of superordinate categories were also inferred using the hierarchical relations (e.g. whenever a "man" is present, a "human" is also present). Different labeled movie frames are shown in Table 2.2. Using labeled frames, a "stimulus matrix" was built and filled with binary elements, in which each element represents the presence of the corresponding category in the corresponding time instance in the movie clip. Procedure details are explained in Chapter 2.
The stimulus matrix was temporally down-sampled by a factor of two to match

the sampling rate of the fMRI scans.

Movie Frame	Labels	Movie Frame	Labels
	• insect.n.01 • leaf.n.01 • tree.n.01		• snow.n.01 • snowmobile.n.01 • jump.v.01 • ski-jump.n.01 • man.n.01
	• man.n.01 • truck.n.01 • run.v.01 • policeman.n.01 • grass.n.01		• small-boat.n.01 • man.n.01 • woman.n.01 • flower.n.01 • tree.n.01 • sail.v.01

Table 2.2: **Labeling movie frames for presence of object and action categories.** Each one second frame of the movie clip was labeled using the words from WordNet lexicon providing a total of 831 object and action categories. Categories were then used to construct the stimulus matrix.

2.6 Hemodynamic response

Recalling that what we measure as the BOLD response is an aftermath of neural firings rather than direct neural activity, the hemodynamic response of the brain that maps neural activity to the BOLD signal should be taken into account. To consider the delay resulting from the hemodynamic response, we concatenated 4, 6 and 8 seconds delayed stimulus vectors. Under this scheme, these vectors indicate

presence of categories four, six and eight seconds earlier. This is equivalent to convolution of the stimulus matrix with a three-tap finite impulse response filter.

2.7 Nuisance regressors

2.7.1 Head-motion and physiological noise

Although passing the data through preprocessing steps prunes it for some of the high-frequency noise and slow drifts, we further compensated for residual physiological-noise and head-motion effects by regressing out nuisance regressors. The cardiac and respiratory activity were collected using a pulse oximeter and a pneumatic belt during the runs. They was used to estimate two regressors to capture respiration and nine regressors to account for cardiac activity [59]. Affine motion parameters estimated from momentarily head motions during the motion-correction preprocessing stage were taken as the head-motion regressors. The collection of these regressors were used to regress out the head-motion and physiological-noise modulations from the BOLD responses.

2.7.2 Motion energy correlation

To account for spurious correlation between motion-energy and object-action categories in the natural movies, a nuisance regressor that describes the total motion energy during each one-second of the movie clip was employed as described in [42]. This regressor was formed by taking the mean energy across each one second frame of the movie clip, calculated by applying multiple Gabor filters with different orientation and scales to the frame. The Gabor filter bank consisted of several stages of filtering and transformation. First stage comprised of transforming the stimuli into the International Commission on Illumination LAB color space and removing the color channel, retaining the luminance information. Luminance

channel was then passed through 6555 spatiotemporal Gabor filters with different scales, orientations and spatial and temporal frequencies. At last, the Gabor motion-energy was assessed by squaring and summing outputs of quadrature filter pairs, and the results passed through a logarithm compressive nonlinearity and temporally down-sampled to the fMRI acquisition rate [42]. The resulting time series was taken as the motion-energy regressor.

2.8 Category weights

Category weights were assessed using regularized ridge regression (also known as "Tikhonov Regularization") of stimulus matrix and the recorded voxel-wise BOLD signals. Figure 2.5 illustrates the process of assessing category weights. The optimal regularization parameter for each voxel was picked by doing a 20-fold cross validation. In each fold, data was randomly split such that 90 percent of time samples were used to find weights using different regularization parameters in the range $\lambda_i = [1, 10 \times 2^{20}]$ and the Pearsons correlation coefficient of predicted and recorded BOLD responses in the remaining 10 percent of samples was taken as the prediction score. Cross-validation was performed for each of the candidate regularization parameters (λ_i) and the regularization parameter that maximizes the mean prediction score across 20 folds was picked as the optimal value (see Appendix A). After picking the optimal regularization parameter for each voxel, voxel-wise category weights can be assessed by solving the equation:

$$\hat{w}_i = (S^T S + \Gamma^T \Gamma)^{-1} S^T r_i \tag{2.2}$$

in which S is the stimulus matrix containing time series of presence of each category in rows, Γ is the diagonal matrix containing voxels-wise optimal regularization parameters, r_i is the matrix containing voxel-wise BOLD response in rows and $\hat{w}_i$ is the matrix of voxel-wise category weight vectors for $i = H, V, B$ indicating "attend to humans" (H condition), "attend to vehicles" (V condition) and "attend to both humans and vehicles" (B condition). Category weight matrices assessed in this modeling framework describe the contribution of each of the object and action categories to the voxel BOLD responses.

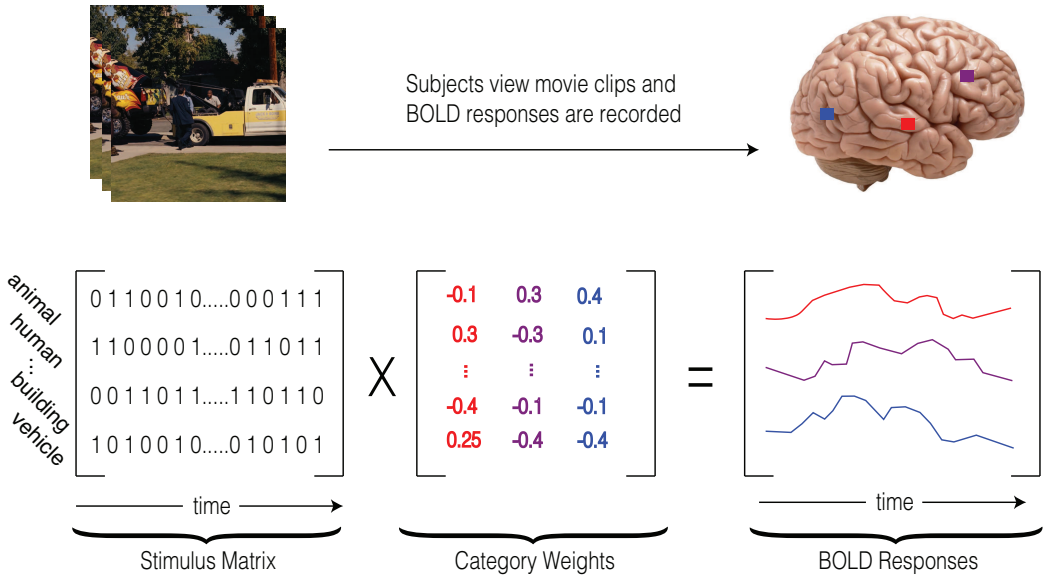

Figure 2.5: **General procedure of obtaining model weights.** Voxel-wise category weights were assessed using regularized ridge regression of stimulus matrix and voxel-wise BOLD responses. Each element in the category weight matrix accounts for the contribution of an specific category in the voxel BOLD response.

2.8.1 Target detection effect

To account for spurious correlation between BOLD response and target detection, a target regressor was used indicating the presence of the target in each one-second frame of the movie stimulus. This target regressor was equivalent to the "human" category regressor for the search for "humans" task, and to the "vehicle" category regressor for the search for "vehicles" task. For the divided attention task, the target regressor was taken as the union of the binary "human" and "vehicle" category regressors. To regress out effects of target detection from BOLD responses, the stimulus matrices were aggregated across the three search tasks, and model fits were performed on this aggregated stimulus matrix (Figure 2.6). Using the aggregated stimulus and response matrices, the model can be

described as

$$S_{agr} = \left[\begin{array}{c} diag(S,S,S) \\ \hdashline T \end{array}\right] \tag{2.3}$$

$$T = \left[\begin{array}{c:c:c} s_H & s_H \oplus s_V & s_V \end{array}\right] \tag{2.4}$$

$$R_{agr} = \left[\begin{array}{c} r_H \\ \hdashline r_B \\ \hdashline r_V \end{array}\right] \tag{2.5}$$

$$\hat{W}_{agr} = \left[\begin{array}{c} \hat{w}_H \\ \hdashline \hat{w}_B \\ \hdashline \hat{w}_V \end{array}\right] = (S_{agr}^T S_{agr} + \Gamma^T \Gamma)^{-1} S_{agr}^T R_{agr} \tag{2.6}$$

in which s_H and s_V are the rows from the stimulus matrix S corresponding to the "human" and the "vehicle" categories respectively and $\oplus$ performs the binary *OR* operation. r_H, r_V and r_B are the BOLD responses collected in the three search tasks and $\hat{W}_{agr}$ stores the model weights for the search tasks.

2.9 Projection onto semantic space

To estimate a semantic space of category representation, principal components analysis (PCA) was performed on voxel-wise category weights in each subject. Principal components analysis projects voxel category weights (category-tuning vectors) onto a space in which data points are uncorrelated. Thus, semantically similar categories are projected to nearby points in the semantic space, and the categories with less semantic similarity are projected to distant points. To avoid overfitting, we performed PCA on the model weights pooled across the two single-target attention conditions. Due to existence of multiply many object categories that humans can perceive, the real semantic representation underlying category tuning profiles in the brain is high-dimensional. Thus, the collection of principal components (PCs) that explain at least 90 percent of the variance in the data were selected for further analysis in each subject. This gave between 36 to 47 principal components for the five subjects. Semantic-tuning profiles were then

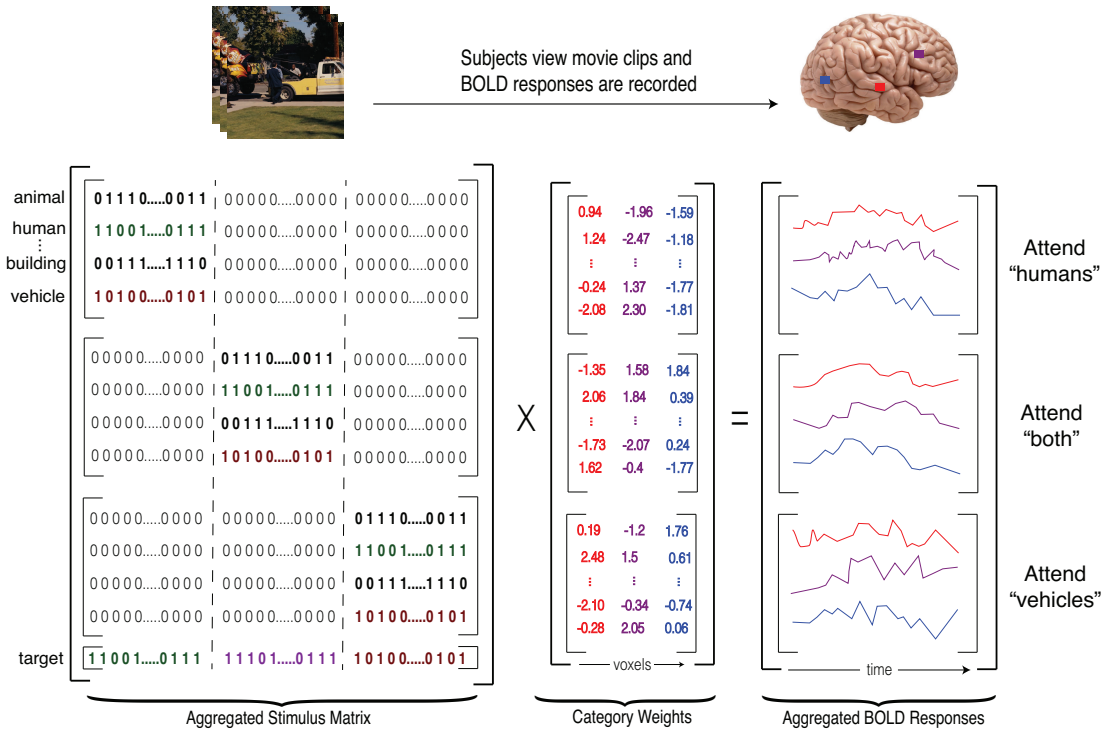

Figure 2.6: **Control for target detection effects.** To account for the possible modulation of BOLD responses by target detection, a target regressor was added to the model. Target regressor was equivalent to the "human" regressor for the "search for humans" task, the "vehicle" regressor for the "search for vehicles" task and the binary union of the "humans" and the "vehicle" regressors for the "search for both humans and vehicles" task. By populating the aggregated category model, model fitting was performed simultaneously for the three attention conditions and target detection effects were regressed out of the BOLD responses.

obtained by projecting category models for each task onto the semantic space defined by these PCs.

2.10 Linearity analysis

We tested whether the semantic-tuning profile during divided attention can be described as a weighted average of the semantic-tuning profiles during single-target tasks. We performed an ordinary least-squares analysis among semantic-tuning profiles in three attention conditions to linearly describe the tuning profile in divided-attention condition using tuning profiles in the single-target attention conditions. The problem can be formulated as

$$w_{P_B} = [1 \quad w_{P_H} \quad w_{P_V}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_1 \end{bmatrix} = XB \tag{2.7}$$

in which w_{P_B}, w_{P_H} and w_{P_V} are the voxel-wise semantic tuning profiles for the B condition, H condition and V condition respectively. Regression weights β_1 , β_2 and β_3 can be assessed by solving the equation

$$\hat{B} = (X^T X)^{-1} X^T w_{P_B} \tag{2.8}$$

and a prediction of w_{P_B} results from

$$\hat{w}_{P_B} = X\hat{B} \tag{2.9}$$

We took the Pearsons correlation between predicted and actual semantic-tuning profiles in divided attention condition (w_{P_B}), as the voxel-wise linearity index. Figure 2.7 illustrates the procedure of obtaining the linearity index.

Higher linearity index means that the tuning profiles in divided attention con-

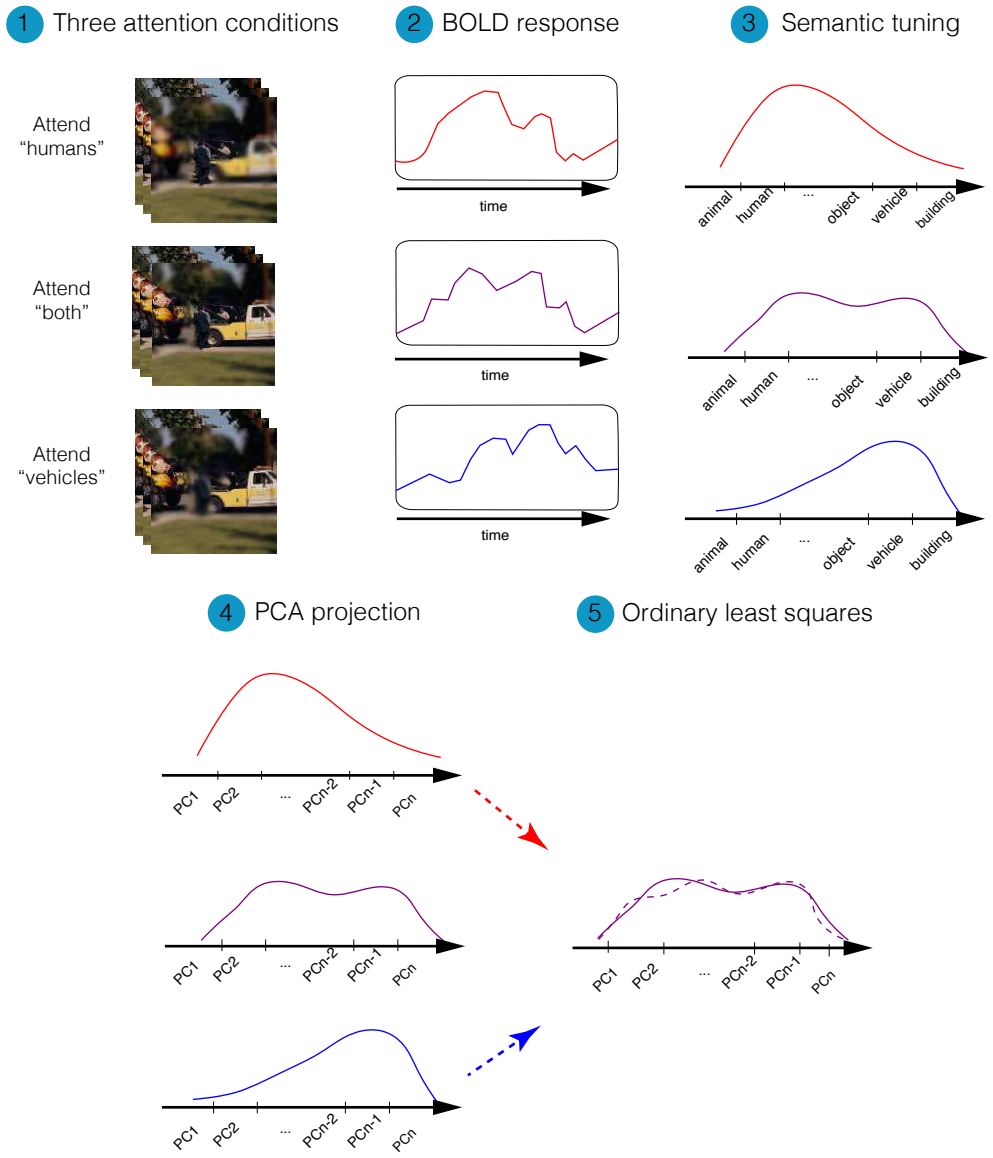

Figure 2.7: **Linearity index calculation procedure.** Voxel-wise linearity index was measured based on the correlation of the predicted and actual semantic tuning profile in divided attention condition (B condition). Tuning in B condition was predicted using an ordinary least-square on tuning profiles in attend to "humans" (H condition) and attend to "vehicles" (V condition). Dashed lines indicate predicted and solid line indicates measured tuning profiles. Thus, the linearity index indicates the accuracy in linear prediction of semantic-tuning profiles of divided attention condition using semantic-tuning profiles in the isolated attention conditions.

dition can be better described as a weighted average of tuning profiles in isolated attention conditions.

2.11 Category weight shift analysis

We have hypothesized that the weights in the weighted average are biased towards the preferred object category in object-selective cortical areas. To put this hypothesis into test quantitatively, multi-voxel distribution of semantic tuning was examined. Specifically, the tuning distribution during divided attention was regressed onto the tuning distributions during the two single-target tasks. We then computed a shift index to measure the direction and magnitude of attentional bias in semantic tuning. The procedure is illustrated in Figure 2.8. In this analysis, we projected the semantic-tuning pattern during divided attention condition onto the hyperplane defined by the tuning distributions during the single-target attention conditions. This projection was performed via ordinary least-squares.

Projections were obtained by solving the problem stated in Equation 2.10 using ordinary least squares:

$$\vec{B} = w_h\hat{H} + w_v\hat{V} + w_c \tag{2.10}$$

in which $\hat{V}$ and $\hat{H}$ are the unit vectors in the direction of $\vec{V}$ and $\vec{H}$ respectively (see Figure 2.8). w_h, w_v and w_c are are the regression weights indicating the contribution of $\vec{V}$ and $\vec{H}$ in the linear combination respectively.

We quantified a normalized "Shift Index" as

$$SI = \frac{w_h - w_v}{|w_h| + |w_v|} \tag{2.11}$$

Positive SI values means that the tuning distribution during divided attention is biased towards the distribution in the attend to "humans" condition and negative values indicate bias towards the distribution in the attend to "vehicles" condition.

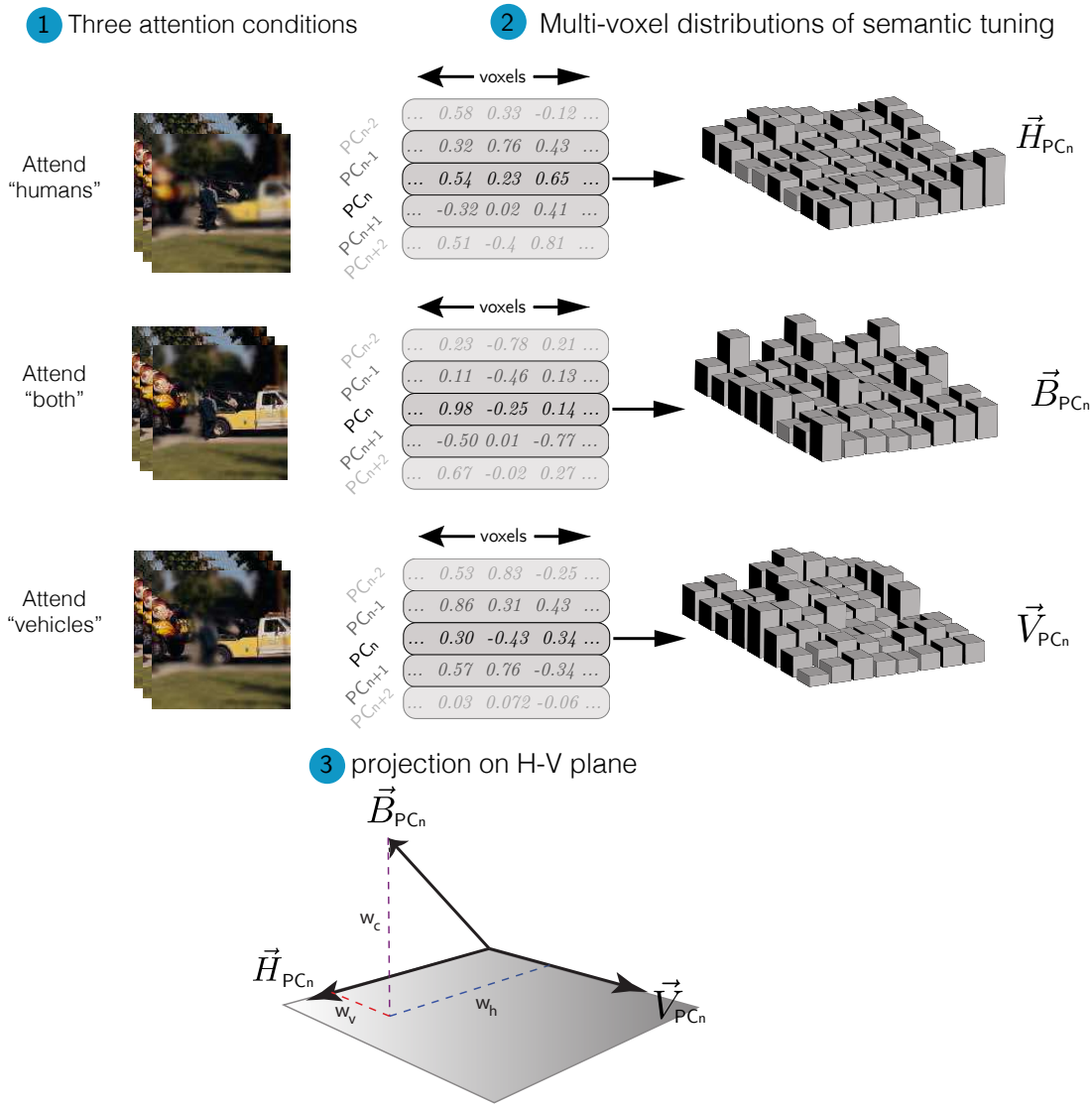

Figure 2.8: **Shift index calculation procedure.** Semantic tuning profile distribution during divided attention was regressed onto the tuning distributions during the two single-target tasks. Regression weights determine the bias in tuning distribution during divided attention towards any of the tuning distributions in the single-target attention tasks. Shift index is then calculated using the regression weights to quantify the bias.

Chapter 3

Results

3.1 Experiment difficulty and subject vigilance

The results of the button press task show that subjects were successful in detecting targets with rates 91%, 90% and 82% in the "search for humans", "search for vehicles" and "search for both humans and vehicles" tasks. Successful hit-rate for attend to "humans", to "vehicles" and to "both" tasks were 0.90 ± 0.10, 0.90 ± 0.09 and 0.82 ± 0.11 respectively. There were no significant differences (one-way ANOVA, $F(2) = 1.01, p = 0.3919$) in the correct hit rates for the three search tasks. This result ensures that following results are not a mere consequence of difference in experiment difficulty.

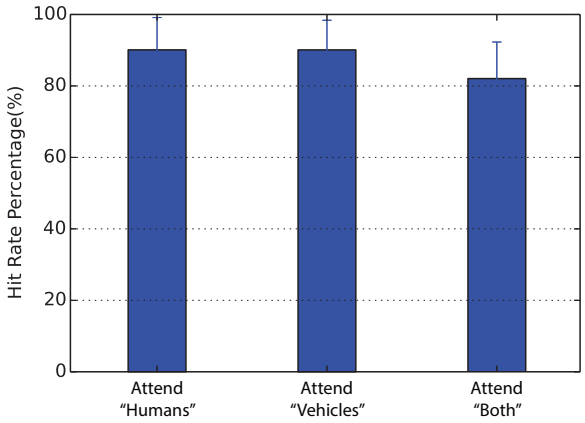

Figure 3.1: **Behavioral performance under different attention conditions.** Performance of behavioral task (i.e. button press task) is calculated for the three attention conditions. Error bars represent standard error of the mean across five subjects. There were no significant differences in the correct hit rates for the three search tasks.

3.2 Category model performance

In order to make conclusions using the assessed category weights, the underlying category model should have a reasonable accuracy.
After picking the optimal voxel-wise regularization parameters, nuisance regressors described in section 2.7 were removed from the stimulus matrix and voxel-wise models were fit. Voxel-wise BOLD responses were then predicted using the estimated category weights. Performance of the category model fit in each voxel was quantified by the Pearson's correlation between predicted and actual BOLD responses.

Figure 3.2 shows the prediction score on cortical flat maps for the five subjects of study. Flat maps are produced using Pycortex [25] software package. Voxels that appear in bright yellow have high prediction scores and the models are not significant in gray areas (t-test, $p > 0.05$ FDR corrected). The ROI boundaries on the maps are drawn by specifying boundaries on contrast maps, gathered from localizer scans. It is seen that the category model predicts fairly well in most of the dorsal and ventral visual streams.

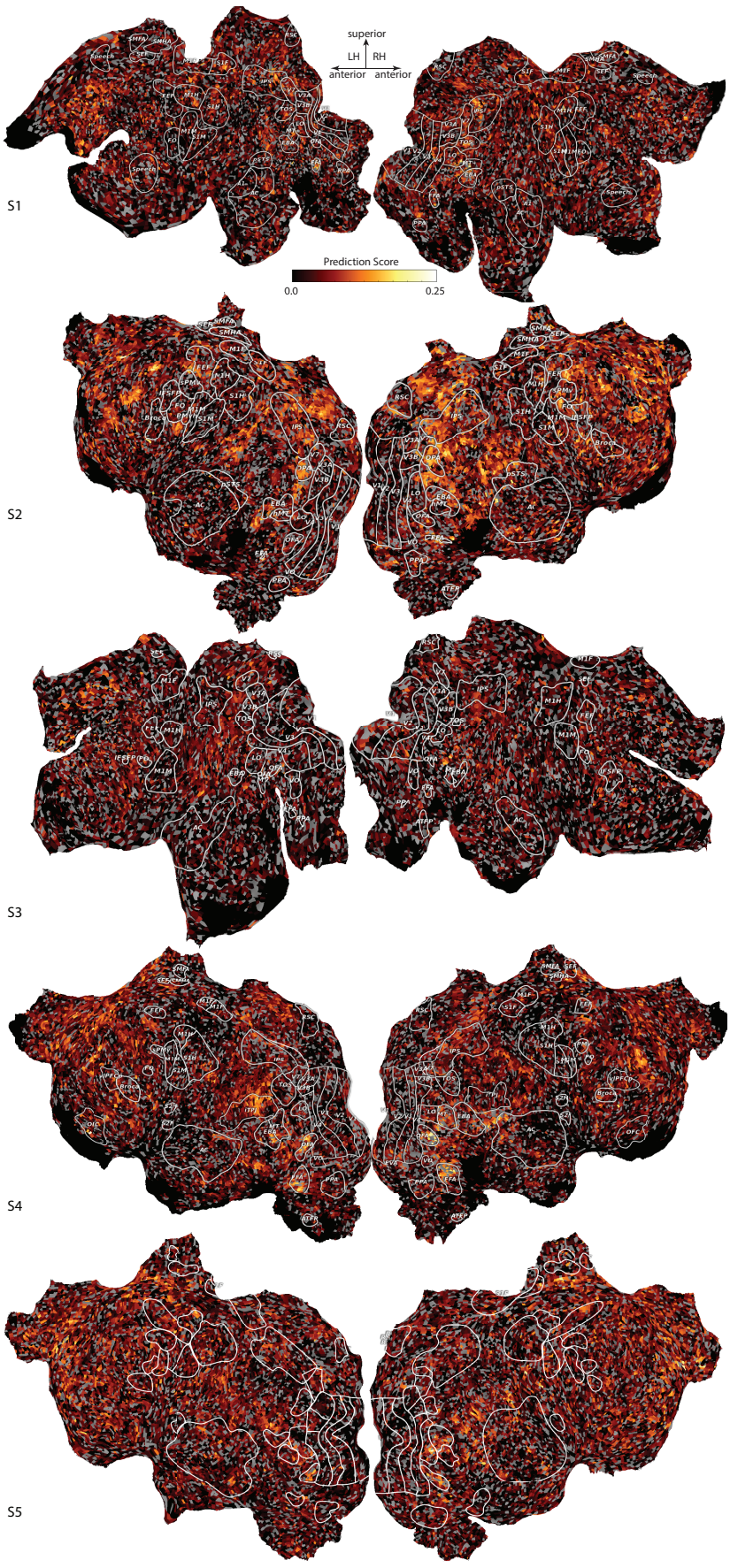

Figure 3.2: **Cortical flat maps of prediction scores for the five subjects.** To ensure the validity of the category model used across the study, prediction score of the category model is illustrated on the cortical flat map for the five subjects(S1-S5). Prediction scores were quantified as the Pearson's correlation between the predicted and measured BOLD responses. Brighter regions have higher prediction score and darker areas have lower prediction score. Model predictions are not significant in gray areas (t-test, $p > 0.05$ FDR corrected). The cortical regions identified from localizer sessions' data are drawn and labeled on the maps. Voxels across much of the dorsal and ventral visual streams are well predicted.

Studying the distribution of category weights across the cortex also provides valuable insight into attentional modulation in different brain areas. Figure 3.3 visualizes the distribution of the mean weights of "humans" and "vehicle" categories for a representative subject in three attention conditions on flat maps. It is seen that voxels in areas that has been shown to be selective to "human"-related categories during passive viewing (i.e. FFA, EBA) [22] appear in red color, meaning that they are tuned for the human categories. Meanwhile voxels in areas that are shown to be selective to visual scenes during passive viewing (i.e. PPA, RSC) [22] appear in green color and are tuned to vehicle categories. It is seen that many voxels across the ventral and dorsal visual streams preserve their tuning towards the preferred object category while dividing attention to "both humans and vehicles", whereas some voxels in anterior prefrontal cortex shift their tuning away from both human and vehicle categories. This finding suggests that these areas are involved in distractor detection [6].

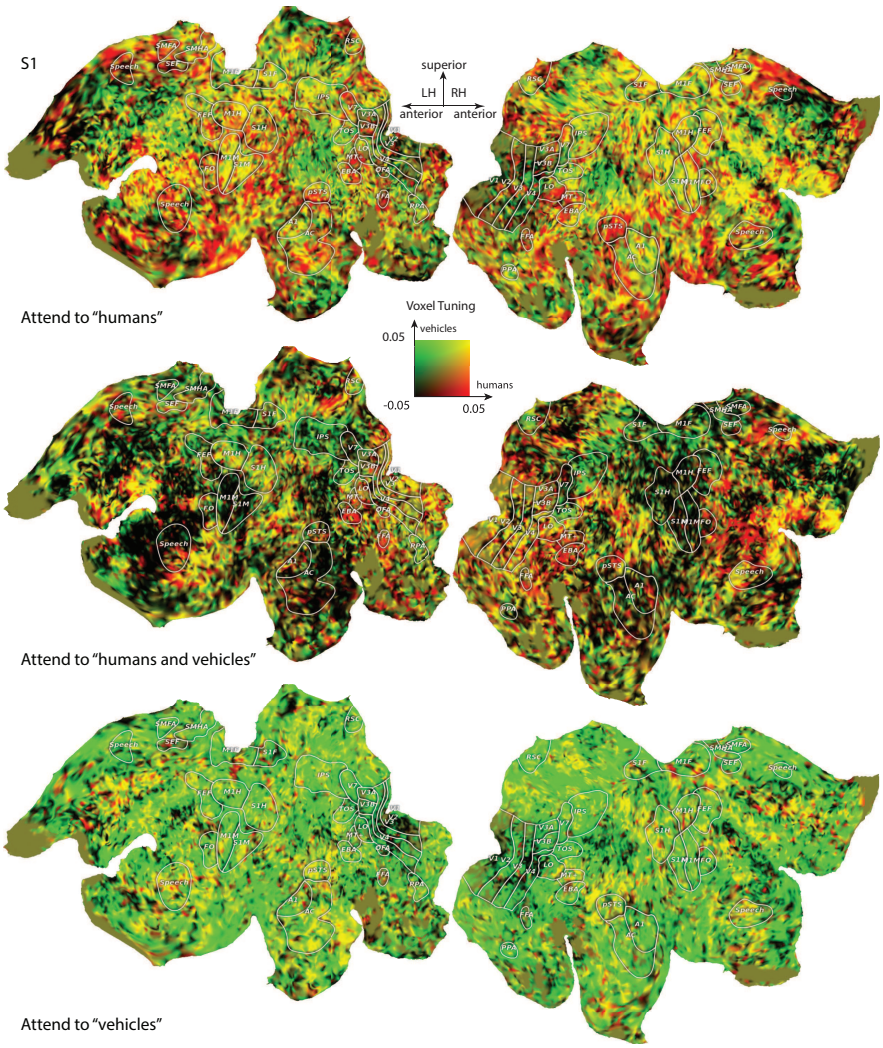

Figure 3.3: **Cortical flat maps of the mean category weights for human and vehicle categories in subject S1.** Many cortical voxels shift their category-tuning to enhance representation of the attended category. In the divided attention condition, many voxels shift their tuning away from the attended target categories. Voxels in areas selective for human categories (FFA, EBA), and vehicle categories (PPA, RSC) retain their category-tuning for the preferred object category.

3.3 The continuous semantic space

To have an idea of the distinctive properties of the principal components of the semantic space, we have illustrated the first four PC loadings for subject S1 in Figure 3.4. It is seen that first principal component distinguishes between categories with high motion-energy (e.g. humans, cars) and still categories (e.g. sky, land). Second PC distinguishes between humans and vehicles. Third PC weighs more to activity verbs and the fourth PC weighs more to animal categories and activity-related verbs.

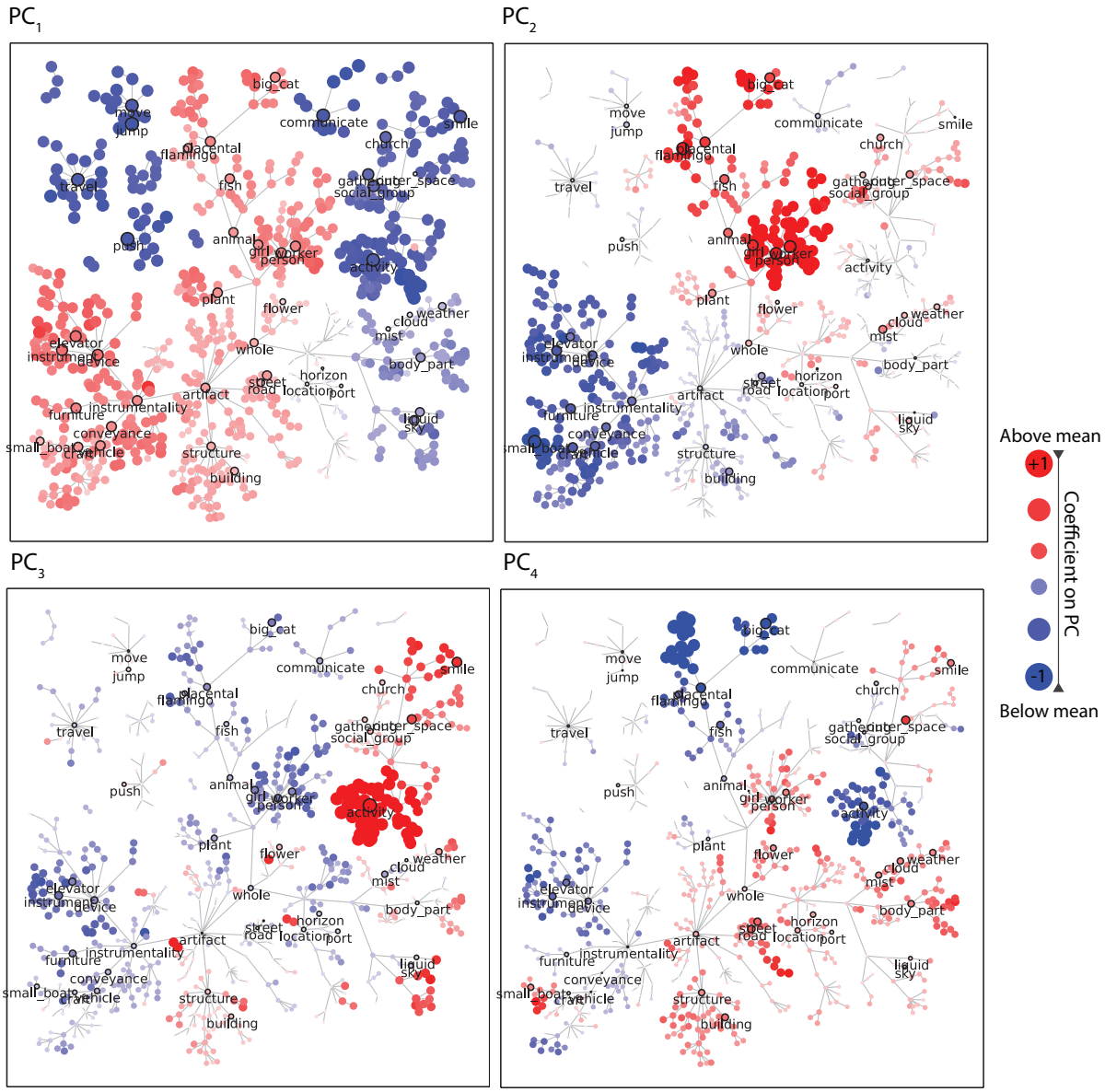

Figure 3.4: **Graphical visualization of semantic space.** Coefficients of all 831 categories in the first four PCs of subject S1 are shown in Wordnet graphical structure. Coefficients above mean are colored in red and coefficients below mean are shown in blue. Radius of dots represent the absolute value of the coefficient. First principal component distinguishes between categories with high motion-energy (e.g. humans, cars) and still categories (e.g. sky, land). Second PC distinguishes between humans and vehicles. Third PC weighs more to activity verbs and the fourth PC weighs more to animal categories and activity-related verbs. The display scale for the PC loadings is adjusted to the -1 and +1 range.

In all further analysis, collection of PCs explaining at least 90 percent of the variance in the data are taken as the axes of the semantic space.

3.4 Linearity index across the cortex

Linearity index indicates how well the semantic-tuning profiles in divided attention condition can be predicted by a weighted average of tuning profiles during isolated attention conditions. We hypothesized that the strongly category selective areas (such as FFA, PPA, EBA, RSC) be less flexible to changes in the attention state, and thus less linear than the attention-control areas (such as IPS, FEF, FO) [11, 62]. Figure 3.5 shows the average linearity index in ROIs of study.

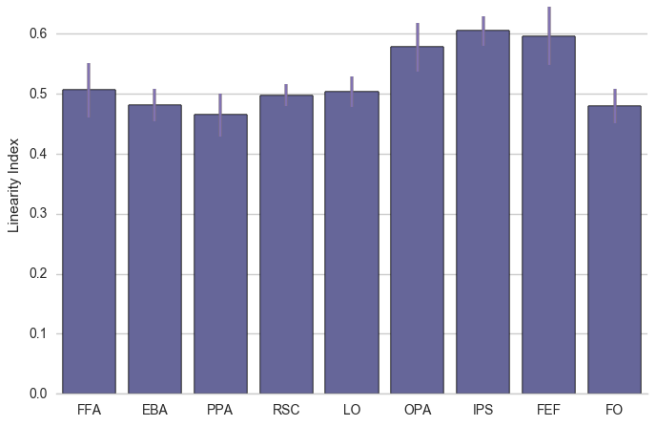

Figure 3.5: **Linearity index calculated in different ROIs.** Linearity index is averaged across voxels of each ROI. This index indicates the accuracy of prediction of the tuning profile in divided attention condition using a weighted average of tuning profiles while attending to isolated categories. Error bars represent standard error of the mean across five subjects.

We find that the average linearity index is 0.49 ± 0.03 ($mean \pm s.d.$, averaged across subjects) in category-selective areas (FFA, PPA, EBA), and 0.56 ± 0.02 in attentional-control areas (IPS, FEF, FO). This result suggests that, in higher visual cortex, a substantial portion of semantic tuning during divided attention can be expressed as a weighted average of tuning profiles during attention to single targets. One-way ANOVA on the linearity index in these ROIs shows significance

variability in linearity across cortical regions ($F(8) = 2.43, p = 0.0327$) and the linearity index in attentional-control areas is significantly higher than category-selective areas (t-test, $p < 10^{-6}$). This is consistent with previous studies that suggests that the representation at later stages of visual processing are more flexible than representations at earlier stages [11].
Flat maps of voxel-wise linearity index across the cortex for the five subjects is illustrated in Figure 3.6. Voxels with model prediction score below zero are colored in gray. It is seen that many voxels in lateral part of parietal lobe and voxels in the frontal cortex have high linearity index.

3.5 Bias in semantic representations

It has been previously suggested that category-selective ROIs in ventral-temporal cortex including FFA and PPA show bias towards their preferred category during divided attention and clutter in the visual scene [50]. We hypothesized that the semantic tuning in category-selective areas in ventral-temporal cortex will be biased towards the preferred object category, but not towards any specific category in attention control areas. Inspecting the category-tuning vectors across functional ROIs reveals bias towards preferred object category in strongly object-selective areas such as FFA and PPA. Tuning profiles for two representative voxels in FFA and PPA are shown in Figure 3.7 separately for the three search tasks.
As seen in Figure 3.7, in both FFA and PPA, category-tuning profile for the B condition appears to lie between the tuning vectors for the V and H conditions. As expected, the FFA voxel responds strongly to human categories whereas the PPA voxel responds strongly to vehicle categories.

To study this bias in the semantic space we calculated the average shift Index as formulated in Chapter 2. Shift index of separate object categories are calculated by projecting only the corresponding category weights onto the semantic space and averaging the shift index over projected weights.
To get an idea of the semantic relationship between the object categories which

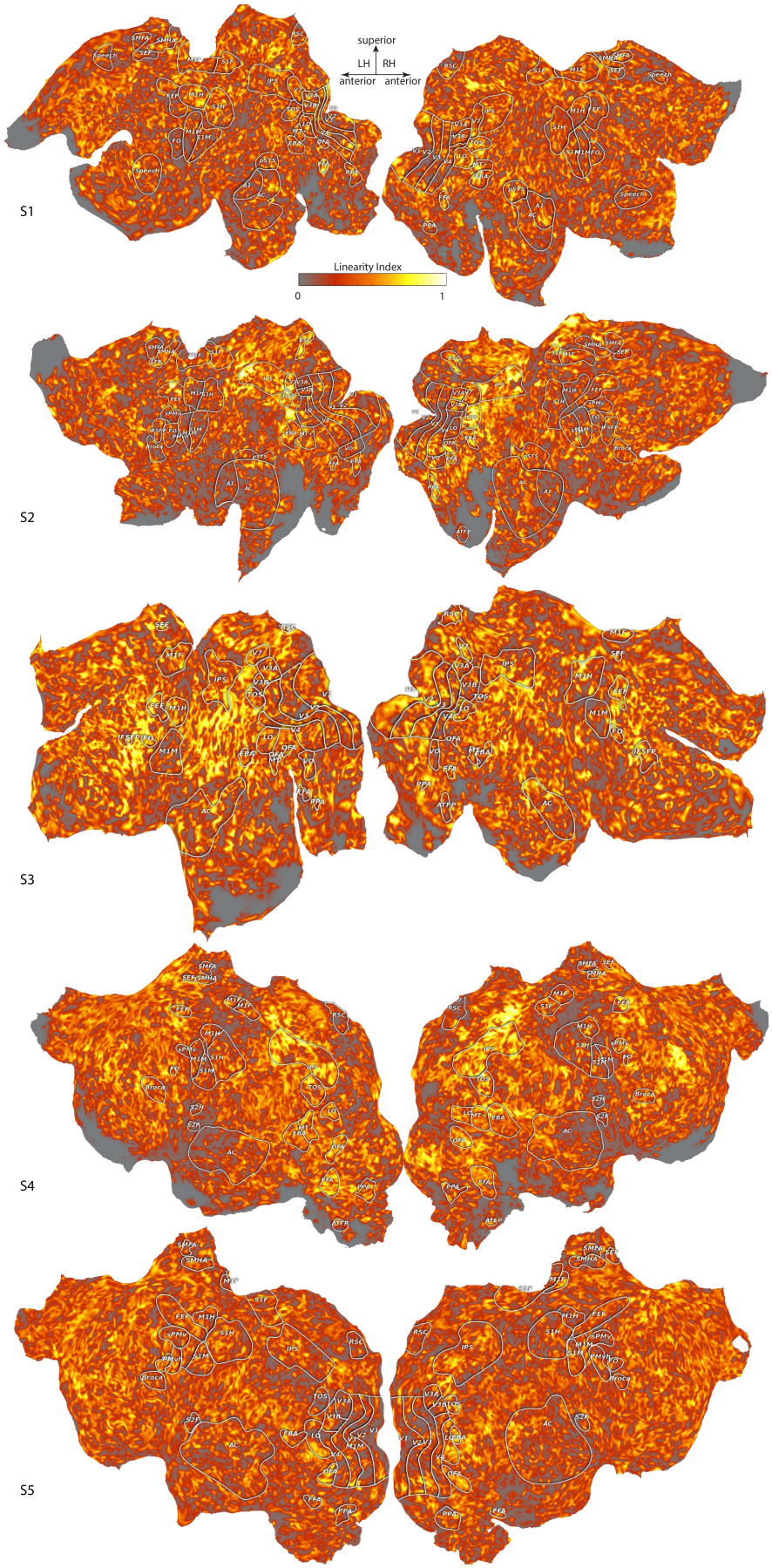

Figure 3.6: **Cortical flat maps of linearity index for five subjects.** The voxel-wise linearity index is illustrated on the cortical flat map for the five subjects(S1-S5). Linearity index indicates the accuracy of describing the semantic-tuning profiles during divided attention condition using a weighted average of tuning profiles during the two isolated attention conditions. Voxels in which prediction score of the category model is less than zero are colored in gray. Brighter voxels have higher linearity and darker ones have low linearity. Many voxels in lateral parietal lobe and frontal lobe have high linearity index.

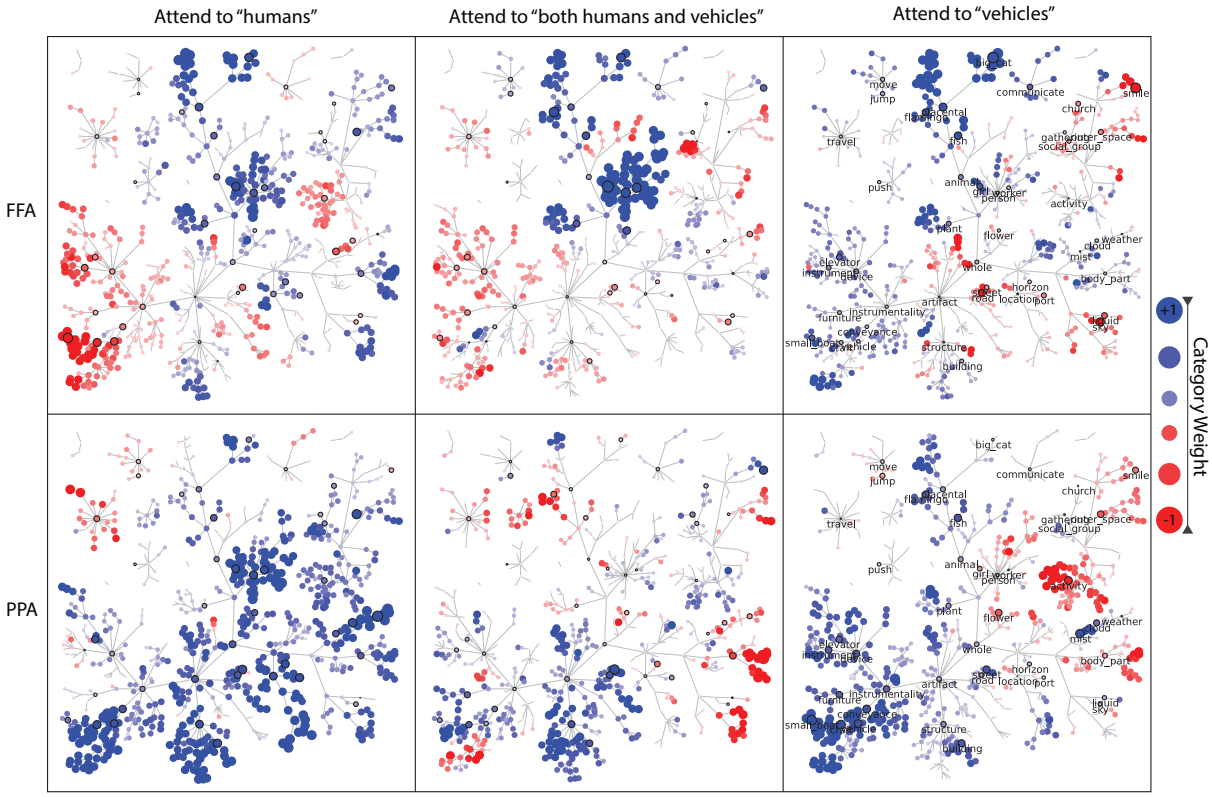

Figure 3.7: **Category-tuning profiles for representative voxels in FFA and PPA.** Category- tuning profiles of 831 object and action categories during different attention conditions ("Attend V": search for "vehicles", "Attend H": search for "humans", "Attend B": search for "both") for one FFA voxel **(top row)** and one PPA voxel **(bottom row)** from subject S1 are represented on the WordNet graph. The display scale for the tuning profiles is adjusted to the -1 and +1 range. category-tuning profiles are biased toward preferred object category during divided attention.

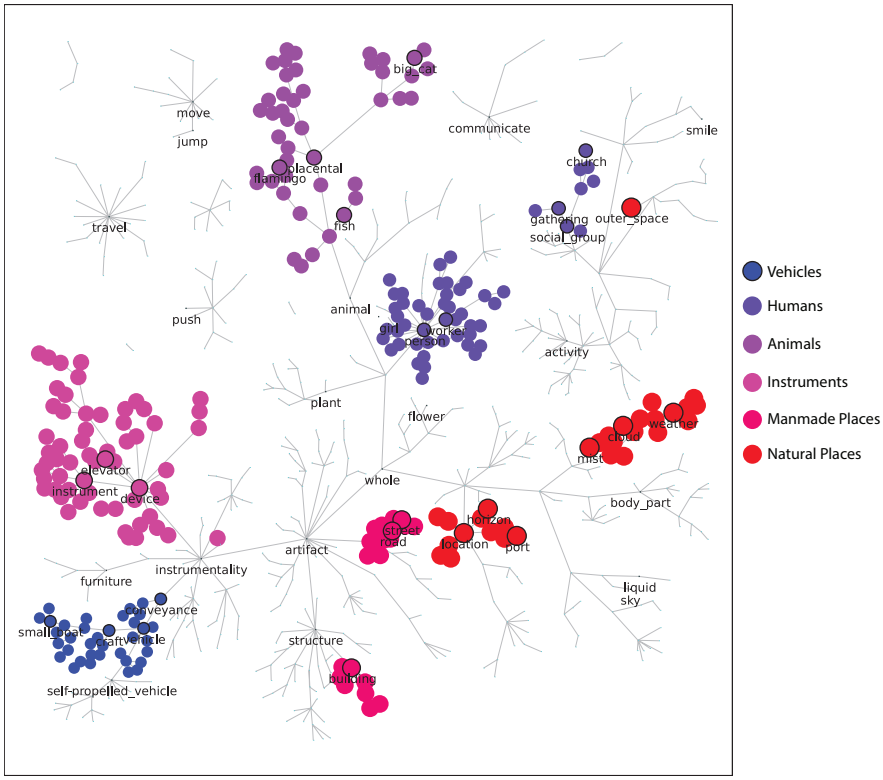

Figure 3.8: **Placement of object categories of study on the WordNet hierarchy graph.** Six distinct object categories are studied for their weights shift behavior in this study. In this figure, some of the representative nodes are labeled to visualize their semantic relationship.

will be discussed in the following, separate distinct object categories are represented on the WordNet graph in Figure 3.8.

Shift index of "target" categories, the union of "human" and "vehicle" categories is shown in Figure 3.9. As seen in Figure 3.9, the tuning profiles in category-selective areas , FFA, PPA and EBA, are biased towards the preferred target category, while no significant bias (bootstrap, $p > 0.05$) is seen in general object-selective area LO and attention-control areas, IPS, FEF and FO.

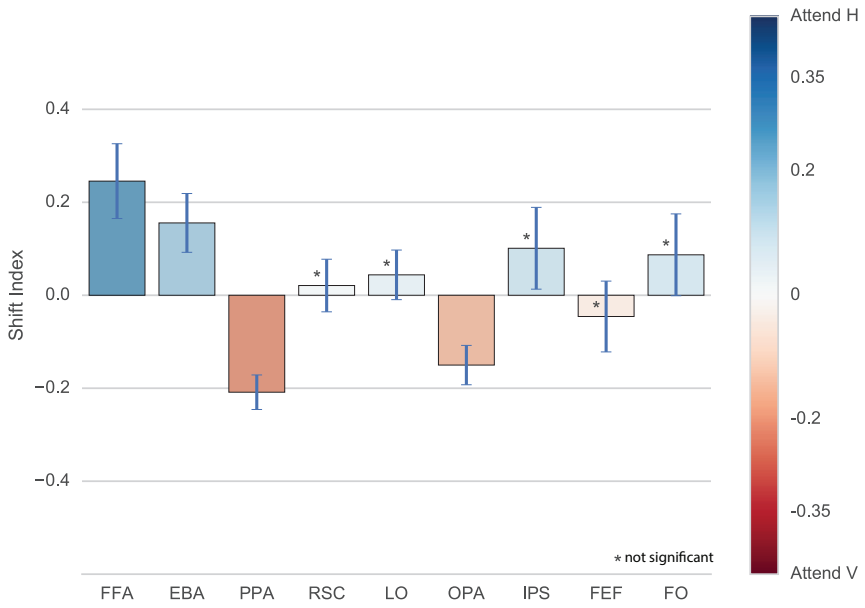

Figure 3.9: **Average shift index of "target" categories.** Average weights shift of target categories is calculated by predicting the tuning profile distribution during divided attention condition as a weighted average of tuning profile distributions during isolated attention conditions for the union of vehicle and human categories. Results are averaged over five subjects. Error bars indicate standard error across subjects. the weights in object-selective areas, FFA, PPA, EBA, are biased towards the preferred target, while no significant bias ($p > 0.05$) is seen in general object-selective area LO and attention-control areas, IPS, FEF, FO. Non-significant results are marked with asterisk.

To ensure that this effect is not due to the mere object detection, we assessed the shift index of non-target object categories, i.e. all categories excluding the target ones. The shift index of non-target categories is depicted in Figure 3.10. It is seen that the object-selective areas preserve their bias towards the preferred attention condition even without the target categories included.

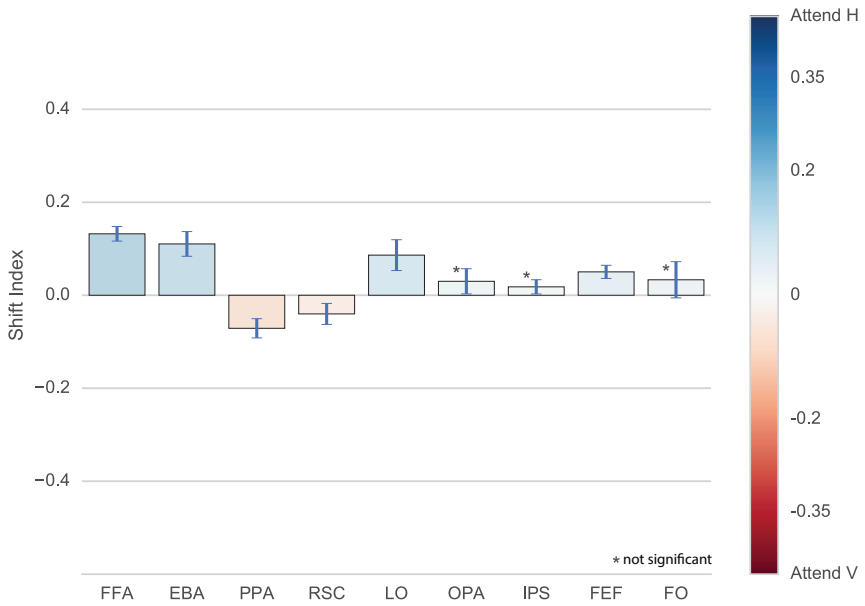

Figure 3.10: **Average shift index of "non-target" categories, control analysis.** Average shift index of non-target categories indicate that the bias observed in object-selective areas is preserved even without targets included, thus the effect is not a result of target detection. Results are averaged over five subjects. Error bars indicate standard error across subjects. Non-significant results ($p > 0.05$) are marked with asterisk.

Examining the shift index for "human" and "vehicle" categories individually reveals some interesting patterns. Shift index of human and vehicle categories is shown in Figure 3.11.

Figure 3.11 also shows that weights are shifted away from the target category in IPS and this is a significant effect ($p < 0.05$). Another interesting result is the effect in OPA. We were expecting OPA to show similar behavior to the other scene-selective area RSC, but it does not follow the patterns in RSC.

The bias in category-tuning distributions during divided attention towards category-tuning distributions during any of the single-target tasks is also consistent with the results of bias in semantic-tuning distributions. We computed shift index using directly the category-tuning distributions and illustrated them in WordNet graphs of Figures 3.12 and 3.13. The category-tuning distribution of

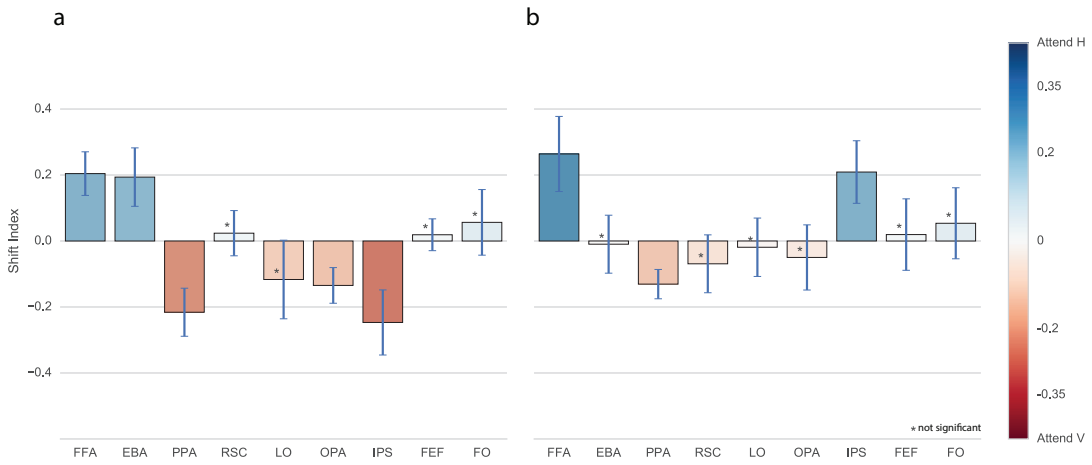

Figure 3.11: **Average shift index of separate target categories.** Average shift index of human**(a)** and vehicle**(b)** categories shows a bias towards the target category in category selective areas and away from it in IPS. Results are averaged over five subjects. Error bars indicate standard error across subjects. Non-significant results ($p > 0.05$) are marked with asterisk.

"target" categories during divided attention in the category-selective areas FFA, PPA and EBA are biased towards the tuning distribution while attending to the preferred object category and bias is away from the target category in IPS.

Average shift index of separate non-target categories grouped as "animals", "instruments", "man-made scenes" and "natural scenes" are depicted in Figure 3.14. "Animal" categories are semantically closer to humans than to vehicles, so it is predictable that their tuning distribution be biased towards the distribution in "attend to humans" condition in the category-selective areas. On the other hand, "instrument" categories are semantically closer to vehicles than to humans, thus the tuning distributions in category-selective areas are biased towards the distribution in "attend to vehicles" condition. The bias in tuning distribution of man-made scenes toward the distribution in "attend to humans" condition in FFA can be because of the semantic proximity of man-made scenes to humans, and the opposite is seen for the tuning distribution of natural scenes in the PPA. This suggests that during divided attention, semantic tunings distribution of categories which are semantically close to the targets are also biased towards the target categories.

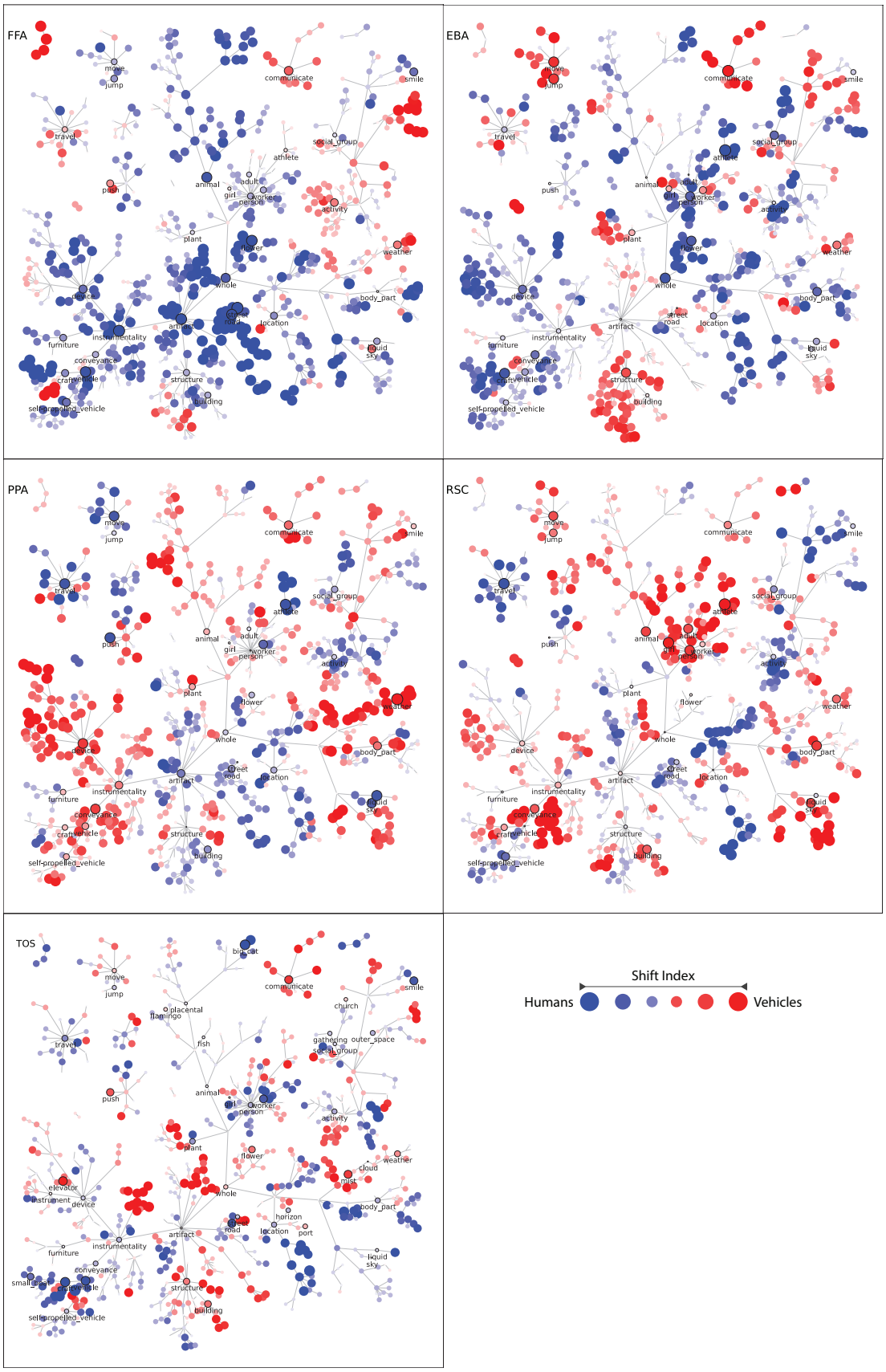

Figure 3.12: (Caption on the following page.)

Figure 3.12: **Shift index profile of categories in object-selective areas.** Semantic-tuning distribution of "target" categories during divided attention in the category-selective areas FFA, PPA and EBA are biased towards the tuning distribution while attending to the preferred object category. Results are averaged over five subjects.

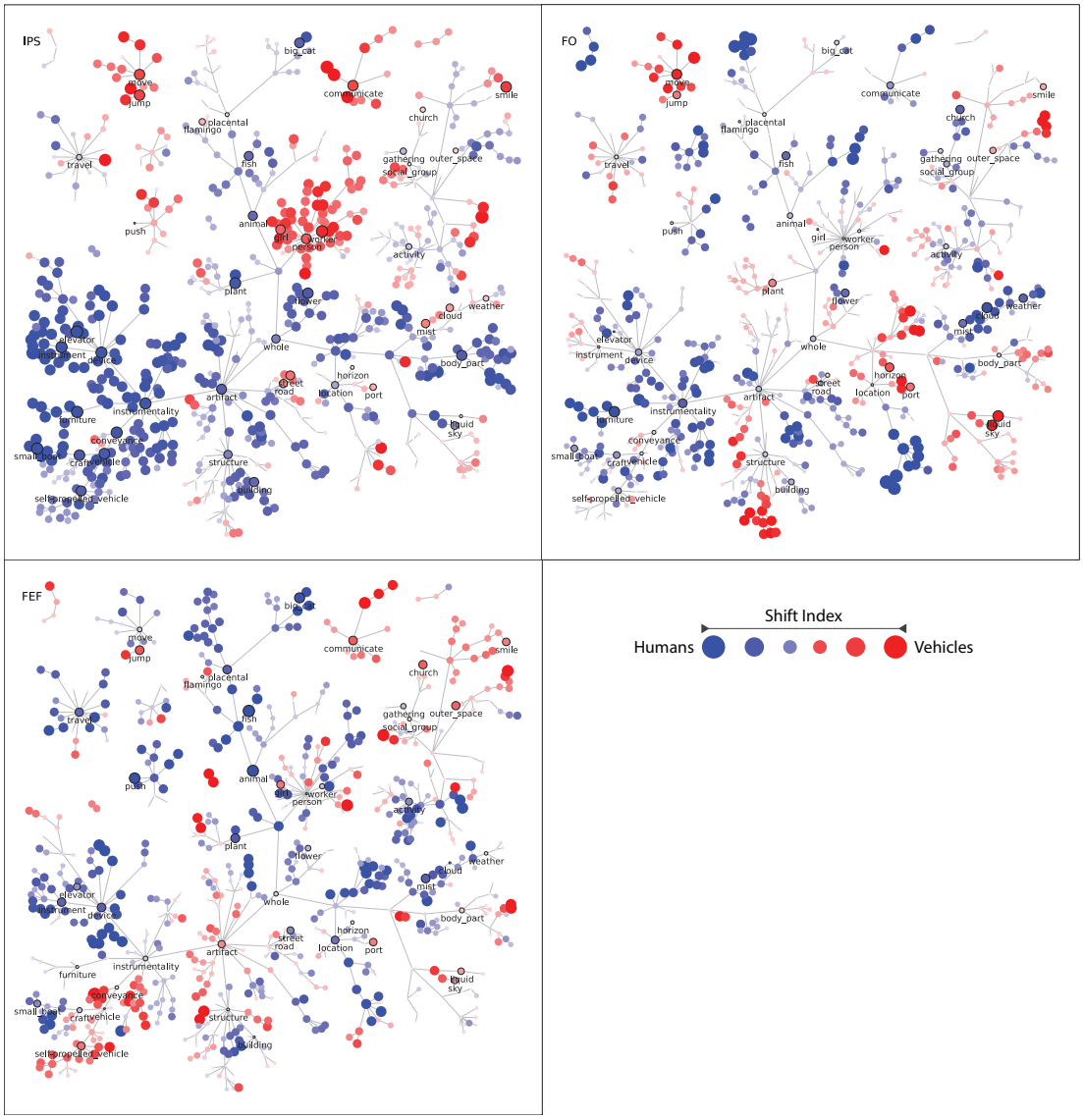

Figure 3.13: **Shift index profile of categories in attention-control areas.** Semantic-tuning distribution of "target" categories during divided attention in IPS is biased away from the tuning distribution while attending to the target category. Results are averaged over five subjects.

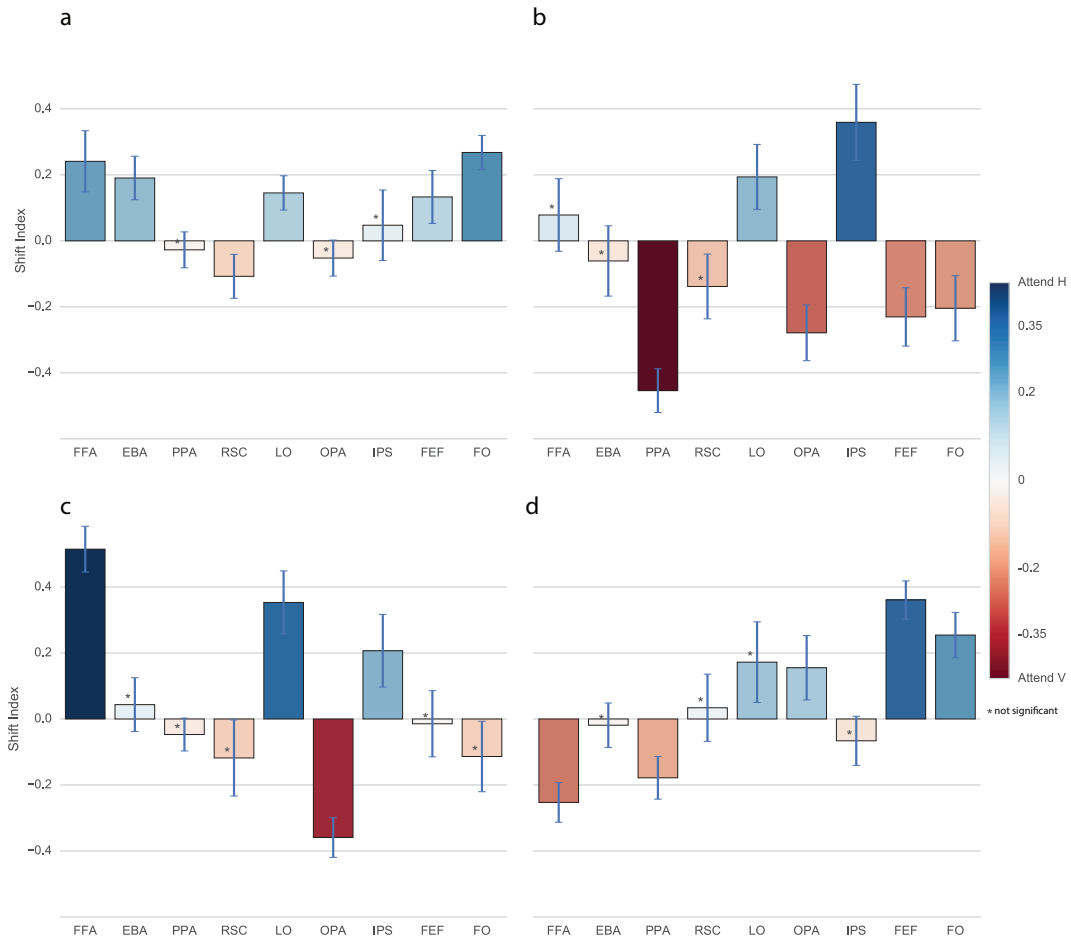

Figure 3.14: **Average shift index of different object categories.** Average shift index of animals**(a)**, instruments**(b)**, man-made scenes**(c)** and natural scenes**(d)** categories shows bias towards H condition in animal features and towards V condition in instrument features. Bias towards H condition in man-made place features is seen whereas the bias for natural places is towards V condition. Results are averaged over five subjects. Error bars indicate standard error across subjects. Non-significant results are marked with asterisk.

Chapter 4

Summary and Conclusion

In this thesis, we tested the predictions of the biased competition theory using a voxel-wise computational analysis of the brain activity collected during natural vision. Our results indicate that the cortical semantic tunings during divided attention among multiple object categories is described as a weighted average of semantic tunings when attending to each of the target object categories solely. This result is consistent with the prediction of the biased competition theory. We have further shown that performance of the weighted average model is higher in the attention control cortical areas compared to the object selective areas. Furthermore, semantic representation while dividing attention among multiple targets is biased with varying extent and direction across the cortex. The competition in the object-selective areas is biased in favor of the preferred object category and not toward any of the target categories in the areas belonging to the attention control network. These results are consistent with neurophysiological studies and previous fMRI evidence for biased competition [18, 52, 2, 3].

An interesting finding here is that the linearity index in OPA is unexpectedly high. Similar to PPA and RSC, OPA is also shown to be selective for visual scenes [15]. However, recent evidence suggests that OPA has a different functional role than PPA and RSC in scene representation. A recent fMRI study implies that OPA represents scenes using the local representations of their constituent objects,

whereas PPA and RSC represent global scene properties [33]. Thus, the high linearity index that we measure in OPA might reflect the dependence of scene representations to individual representations of local partial elements in OPA.

There are multiple studies on the relationship between cortical representation during divided attention among multiple targets and the representations while attending to each of them solely. While some of these studies favor a simple mean relationship [43, 63], others suggest a nonlinear relationship in the sense that response during divided attention is equal to the maximum among the two responses during single-target attention tasks [29]. The analyses used in this thesis enabled us to investigate more general relationships including the ones proposed by previous studies as special cases of the weighted average. Under this scheme we are also able to assess the weights in the weighted average for each of our diverse set of object and action categories separately. Our results are consistent with a recent fMRI study on multivoxel BOLD responses showing that response patterns in object-selective areas while viewing two objects simultaneously is a weighted average of the response patterns elicited during viewing each of them alone [51]. However, our work here suggests that the predictions of biased-competition theory also hold for the attentional modulations of semantic representations. Our results show that semantic representation during divided attention is biased towards the preferred object category in areas selective to the target categories, but there is no significant bias towards any of the target categories in the general-object-selective area LO and in the attention-control areas.

We have shown that the shift index in the IPS is positive for vehicle-related categories and negative for the human-related ones, meaning that representation in IPS is biased towards the distractor object category. This is in line with studies stating that IPS enhances target detection by identifying non-target features and thus, optimizing target detection [16, 5].

The attentional bias in semantic representation of non-target categories show that semantic-tuning distributions in FFA, PPA and EBA are biased towards the preferred object category even in the absence of targets. This is in line with a

recent study showing that that representations in object-selective areas (FFA, PPA) show strong robustness to divided attention and clutter in the scene, but this does not hold for the lateral occipital complex [50].

Analysis of the diverse set of object categories indicates that the representation of the categories which are semantically close to the targets are also biased towards the target categories in object-selective areas.

Target-presence regressor was utilized with the assumption that it would regress out the modulations raising form mere target presence from the BOLD responses. Low category model prediction scores in retinotopic areas confirm this assumption. Further, no significant experiment difficulty was observed between the three attention tasks. Absence of target-presence modulations in the BOLD responses guarantee that the observed biases are not mere consequences of target detection.

4.1 Limitations and future work

Although we have based our analysis on voxel-wise semantic tunings and attention to semantic object categories and the subjects were asked to fixate on the center of screen during the experiment, presence of target object categories in different positions across the receptive field and their varying frequency of appearance might alter voxel tunings, and the major results of this study. Analysis of the eye tracking data gathered during the experiment sessions would thus be an immediate future step in this study. Color difference between humans and vehicles might be a confounding factor in the analysis. Vivid colors of vehicles can modulate BOLD responses through bottom-up mechanisms. Thus, doing the same experiment with black and white movies would be another possible control analysis.

The target object categories of this study (humans and vehicles) are very distinct. However, the way that attention enhances target detection depends

on similarity of target and distractors. If the target and distractor are similar, attention enhances the representation of features that distinguish target and distractor. An extension to this work could be studying the linearity while target and distractor are not this much distinct.

Attentional effects on semantic representations are multi-modality rather being limited to natural vision. Thus, an important future work in the direction of this thesis will be to study the linearity in other modalities. For instance, a similar experiment in which subjects listen to stories and divide their attention among specific categories could be used to investigate biased competition in the audition.

Bibliography

[1] Geoffrey K Aguirre, E Zarahn, and D'Esposito, M. An area within human ventral cortex sensitive to Building stimuli. *Neuron*, 21(2):373–383, 1998. ISSN 0896-6273. doi: 10.1016/S0896-6273(00)80546-2.

[2] Diane M Beck and Sabine Kastner. Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, 8(8):1110–1116, 2005. ISSN 1097-6256. doi: 10.1038/nn1501.

[3] Diane M Beck and Sabine Kastner. Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, 2007.

[4] Diane M Beck and Sabine Kastner. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, 49(10):1154–1165, 2009. ISSN 0042-6989. doi: 10.1016/j.visres.2008.07.012.

[5] M Behrmann, JJ Geng, and S Shomstein. Parietal cortex and attention. *Current opinion in neurobiology*, 2004. ISSN 0959-4388. doi: 10.1016/j.conb.2004.03.012.

[6] Christoph Bledowski. Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage*, 22(2):530–540, 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2003.12.034.

[7] SL Bressler, W Tang, and CM Sylvester. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *Journal of Neuroscience*, 2008. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1776-08.2008.

[8] Richard B Buxton, Kamil Uludağ, David J Dubowitz, and Thomas T Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23 Suppl 1:S220–33, 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.07.013.

[9] Tolga Çukur, Alexander G Huth, and Jack L Gallant. Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci*, 16(6):763–770, 2013. ISSN 1097-6256. doi: 10.1038/nn.3381.

[10] JD Connolly and MA Goodale. A comparison of frontoparietal fMRI activation during anti-saccades and anti-pointing. *Journal of Neurophysiology*, 2000.

[11] Erik P Cook and John HR Maunsell. Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. *Journal of Neuroscience*, 22(5):1994–2004, 2002.

[12] Maurizio Corbetta and Gordon L Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*, 3(3):201–215, 2002. ISSN 1471-003X. doi: 10.1038/nrn755.

[13] Maurizio Corbetta, Erbil Akbudak, Thomas E Conturo, Abraham Z Snyder, John M Ollinger, Heather A Drury, Martin R Linenweber, Steven E Petersen, Marcus E Raichle, David C Van Essen, et al. A common network of functional areas for attention and eye movements. *Neuron*, 21(4):761–773, 1998.

[14] G Deco and TS Lee. A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, 2002.

[15] Daniel D Dilks, Joshua B Julian, Alexander M Paunov, and Nancy Kanwisher. The occipital place area is causally and selectively involved in scene perception. *The Journal of Neuroscience*, 33(4):1331–1336, 2013. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.4081-12.2013.

[16] Tobias H Donner, Andreas Kettermann, Eugen Diesch, Arno Villringer, and Stephan A Brandt. Parietal activation during visual search in the absence of multiple distractors. *NeuroReport*, 14(17):2257–2261, 2003.

[17] PE Downing, Y Jiang, M Shuman, and N Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–3, 2001. ISSN 0036-8075. doi: 10.1126/science.1063414.

[18] J Duncan. Selective attention and the organization of visual information. *Journal of Experimental Psychology*, 1984.

[19] J Duncan. Cooperating brain systems in selective perception and action. *Attention and Performance*, 1996.

[20] MP Eckstein, JP Thomas, and J Palmer. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception and Psychophysics*, 2000. doi: 10.3758/BF03212096.

[21] Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998. ISSN 0028-0836. doi: 10.1038/33402.

[22] Russell A Epstein. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10):388–396, 2008. ISSN 1364-6613. doi: 10.1016/j.tics.2008.07.004.

[23] Michael W Eysenck and Mark T Keane. *Cognitive psychology: A student's handbook*. Taylor & Francis, 2000.

[24] D.H. ffytche and S Zeki. Brain activity related to the perception of illusory contours. *NeuroImage*, 3(2):104–108, 1996. ISSN 1053-8119. doi: 10.1006/nimg.1996.0012.

[25] James S Gao, Alexander G Huth, Mark D Lescroart, and Jack L Gallant. Pycortex: an interactive surface visualizer for fMRI. *Frontiers in neuroinformatics*, 9, 2015. doi: 10.3389/fninf.2015.00023.

[26] MA Goodale and AD Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 1992.

[27] Kathleen A Hansen, Kendrick N Kay, and Jack L Gallant. Topographic organization in and near human visual area v4. *The Journal of Neuroscience*, 27(44):11896–11911, 2007.

[28] Uri Hasson, Michal Harel, Ifat Levy, and Rafael Malach. Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, 37(6):1027–1041, 2003.

[29] Hilary W Heuer and Kenneth H Britten. Contrast dependence of response normalization in area MT of the rhesus macaque. *Journal of Neurophysiology*, 88(6):3398–3408, 2002. ISSN 0022-3077. doi: 10.1152/jn.00255.2002.

[30] T Higo, RB Mars, and ED Boorman. Distributed and causal influence of frontal operculum in task control. *PNAS*, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1013361108.

[31] Alexander G Huth, Shinji Nishimoto, An T Vu, and Jack L Gallant. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6):1210, 2012. ISSN 0896-6273. doi: 10.1016/j.neuron.2012.10.014.

[32] Mark Jenkinson, Christian F Beckmann, Timothy E Behrens, Mark W Woolrich, and Stephen M Smith. FSL. *Neuroimage*, 62(2):782–90, 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2011.09.015.

[33] Frederik S Kamps, Vishal Lall, and Daniel D Dilks. The occipital place area represents first-person perspective motion information through scenes. *Neuroimage*, 83:17–26, 2016. ISSN 0010-9452. doi: 10.1016/j.cortex.2016.06.022.

[34] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. Elsevier, 2000.

[35] Nancy Kanwisher, McDermott, Josh, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302–4311, 1997. ISSN 0270-6474.

[36] Sabine Kastner and Leslie G Ungerleider. The neural basis of biased competition in human visual cortex. *Neurophysiologia*, 39(12):1263–1276, 2001. ISSN 0028-3932. doi: 10.1016/S0028-3932(01)00116-6.

[37] Z Kourtzi, M Erb, W Grodd, and HH Bülthoff. Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral Cortex*, 2003. doi: 10.1093/cercor/13.9.911.

[38] LJ Lanyon and SL Denham. A biased competition computational model of spatial and object-based attention mediating active visual search. *Nurocomputing*, 2004.

[39] George A Miller. WordNet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995. doi: 10.1145/219717.219748.

[40] DJ Mort, RJ Perry, SK Mannan, and TL Hodgson. Differential cortical activation during voluntary and reflexive saccades in man. *Neuroimage*, 2003.

[41] René M Müri. MRI and fMRI analysis of oculomotor function. *Progress in brain research*, 151:503–26, 2006.

[42] Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19):1641–1646, 2011. ISSN 0960-9822. doi: 10.1016/j.cub.2011.08.031.

[43] MacEvoy, Sean P and Russell A Epstein. Constructing scenes from objects in human occipitotemporal cortex. *Nature neuroscience*, 14(10):1323, 2011. doi: 10.1038/nn.2903.

[44] J Palmer. Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vision research*, 34(13):1703–21, 1994. ISSN 0042-6989.

[45] J Palmer, CT Ames, and DT Lindsey. Measuring the effect of attention on simple visual search. *Journal of experimental psychology. Human perception and performance*, 19(1):108, 1993. doi: 10.1037/0096-1523.19.1.108.

[46] A Pasupathy and CE Connor. Population coding of shape in area v4. *Nature Neuroscience*, 2002. doi: 10.1038/972.

[47] Tomá Paus. Location and function of the human frontal eye-field: A selective review. *Neurophysiologia*, 34(6):475–483, 1996. ISSN 0028-3932. doi: 10.1016/0028-3932(95)00134-4.

[48] F Pestilli and M Carrasco. Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision research*, 2005.

[49] William Prinzmetal, McCool, Christin, and Samuel Park. Attention: Reaction time and accuracy reveal different mechanisms. *J Exp Psychology Gen*, 134(1):73–92, 2005. ISSN 0096-3445. doi: 10.1037/0096-3445.134.1.73.

[50] Leila Reddy and Nancy Kanwisher. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current biology : CB*, 17(23):2067, 2007. ISSN 0960-9822. doi: 10.1016/j.cub.2007.10.043.

[51] Leila Reddy, Nancy G Kanwisher, and VanRullen, Rufin. Attention and biased competition in multi-voxel object representations. *PNAS*, 106(50): 21447, 2009. ISSN 1091-6490 (Electronic){\}r1091-6490 (Linking). doi: 10.1073/pnas.0907330106.

[52] JH Reynolds and L Chelazzi. Competitive mechanisms subserve attention in macaque areas v2 and v4. *Journal of Neuroscience*, 1999.

[53] John H Reynolds and David J Heeger. The normalization model of attention. *Neuron*, 61:168, 0. ISSN 0896-6273. doi: 10.1016/j.neuron.2009.01.002.

[54] Seong and Kamil Uğurbil. Comparison of blood oxygenattion and cerebral blood flow effect in fMRI: estimation of relative oxygen consumption change. *Magnetic Resonance in Medicine*, 38(1):59–65, 1997. ISSN 1522-2594. doi: 10.1002/mrm.1910380110.

[55] John T Serences, Jens Schwarzbach, Susan M Courtney, Xavier Golay, and Steven Yantis. Control of object-based attention in human cortex. *Cerebral Cortex*, 14(12):1346–1357, 2004. ISSN 1047-3211. doi: 10.1093/cercor/bhh095.

[56] J Sirosh and R Miikkulainen. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 1997. doi: 10.1162/neco.1997.9.3.577.

[57] MW Spratling. Predictive coding as a model of biased competition in visual attention. *Vision research*, 2008.

[58] P Verghese. Visual search and attention: a signal detection theory approach. *Neuron*, 31(4):523–35, 2001. ISSN 0896-6273. doi: 10.1016/S0896-6273(01)00392-0.

[59] Timothy D Verstynen and Vibhas Deshpande. Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. *Neuroimage*, 55(4):1633–1644, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.11.090.

[60] M Waterston and C Pack. Enhanced depth perception following high-frequency repetitive transcranial magnetic stimulation of human area V2/V3. *Journal of Vision*, 8(6):791–791, 2008. ISSN 1534-7362. doi: 10.1167/8.6.791.

[61] S Yantis, J Schwarzbach, and JT Serences. Transient neural activity in human parietal cortex during spatial attention shifts. *Nature* , 2002.

[62] Davide Zoccolan, Minjoon Kouh, Tomaso Poggio, and DiCarlo, James J. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(45):12292, 0. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1897-07.2007.

[63] Davide Zoccolan, David D Cox, and DiCarlo, James J. Multiple object response normalization in monkey inferotemporal cortex. *The Journal of Neuroscience*, 2005. doi: 10.1523/JNEUROSCI.2058-05.2005.

Appendix A

Optimal regularization parameters

Optimal voxel-wise regularization parameters were chosen using a cross validation procedure as described in chapter 2. However, there is always a possibility that the optimal regularization parameter be outside the test rage. To verify this, we have assessed the mean prediction score of the category model for different values of the regularization parameter, for different cross-validation folds. Figure A.1 illustrates the model prediction score for the range of tested regularization parameters in the five subjects. Candidate regularization parameters are $10 * 2^{\lambda_i}$ where λ_i ranges from 0 to 20. Peaks of prediction score fall in the proposed range for all subjects.

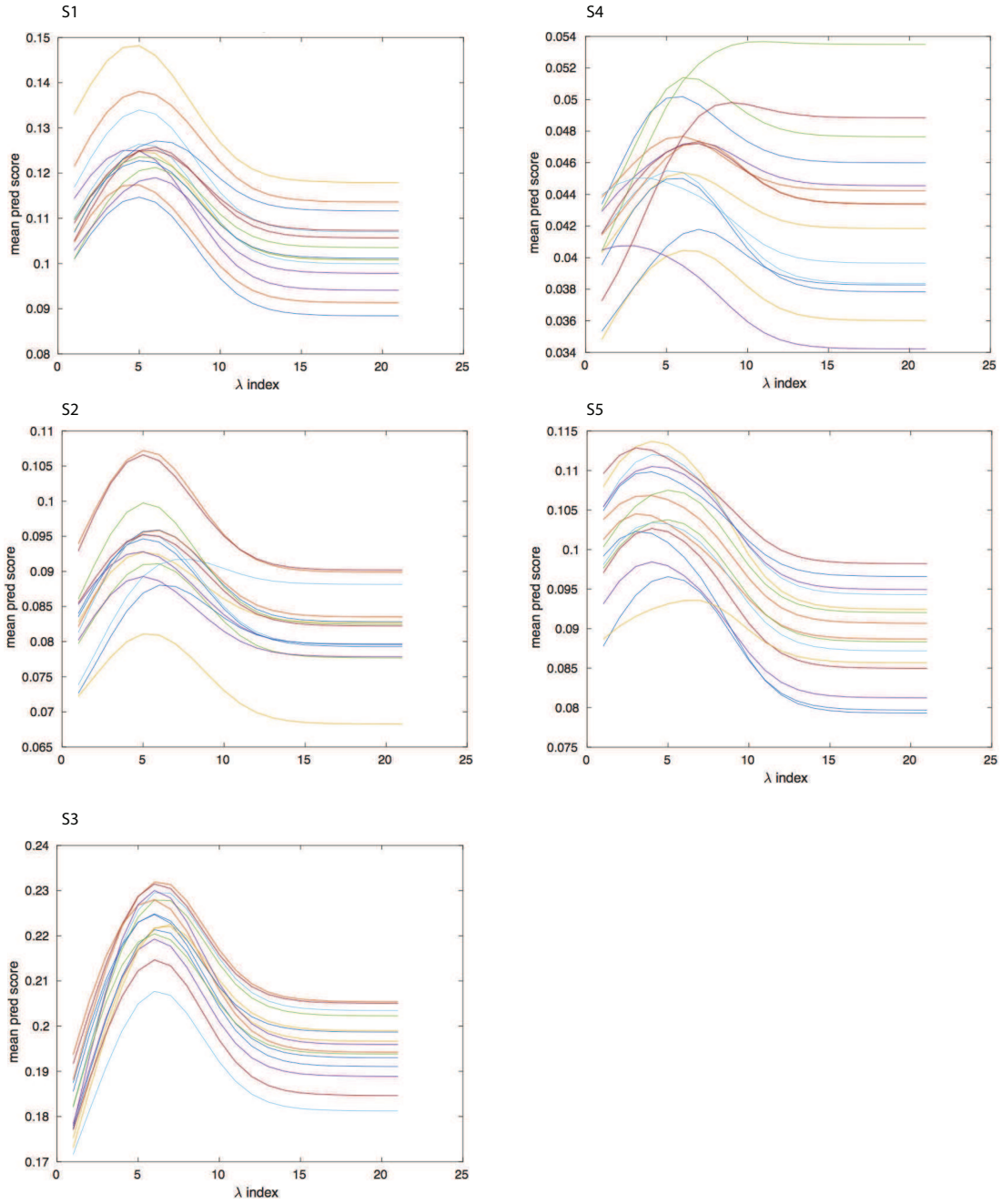

Figure A.1: **Verifying the search range of regularization parameter.** To ensure the validity of the search range in picking the best regularization parameters, mean prediction score for category model with different regularization parameters (λ_i) is shown for different cross-validation folds. In all subjects, peaks of prediction score occur for a λ_i in the search range.

Appendix B

Voxel-wise shift index

Voxel-wise shift index is calculated across the cortex. Shift index in each voxel is calculated using the semantic-tuning distribution across the voxels in a surrounding neighborhood of 26 voxels (a 3 by 3 by 3 cube, with the central voxel being the voxel of study). Voxel-wise shift indices give an overall picture of the distribution of bias in different regions of the cortex.

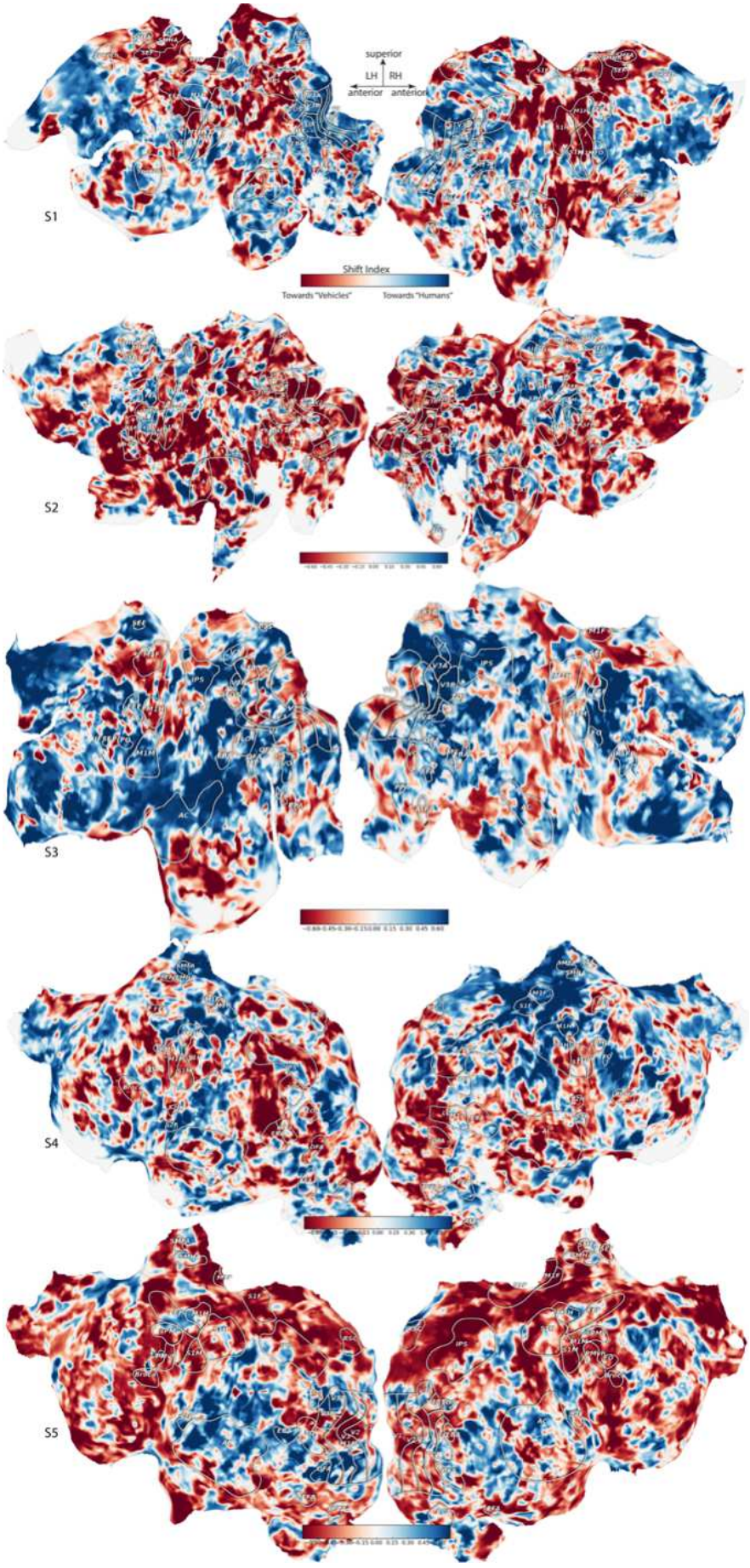

Figure B.1: **Cortical flat maps of shift index of "human" categories for the five subjects.** The voxel-wise shift index of "human" categories is illustrated on the cortical flat map for the five subjects(S1-S5). Positive shift index values mean that the semantic-tuning distribution of divided attention is more close to semantic-tuning distribution during the attend to "humans" condition and negative values indicate bias towards semantic-tuning distribution while attending to "vehicles" condition.

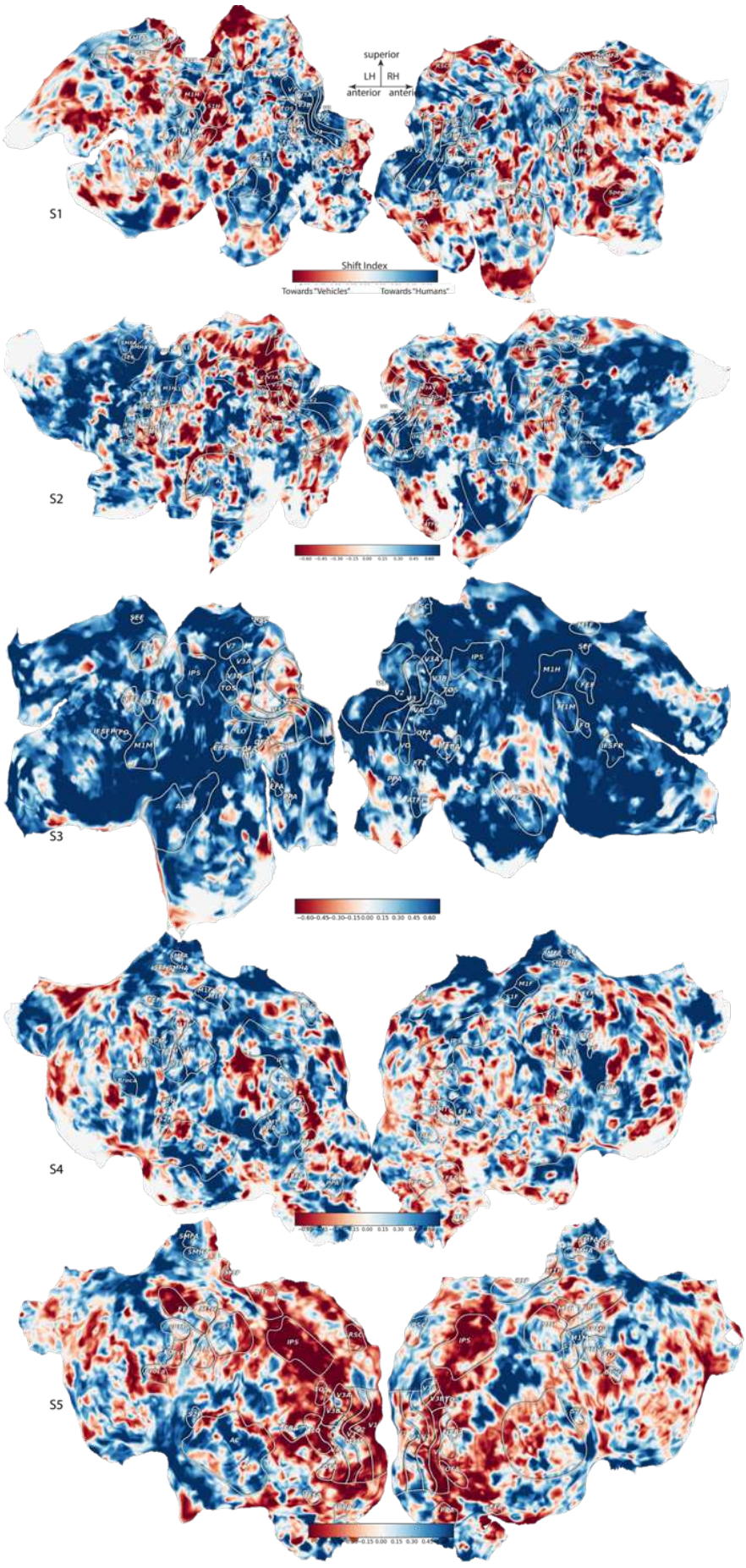

Figure B.2: **Cortical flat maps of shift index of "vehicle" categories for the five subjects.** The voxel-wise shift index of "vehicle" categories is illustrated on the cortical flat map for the five subjects(S1-S5). Positive shift index values mean that the semantic-tuning distribution of divided attention is more close to semantic-tuning distribution during the attend to "humans" condition and negative values indicate bias towards semantic-tuning distribution while attending to "vehicles" condition.

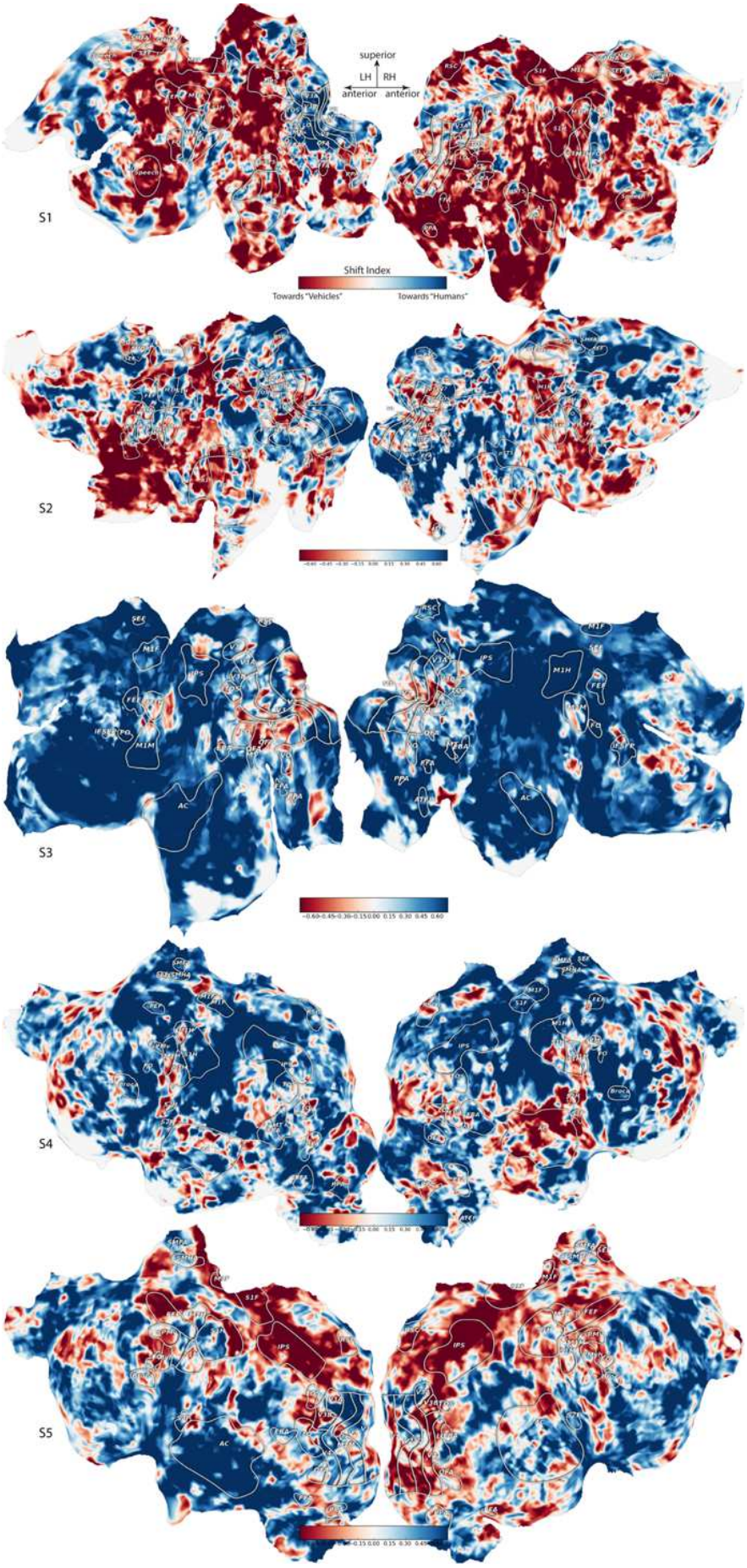

Figure B.3: **Cortical flat maps of shift index of "animal" categories for the five subjects.** The voxel-wise shift index of "animal" categories is illustrated on the cortical flat map for the five subjects(S1-S5). Positive shift index values mean that the semantic-tuning distribution of divided attention is more close to semantic-tuning distribution during the attend to "humans" condition and negative values indicate bias towards semantic-tuning distribution while attending to "vehicles" condition.

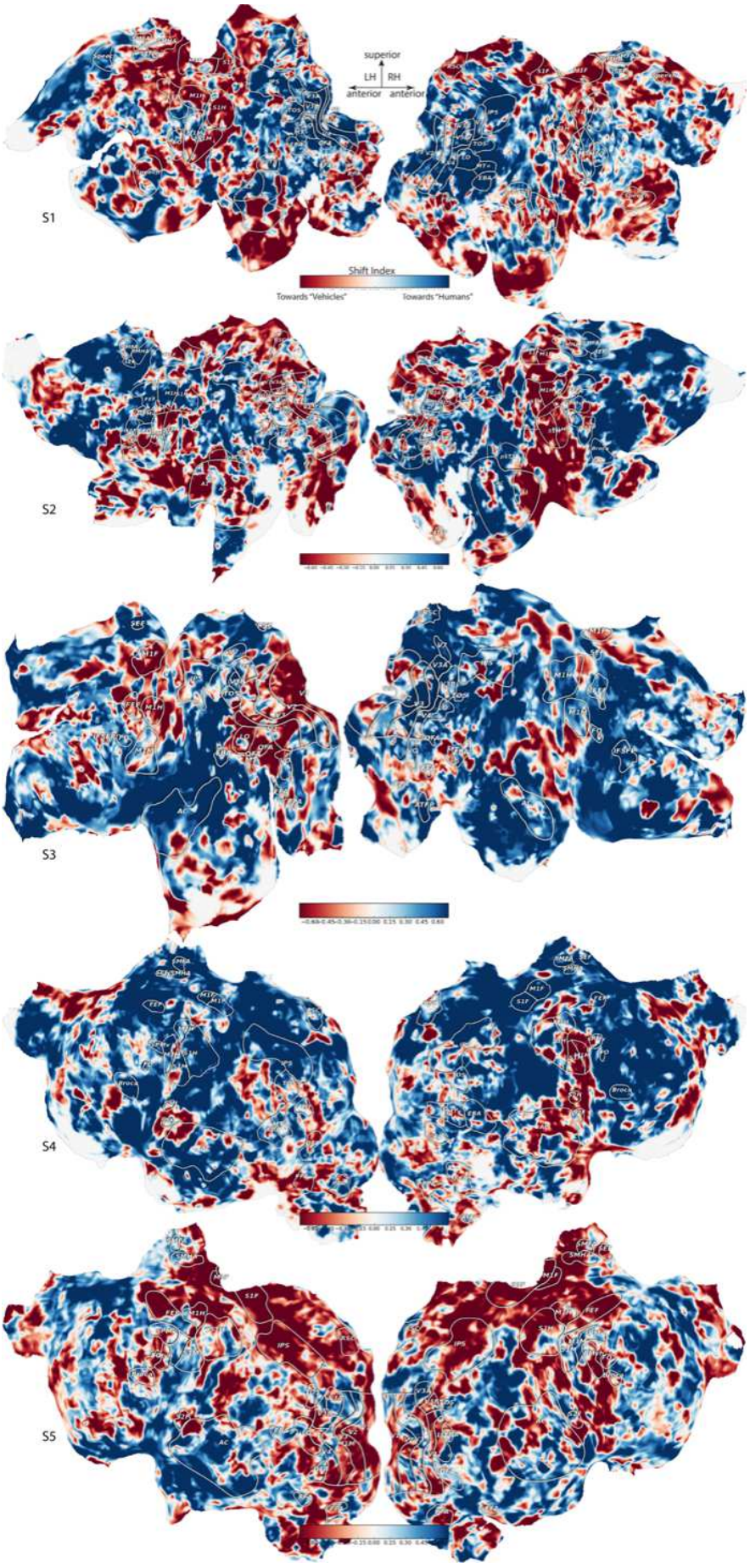

Figure B.4: **Cortical flat maps of shift index of "instrument" categories for the five subjects.** The voxel-wise shift index of "instrument" categories is illustrated on the cortical flat map for the five subjects(S1-S5). Positive shift index values mean that the semantic-tuning distribution of divided attention is more close to semantic-tuning distribution during the attend to "humans" condition and negative values indicate bias towards semantic-tuning distribution while attending to "vehicles" condition.

Appendix C

Object and action categories

There is a diverse set of object and action categories studied in this thesis. Important groups of object categories which are used in this thesis are listed here.

VEHICLE	HUMAN	ANIMAL
bus.n.01	contestant.n.01	whale.n.02
carriage.n.02	institution.n.01	rhinoceros.n.01
airplane.n.01	spectator.n.01	vertebrate.n.01
conveyance.n.03	church.n.01	heron.n.02
subway_train.n.01	clergyman.n.01	ungulate.n.01
starship.n.01	priest.n.01	elasmobranch.n.01
helicopter.n.01	male.n.02	coyote.n.01
vehicle.n.01	metro.n.01	zebra.n.01
ship.n.01	gathering.n.01	camel.n.01
motor_scooter.n.01	enemy.n.02	reptile.n.01
small_boat.n.01	organization.n.01	shark.n.01
military_vehicle.n.01	creator.n.02	meerkat.n.01
public_transport.n.01	preacher.n.01	diapsid.n.01
school_bus.n.01	oarsman.n.01	horse.n.01
bicycle.n.01	leader.n.01	viverrine.n.01
rocket.n.01'	player.n.01	wolf.n.01
jet.n.01	male_child.n.01	wading_bird.n.01
container_ship.n.01	soldier.n.01	equine.n.01
cargo_ship.n.01	window_washer.n.01	bird.n.01
baby_buggy.n.01	rider.n.01	mammal.n.01
boat.n.01	parachutist.n.01	canine.n.02
fighter.n.02	student.n.01	fish.n.01
jetliner.n.01	cowboy.n.01	flamingo.n.01
spacecraft.n.01	defender.n.01	aquatic_bird.n.01
sled.n.01	watchman.n.01	seabird.n.01
dinghy.n.01	religious.n.01	feline.n.01
vessel.n.02	person.n.01	snake.n.01
warship.n.01	parishioner.n.01	big_cat.n.01
canoe.n.01	motorcyclist.n.01	pelican.n.01
train.n.01	pilot.n.01	placental.n.01
barge.n.01	avant-garde.n.01	cat.n.01
heavier-than-air_craft.n.01	shopper.n.01	penguin.n.01
sailboat.n.01	worker.n.01	aquatic_mammal.n.01
luge.n.01	farmer.n.01	hippopotamus.n.01
wheeled_vehicle.n.01	nun.n.01	cattle.n.01
craft.n.02	pope.n.01	dog.n.01
	golfer.n.01	bison.n.01
	female.n.02	carnivore.n.01
	spiritual_leader.n.01	odd-toed_ungulate.n.01
	female_child.n.01	lion.n.01

VEHICLE	HUMAN	ANIMAL
	traveler.n.01	giraffe.n.01
	crowd.n.01	ruminant.n.01
	social_group.n.01	even-toed_ungulate.n.01
	religious_person.n.01	bovid.n.01
	skydiver.n.01	old_world_buffalo.n.01
	skilled_worker.n.01	
	guard.n.01	
	pedestrian.n.01	
	demonstrator.n.03	
	child.n.01	
	policeman.n.01	
	golf_club.n.01	
	diver.n.01	

INSTRUMENTS	MAN-MADE SCENES	NATURAL SCENES
spectacles.n.01	garage.n.01	horizon.n.01
headlight.n.01	church.n.02	wind.n.01
source_of_illumination.n.01	stairs.n.01	snow.n.01
control.n.09	way.n.06	snow.n.02
candle.n.01	expressway.n.01	discharge.n.05
timer.n.01	path.n.02	geographic_point.n.01
ski_jump.n.01	hotel.n.01	peak.n.04
rotating_mechanism.n.01	doorway.n.01	weather.n.01
measuring_instrument.n.01	road.n.01	rooftop.n.01
mace.n.01	temple.n.01	smoke.n.01
mechanical_device.n.01	place_of_worship.n.01	cloud.n.02
ski.n.01	building.n.01	region.n.01
combine.n.01	driveway.n.01	port.n.01
lifting_device.n.01	stairway.n.01	fog.n.01
goggles.n.01	crossing.n.05	mist.n.01
treadmill.n.01	street.n.01	precipitation.n.03
weapon.n.01	house.n.01	location.n.01
ramp.n.01	highway.n.01	sunset.n.02
shelf.n.01	entrance.n.01	atmospheric-phenomenon.n.01
support.n.10	skyscraper.n.01	flood.n.01
medical_instrument.n.01	sidewalk.n.01	hole.n.01

INSTRUMENTS	MAN-MADE SCENES	NATURAL SCENES
lantern.n.01		natural_phenomenon.n.01
sunglasses.n.01		outer_space.n.01
spear.n.01		cloud.n.01
segway.n.01		aerosol.n.01
button.n.01		physical_phenomenon.n.01
harvester.n.02		point.n.02
machine.n.04		line.n.11
crane.n.04		rain.n.01
lamp.n.01		opening.n.01
hook.n.04		
automaton.n.02		
restraint.n.06		
blade.n.08		
speedometer.n.01		
seat_belt.n.01		
push_button.n.01		
stroboscope.n.01		
laser.n.01		
fan.n.01		
machine.n.01		
fastener.n.02		
electro-acoustic_transducer.n.01		
stringed_instrument.n.01		
microphone.n.01		
musical_instrument.n.01		
mechanism.n.05		
device.n.01		
surgical_instrument.n.01		
elevator.n.01		
bomb.n.01		
timepiece.n.01		
streetlight.n.01		
electrical_device.n.01		
guitar.n.01		
gun.n.01		
optical_instrument.n.01		
missile.n.01		
instrument.n.01		

Table C.1: **Distinct groups of object categories used in the thesis.** Object and action categories belonging to the six groups of categories studied in this thesis are presented here.