**Title**
Computational Modeling of Cortical and Behavioral Responses to Emotional Stimuli



**Author**
Abdel-Ghaffar, Samy Ahmed Mansour

**Publication Date**
2018

Peer reviewed|Thesis/dissertation

# Computational Modeling of Cortical and Behavioral Responses to Emotional Stimuli

By

Samy Abdel-Ghaffar

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sonia Bishop
Professor Jack Gallant
Professor Bruno Olshausen
Professor David Whitney

Fall 2018

Computational Modeling of Cortical and Behavioral Responses to Emotional Stimuli

Abstract

Computational Modeling of Cortical and Behavioral Responses to Emotional Stimuli

by

Samy Abdel-Ghaffar

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Sonia Bishop, Chair

Emotions are stereotyped responses to situations of high survival value. Recognition of these high survival value situations is a necessary precondition to emotional elicitation. As vision is the primary human sensory modality, visual perception of an emotional situation is often responsible for the elicitation of emotion in humans. This raises the question: how does the human brain extract the emotional content from patterns of light on the retina in order to initiate appropriate behavioral responses that are adaptive for survival? And how are these emotional scenes differentially processed from non-emotional scenes? In order to address these questions I conducted a series of experiments using techniques and theory from a broad range of disciplines including cognitive neuroscience, machine learning, psychophysics, and affective science. I motivate, describe, and interpret these experiments in the five chapters of this dissertation.

Chapter 1 contains a brief review of the previous scientific findings that motivated this dissertation. Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging modality which allows researchers to record brain activity with high spatial resolution. Traditional techniques used in the analysis of fMRI data have recently been complimented by computational machine learning techniques, giving neuroscientists powerful new tools to infer the brain's mechanisms and representations. Voxel-wise modeling (VWM) is one such method. Chapter 2 describes a fMRI experiment using a large corpus of naturalistic emotional images and VWM analysis to model brain representations of a combined semantic and emotional feature space. Principal components analysis (PCA) of voxel tuning was then used to uncover the primary dimensions of representation within occipital-temporal cortex (OTC). Alongside animacy, the valence and arousal of animate stimuli are primary dimensions of OTC tuning. Furthermore, this tuning is better able to predict the appropriate behavioral response to the images that subjects viewed than are the semantic and emotional image features used to model the fMRI data. These findings suggest that OTC representations of naturalistic emotional images may be used by other brain regions to elicit appropriate behavioral responses to situations of high survival value depicted in these emotional scenes.

In order to address several theoretical limitations of the study described in chapter 2, I conducted two literature reviews in Chapter 3. The first reviews studies on congenitally blind subjects which suggest that OTC contains supramodal semantic representations that may be idiosyncratically tuned to support appropriate behavior responses. The second reviews the

literature on attention and perception of emotional images and the neural mechanisms that subserve these processes. Additionally, Chapter 3 describes seven fMRI analyses of the data from chapter 2 that address empirical limitations of chapter 2. These analyses include a representational similarity analysis (RSA), univariate SPM analysis, variance partitioning of the VWM model in chapter 2, and several control analyses. The findings from these analyses further boost our claims from chapter 2 that alongside animacy, the valence and arousal of animate stimuli are represented within OTC.

Often situations of high survival value occur quickly, or are only perceived briefly, and thus the human visual system has little information from which action can be taken. Turning from cortical to behavioral responses from naturalistic emotional images, in chapter 4 I describe an experiment using ultra-rapid image presentation to explore the limits of human performance in semantic and emotional valence categorization during a brief glance. Using stimuli and categorization tasks from a broader range of semantic and valence categories than had been done in previous studies, along with controls for observed response bias, I found that humans can accurately categorize both the semantic category (animal, human, object, or building) and emotional valence (negative, neutral or positive) of an image when presented for as little as 17ms and backwards masked. Furthermore, when the image depicted an emotionally negative scene, semantic performance was significantly worse than that for images depicting neutral or positive scenes. I also found that only when subjects successfully categorized the semantic category of an image was valence categorization also above chance across several valence by semantic category conditions. The converse was not true however. This suggests that the semantic information of an image must first be extracted before the emotional information can be extracted. This finding supports the cognitive primacy hypothesis as opposed to the affective primacy hypothesis, two competing hypotheses of visual emotional processing in the field of affective science. Finally, in chapter 5, I interpret the results of the experiments described in chapters 2-4 taken as a whole, and offer closing remarks.

# Dedication

To Jessica Brown, without whom this wouldn't have been possible. And to my family, and friends that are family, for all the support, laughs and consolation.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1. Introduction

You are walking down the street, you turn a corner, and there is a huge dog sitting directly in your path. Is that Noot-Noot, the friendly neighborhood dog? Or is that an unknown dog that will attack? Does it have rabies? Is that a service dog belonging to the woman standing nearby? Or is it a mother simply concerned with protecting her young pups several feet away? You may only have a brief moment to infer answers to these questions, which have huge ramifications for your well being. So what do you do? And more importantly from the scientific perspective, how does your brain determine what to do? Each of the above possibilities affords a different action, and the first stage in determining the appropriate response involves a visual discrimination task. In the remainder of this introduction I will briefly review some of the psychological and neuroscientific literature pertinent to answering the question: how does the human visual system extract semantic and emotional information useful in determining the appropriate behavioral response to situations of high survival value?

Threatening situations demand quick responses to avoid bodily harm. In now foundational work stemming from the study of phobias, Seligman (1971) and Ohman (1975, 2001) have argued that negative, threatening and animate stimuli such as spiders and snakes have been "biologically prepared" by evolution to evoke fast and automatic emotional responses to avoid injury and death. They hypothesized that these negative animate stimuli are conferred with "prioritized processing" across cognitive processes such as perception, attention, memory, and decision making. Decades of research have supported this hypothesis. Negative animate stimuli, many of which are images of fearful or angry human faces, have been shown to demand attentional resources across many tasks. For example, these negative animate images better grab attention when presented as distractor stimuli and are easier to spot in visual search tasks (for a review see Pourtois, Schettino, & Vuilleumier, 2013). Perception of so-called biologically prepared stimuli is also differentially affected relative to non-emotional stimuli. In continuous flash suppression studies, where a fixed image shown to one eye is suppressed by a stream of rapidly changing images presented to the other eye in order to render the fixed image non-conscious, fearful faces break into consciousness faster than do happy faces (Yang, Zald, & Blake, 2007).

Ohman has also hypothesized that a dedicated neural system subserves this observed prioritized processing of biologically prepared stimuli (Ohman & Mineka, 2001). In what has come to be known as the "two-route hypothesis", "subcortical fast route", or "low-road", researchers argue that a subcortical circuit beginning in the retina, connected to the superior colliculus, moving on to the pulvinar and ending at the amygdala is where this prioritized processing occurs (for a review see Tamietto & de Gelder, 2010). Prioritized processing has also been observed in stimuli other than negative animate images (for example see Anderson, 2005, Eastwood et al., 2001), and thus the two-route hypothesis has expanded to account for all emotional stimuli (Tamietto & de Gelder, 2010). The two-route hypothesis is contentious however, and other authors argue that multiple waves or stages of processing within higher order visual regions located in occipital-temporal cortex (OTC) suffice to explain the observed prioritized processing of emotional visual stimuli (Pessoa & Adolphs, 2010; Pourtois, Schettino, & Vuilleumier, 2013). Numerous studies have found evidence for object category selectivity within OTC (for a review see Grill-Spector & Weiner, 2014). Regions such as the fusiform face areas (FFA) are known be selective to faces (Kanwisher, McDermott, & Chun, 1997), the

extrastriate body area (EBA) is selective to bodies (Downing et al, 2001) and the parahippocampal place area (PPA) is selective to landscapes and buildings (Epstein & Kanwisher, 1998), to name a few.

Related to questions of where in the brain prioritized processing of emotional stimuli occurs is a debate concerning the temporal order of semantic vs. emotional information extraction from visual stimuli. The affective primacy hypothesis postulates that emotional information is extracted before, and parallel to, semantic information from visual stimuli (Murphy & Zajonc, 1993; Zajonc, 1980). Authors arguing for the affective primacy hypothesis often invoke the subcortical fast-route as the neural system that subserves this fast and parallel processing, as it is independent from the cortical visual system known to extract semantic information (OTC) and has been thought to contain fewer synaptic junctions making it a fast processing unit (Tamietto & de Gelder, 2010, though see Pessoa & Adolphs, 2010 for neurophysiological evidence that OTC is just as fast). Conversely, the cognitive primacy hypothesis states that semantic information must first be extracted from visual stimuli, then the emotional content of the image can be extracted based on that semantic content (Cohen, Henik, & Mor, 2011; van Steenbergen, Band, & Hommel, 2011). As the OTC is well known to represent visual semantic content, the "two-route" or "multi-wave" hypothesis for prioritized processing of emotional stimuli can be viewed as consistent with the cognitive primacy hypothesis if one considers the possibility that OTC is a region where the the combined semantic and emotional information of visual stimuli is represented. In this view, a fast feed-forward sweep of information processing within OTC first extracts basic semantic information of a visual stimulus that is then sufficient to infer the emotional content of the stimulus conferring it with prioritized processing. Later processing stages within OTC that incorporate reentrant feedback from other brain regions could then further refine the semantic and emotional representations of the visual stimulus. In the studies described in chapters 2-4 of this dissertation we present evidence in support of this view.

# Chapter 2. Occipital-temporal cortical tuning to emotional natural images predicts associated behaviors.

## Introduction

The ability to recognize and respond appropriately to emotionally salient stimuli is essential to a species' evolutionary success. Adaptive responses to dangerous situations carry a survival advantage, while identification of potential mates facilitates reproductive success. There has been considerable interest in identifying the brain mechanisms that support the recognition of emotional stimuli. A popular theory is that a subcortical pathway extending from the retina to the superior colliculus, through the pulvinar, and on to the amygdala enables the rapid detection of danger (LeDoux, 1994; Tamietto and de Gelder, 2010). This theory has been widely adopted to explain the processing of emotional visual stimuli in humans (Dolan and Vuilleumier, 2003; Ohman et al., 2007). However, evidence for this subcortical pathway primarily comes from research in rodents using auditory stimuli (LeDoux, 1994 ). Recent studies have questioned the existence of an equivalent pathway for the processing of visual stimuli in primates (Pessoa and Adolphs, 2010). Further, it is unclear whether such a subcortical pathway could convey the information necessary to guide selection of behavioral responses to different types of emotional stimuli. The proposed subcortical pathway has largely been held to convey information about stimulus valence (positive versus negative value) and to inform engagement in approach versus avoidance. However, the behavioral decisions we make when confronted by emotional stimuli are more complex than this. For example, encountering a large bear versus a weak, diseased, animal should promote different types of avoidance responses, while human infants and potential mates should prompt different types of approach responses. As a result, an optimal mechanism for guiding responses to emotional stimuli likely needs to represent more than stimulus valence alone.

So how might we achieve rapid selection of appropriate responses to different types of emotional stimuli? Here, we argue that a potentially effective solution might entail the combined local representation of emotional and semantic information within the same brain region. The occipital temporal cortex (OTC) is a strong candidate for such a region. Information reaches the OTC as rapidly as it reaches the amygdala, within 80-140ms (Liu et al., 2009). In addition, neuroimaging studies have revealed rich patterns of semantic selectivity in OTC. For example, different parts of OTC preferentially respond to various subclasses of animate and inanimate stimuli, including faces, bodies, places and objects (Kanwisher, McDermott, and Chun, 1997; Gauthier et al., 2000; Downing et al., 2001; Epstein and Kanwisher, 1998; O'Craven and Kanwisher, 2000; Malach et al. 1995; Hansen, Kay and Gallant, 2007; Connolly et al., 2012).

Relatively little is known regarding the representation of stimulus emotional content within OTC. There is some evidence to suggest that stimulus valence and stimulus arousal (emotional evocativeness, intensity) can modulate OTC activity (Chikazoe et al., 2014; Mourão-Miranda et al., 2003; Bradley et al., 2003; Sabatinelli et. al 2007). However, valence and arousal are rarely modeled independently in the same experiment making it difficult to interpret these findings. In addition, there has been little attempt to investigate whether there is combined representation of the semantic and emotional content of stimuli within OTC. A single study to date examined both stimulus animacy and stimulus valence but concluded that these were

represented within different parts of OTC (Chikazoe et al., 2014). Such prior studies have been limited in their scope by adoption of traditional experimental designs with a small number of conditions and relatively few stimuli. If we seek to examine how the semantic and emotional content of stimuli is co-represented, we need a method for modeling brain responses to a wide range of stimuli varying in semantic and emotional features. Outside of the emotion field, a voxel-wise multi-feature encoding model framework has been developed that facilitates analyses of this nature (Kay et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011; Naselaris et al. 2011). In the study reported here, we used this framework to investigate the representation of diverse emotional natural images at a single voxel level, across OTC. In addition, we explored whether this representation might suffice to guide behavioral responses to emotional stimuli.

Our findings revealed representation of natural image semantic and emotional content across much of OTC. Principal components analysis of voxel-wise model weights indicated that many OTC voxels are co-tuned to semantic and emotional image content. Considering the first three components alone, we found co-representation of image animacy and valence as well as image animacy and arousal. Scores on these three components explained 20% of the variance in novel subjects' selection of behavioral responses to match image content. Using scores on the first twenty components explained 40% of variance in novel subjects' selection of behavioural responses to match image content. This prediction of behavioral responses from OTC components based on tuning to both semantic and emotional image content significantly exceeded that achieved using OTC tuning to low-level structural image properties or semantic or emotional content alone. It also significantly exceeded the variance explained by components from PCA performed directly on stimulus features, across images viewed. This is consistent with OTC extracting those aspects of semantic and emotional image content that are of value to guiding behavior.

# Results

## Using a multi-feature encoding modeling approach to investigate the representation of natural image semantic and emotional content.

Our experimental stimuli comprised 1620 images varying widely in semantic and emotional content (see Star Methods for details). We labeled each image for features of interest. Using ridge regression, we fit multi-feature encoding models to the functional magnetic resonance imaging (fMRI) data acquired while subjects viewed the images (see Star Methods and Fig. 2-1A). We tested alternate models by changing the features included as regressors. To have sufficient statistical power to fit the multi-feature encoding models used required the acquisition of a large amount of fMRI data per subject. The six subjects in our study each completed fifty fMRI scans across six 2hr sessions. Thirty training scans, each 7.5 minutes long, were used for model estimation, and twenty test scans, each 6 minutes long, were used for model validation. Images were viewed for 1s each, with a 3s inter-stimulus interval. Subjects maintained foveal fixation and either categorized the valence (negative, positive, neutral) or the broad semantic category (human, animal, building, object, scene) of each image. We used these two different tasks to ensure that cortical tuning to stimulus emotional content was not only observed when task-relevant. Subjects who performed the semantic categorization task in the scanner subsequently categorized the images by valence in a post-scan behavioral session. All

subjects also rated the images for emotional arousal in this post-scan session (see Star Methods for details).

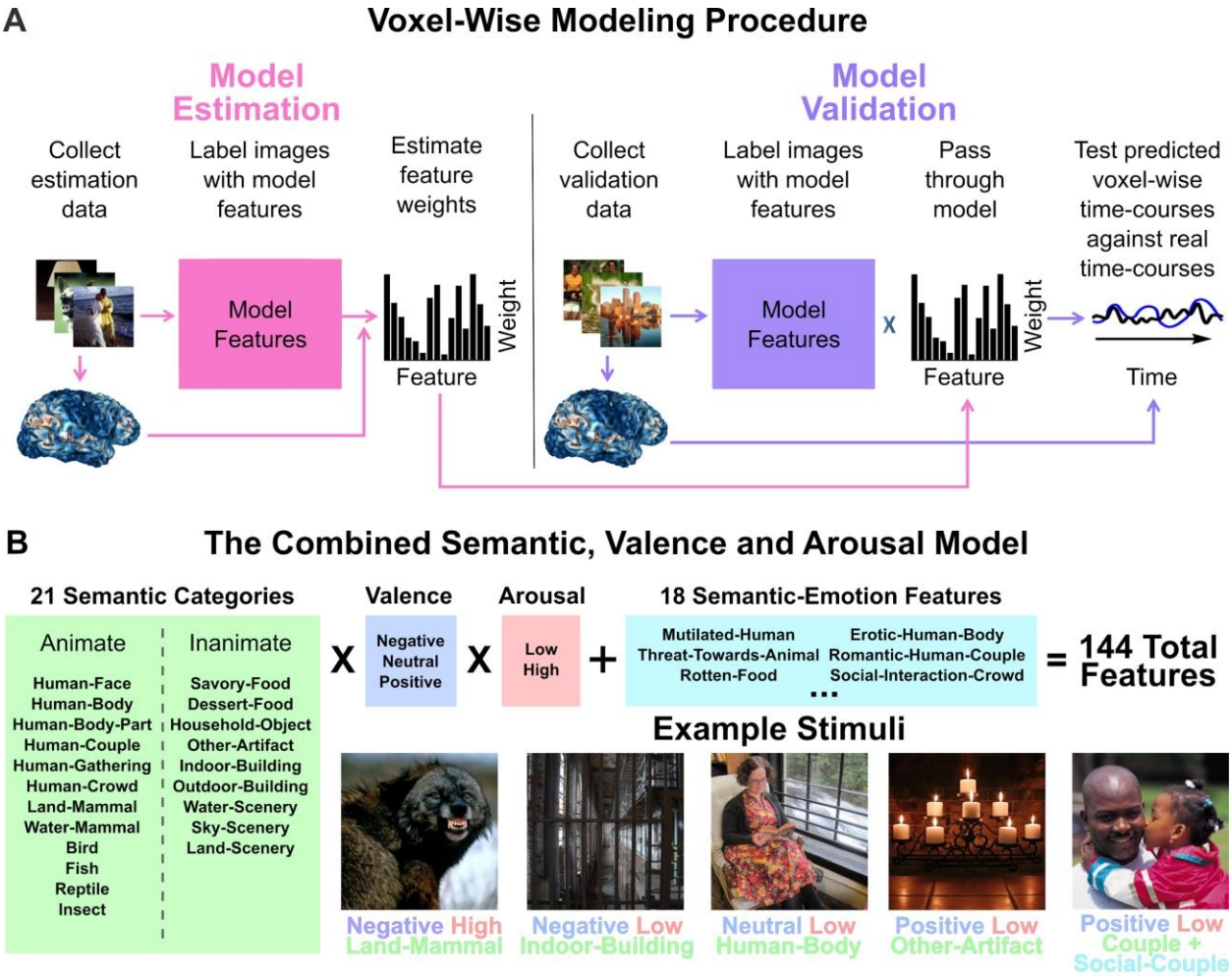


**Figure 2-1. The modeling procedure and the CSVA model.**
(A) BOLD data collected while subjects viewed 1440 images were used for model estimation. Ridge regression was used to fit each model to the BOLD time-series for each voxel, using a finite impulse response function with four 2s time-bins. Weights were estimated for each model feature for each time-bin. These weights characterize each voxel's response profile or 'tuning' to model features. Model validation was conducted using independent fMRI data collected while subjects viewed 180 novel images. Voxel-wise feature weights were used to generate a predicted time-series for each voxel. This was correlated with the recorded BOLD time-series to obtain a metric of model fit that controls for over-fitting, see Star Methods for details. (B) The Combined Semantic, Valence and Arousal (CSVA) model comprises 126 mutually exclusive features denoting image semantic category (21 categories), valence (3 levels) and arousal (2 levels) and 18 additional semantic-emotion (SE) features that carry both semantic and emotional information (e.g. mutilated animal; rotten), see Star Methods for further details. Image semantic category and SE features were labeled by four independent raters; image valence and arousal were assessed by each subject, see Star Methods for further details. Here, five example images are labeled with CSVA features using one subject's valence and arousal responses.

## Mapping tuning to semantic and emotional image features across cortex.

Our primary model, the Combined Semantic, Valence and Arousal (CSVA) model (Fig. 2-1B), describes each image using a combination of mutually exclusive semantic categories, subjects' subjective valence and arousal judgments and a number of additional semantic-emotional features carrying information about both semantic and emotional content (e.g. 'rotten food', 'mutilated animal'); see Star Methods for details. FMRI data from the model estimation runs were concatenated and ridge regression used to fit the CSVA model to each subject's BOLD data using a finite impulse response function with four 2s time-bins. Voxel-wise weights were estimated for each model feature for each time-bin. These weights were applied to the values of feature regressors for the images viewed during the validation scans, generating predicted BOLD time-courses for each voxel. We correlated these predicted time-courses with the observed validation BOLD time-courses to obtain estimates of model prediction accuracy for each voxel (see Star Methods for further details). Figure 2-2 shows the resulting prediction accuracies projected onto the cortex of each individual subject. Figure 2-2A shows whole cortical prediction accuracy maps for each subject. These maps reveal that the CSVA model significantly predicts validation BOLD time-courses across a wide stretch of OTC. Figure 2-2B shows a close up of OTC for one example subject.

In principle, the CSVA model could significantly fit subjects' BOLD data simply through capturing BOLD responses to image semantic or emotional content alone. Given this, we compared the fit of the CSVA model against that of a model containing only the semantic category features from the CSVA model (the Semantic Only model) as well as against that of a model containing only the valence and arousal features (the Emotion Only model). These models were fit to the estimation data in the same manner as for the CSVA model. We then used the validation data to calculate voxel-wise prediction accuracies and compare these between models. A bootstrap procedure was used to compare CSVA model fit against that of each of the other two models (see Star Methods for details). Model comparison was restricted to voxels where at least one of the three models showed a significant fit.

The CSVA model outperformed the Emotion Only model at a group level and for all six subjects considered individually ($ps<0.05$), fig. 2-S1. The CSVA model also outperformed the Semantic Only model at the group level and for four out of six subjects considered individually ($ps<0.05$). CSVA model superiority to the Semantic Only model was apparent in stretches of OTC adjacent to, and overlapping with, regions with known semantic selectivity including the posterior superior temporal sulcus (pSTS), extrastriate body area (EBA), Occipital Face Area (OFA) and Fusiform Face Area (FFA), fig. 2-S2. In a supplementary analysis, we examined whether coding stimulus emotional content differentially improved model fit for animate versus inanimate stimuli. We addressed this by adapting the Semantic Only model such that either animate semantic category regressors or inanimate semantic category regressors were replaced by regressors coding for semantic category, valence and arousal, fig. 2-S3A. All study participants showed a significantly greater increase in prediction accuracies, relative to the Semantic Only model baseline, when coding of emotional content was added for animate as opposed to inanimate stimuli, fig. 2-S3B. These results suggest that co-representation of stimulus emotional and semantic content is stronger for animate than for inanimate stimuli, potentially reflecting the greater evolutionarily relevance of emotional animate stimuli. Many of the voxels where prediction accuracies increased as a result of coding the emotional content of animate versus inanimate stimuli were located within OTC, fig. 2-S3C. Building upon these initial

findings, we used CSVA voxel-wise model weights to further examine the structure of voxel-wise tuning to stimulus animacy and emotional content within OTC. We report these analyses next.

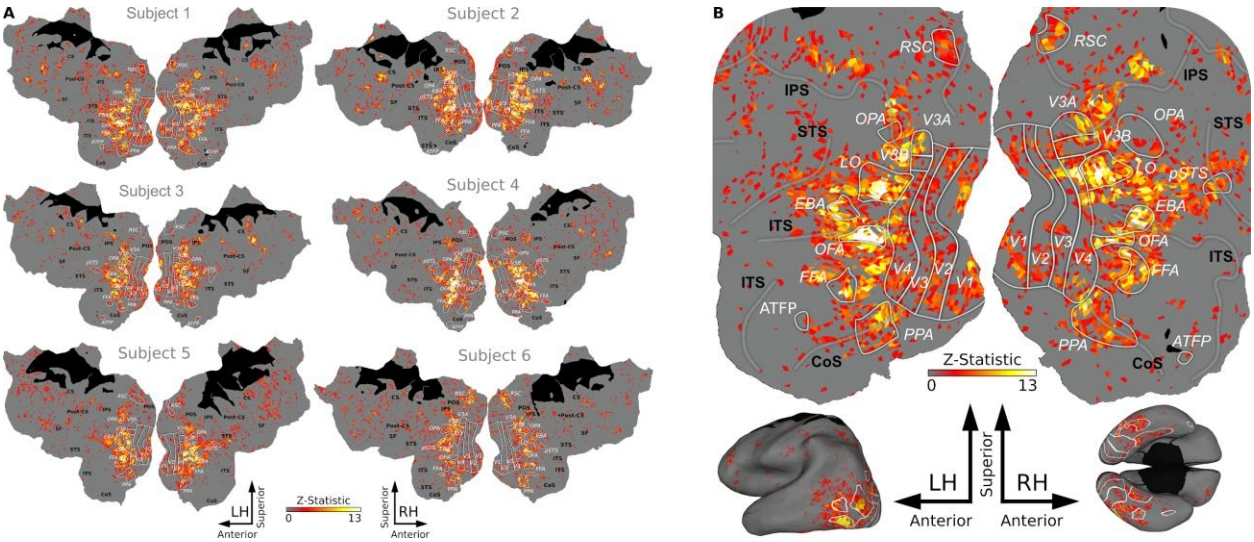


**Figure 2-2. CSVA model voxel-wise prediction accuracy scores mapped onto cortex.**
(A) Voxels where activity was significantly predicted by the Combined Semantic, Valence and Arousal (CSVA) model are shown on cortical maps for all 6 subjects. For each voxel, prediction accuracies were calculated using the z-transformed correlation between the CSVA model predicted time-course and the recorded BOLD time-course for the validation dataset. Significance was assessed by permutation testing, see Star Methods for details. The CSVA model significantly predicts validation BOLD time-courses across much of OTC. This was consistently observed across subjects. (B) Prediction accuracy scores for subject 1; the cortical map is cropped (top) to zoom in on OTC. In addition, prediction accuracy scores are projected onto inflated lateral (bottom left) and ventral (bottom right) cortical surfaces. Note: Regions of interest (ROIs) are labeled in white, sulci in black. RSC: Retrosplenial Complex, OPA: Occipital Place Area, LO: Lateral Occipital cortex, pSTS: Posterior Superior Temporal Sulcus, EBA: Extrastriate Body Area, OFA: Occipital Face Area, FFA: Fusiform Face Area, PPA: Parahippocampal Place Area, ATFP: Anterior Temporal Face Patch. IPS: Intraparietal Sulcus, STS: Superior Temporal Sulcus, ITS: Inferior Temporal Sulcus, CoS: Collateral Sulcus

## PCA on CSVA model feature weights reveals consistent patterns of OTC tuning to stimulus animacy, emotional arousal and emotional valence across subjects.

The modeling framework used here involves the construction and fitting of encoding models as opposed to adoption of a decoding approach. The benefit of this approach is that we can go beyond simply assessing model fit and interrogate voxels' response profiles, i.e. their 'tuning', to different stimulus features. The CSVA model includes over 100 features. As a result, it is unwieldy to examine and interpret spatial patterns of tuning for each individual feature. Instead, we can seek to identify the main structure underlying similarities and differences in feature response profiles, across voxels. This can be achieved by applying principal components analysis (PCA) to model feature weights across a given set of voxels (Huth et al., 2012).

We conducted a group-level PCA on CSVA model feature weights across OTC voxels where the CSVA model fit significantly and predicted validation BOLD time-series better than the Semantic Only model (See Star Methods for details; we note that expanding voxel selection to include all cortical voxels where the CSVA model fit significantly produced a highly similar

PCA solution, fig. 2-S4). Each of the top three PCs from the group-level PCA accounted for significantly more variance than could be explained by covariance between stimulus features alone (Fig. 2-3A); see Star Methods. In addition, these group PCs correlated highly with PCs extracted when PCA was performed separately on each subject's data (Fig. 2-3B), indicating consistency in the structure of OTC tuning to CSVA model features across subjects. We note that this holds for all subjects, including subjects 2 and 4 who categorized images semantically while fMRI data were acquired (see Star Methods).

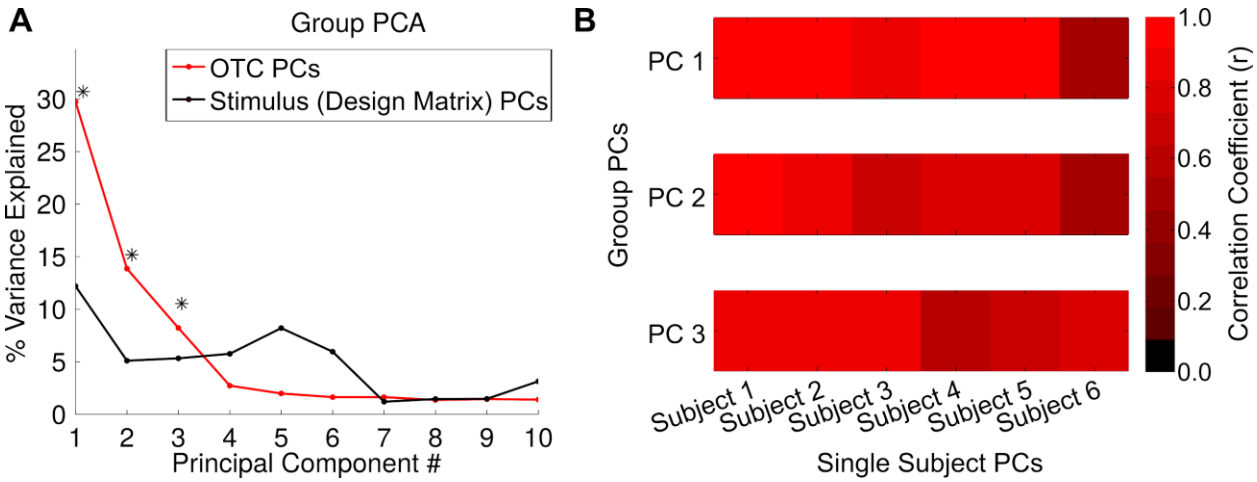


**Figure 2-3. Results of PCA on CSVA model feature weights, across OTC voxels.**
(A) A group-level principal components analysis (PCA) was conducted on the CSVA model feature weights across all OTC voxels where model fit was significant and better than that of the Semantic Only model. The scree plot shows the amount of variance explained by each of the top ten PCs (in red). PCs from a PCA analysis conducted on stimulus features (using the combined design matrix from all 6 subjects) are shown in black. Asterisks indicate group PCs that explain significantly more variance than the stimulus PCs (jackknife test, $p < 10\text{-}8$), see Star Methods for details. (B) Correlations between the top 3 group PCs and the top 3 PCs from single subject PCAs. The significant correlation coefficients between each of the group PCs and the corresponding PCs from the single subject PCA for each subject ($ps < 10\text{-}8$) indicate a shared representational structure across subjects.

To explore OTC co-representation of image animacy and emotional content, we compared the feature loadings of the top three group PCs to those of seven dimensions of theoretical interest (Fig. 2-4). The first of these 'theoretical' dimensions comprised a previously adopted 4-level index of animacy (human, other mammal, non-mammalian vertebrate invertebrate, and inanimate; Connolly et al., 2012). The second comprised a binary index of whether image content was animate or inanimate. The third indexed whether humans were present in each image or not. The remaining four dimensions separately indexed the valence and arousal of animate and inanimate stimuli (see Star Methods for further details). We correlated the feature loadings of these theoretical dimensions with those of the three group PCs; this provides a measure of how well each theoretical dimension explains the information carried by each PC. Bootstrapping was used to determine correlation significance at $p<0.05$ (see Star Methods for further details). This correlational analysis revealed that PC 1 represented both stimulus animacy and stimulus arousal (Fig. 2-4). PC 2 primarily represented the arousal value of animate, but not inanimate, stimuli. PC 3 represented the valence of animate, but not inanimate, stimuli. Both PC2 and PC3 also carried information about animacy in general, though more weakly than PC1.

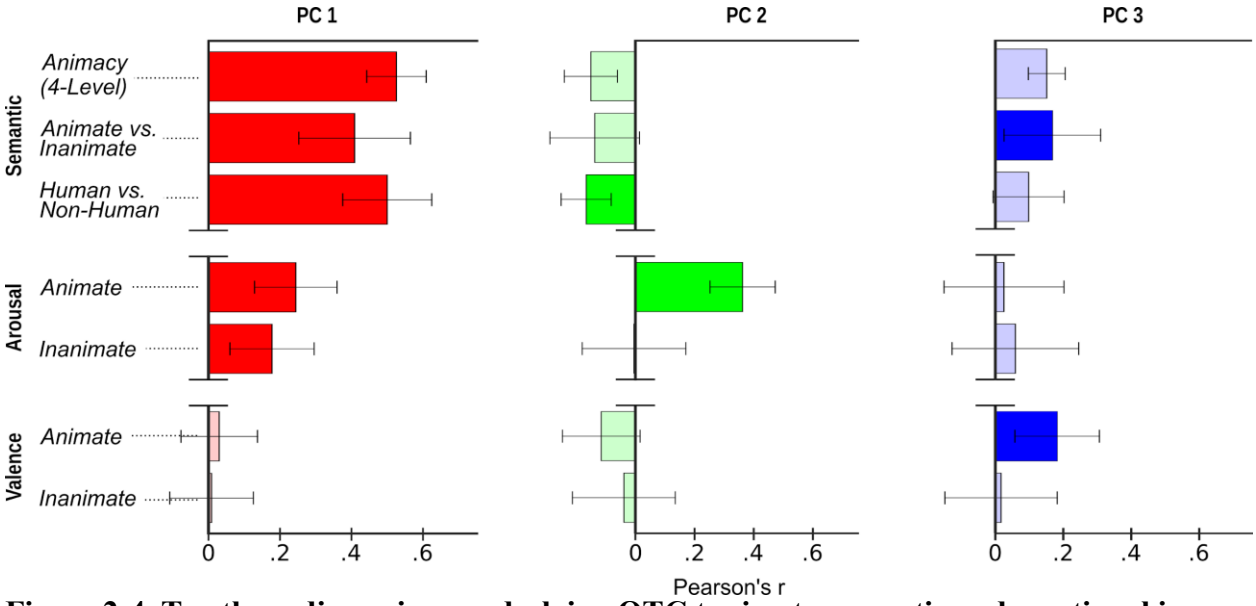


**Figure 2-4. Top three dimensions underlying OTC tuning to semantic and emotional image features carry information about stimulus animacy and its interactions with stimulus arousal and stimulus valence.**

A group-level PCA was conducted on CSVA model feature weights across OTC voxels (see Fig. 2-3). Feature loadings on the top three PCs were correlated with feature loadings on theoretical dimensions of interest (given on the y axis). Bootstrapping was used to determine correlation significance (see Star Methods for further details). Saturated color indicates correlations significant at $p < 0.05$. PC1 carries information about stimulus animacy and the arousal value of both animate and inanimate stimuli. PC2 carries information about the arousal of animate stimuli; PC3 carries information about the valence of animate stimuli. PC2 and PC3 also show some tuning to animacy in general, though more weakly than PC1. Note. Theoretical dimensions: arousal is coded as high (+1) or low (0), valence is coded as positive (1), neutral (0), negative (-1), see Star Methods for details.

To visualize the spatial structure of tuning captured by these PCs, we projected voxel-wise PC scores onto maps of OTC for each subject (Fig. 2-5 and fig. 2-S5). These maps show clear hemispheric symmetry in voxel-wise feature tuning across OTC; they also reveal commonalities in spatial transitions in tuning across subjects.

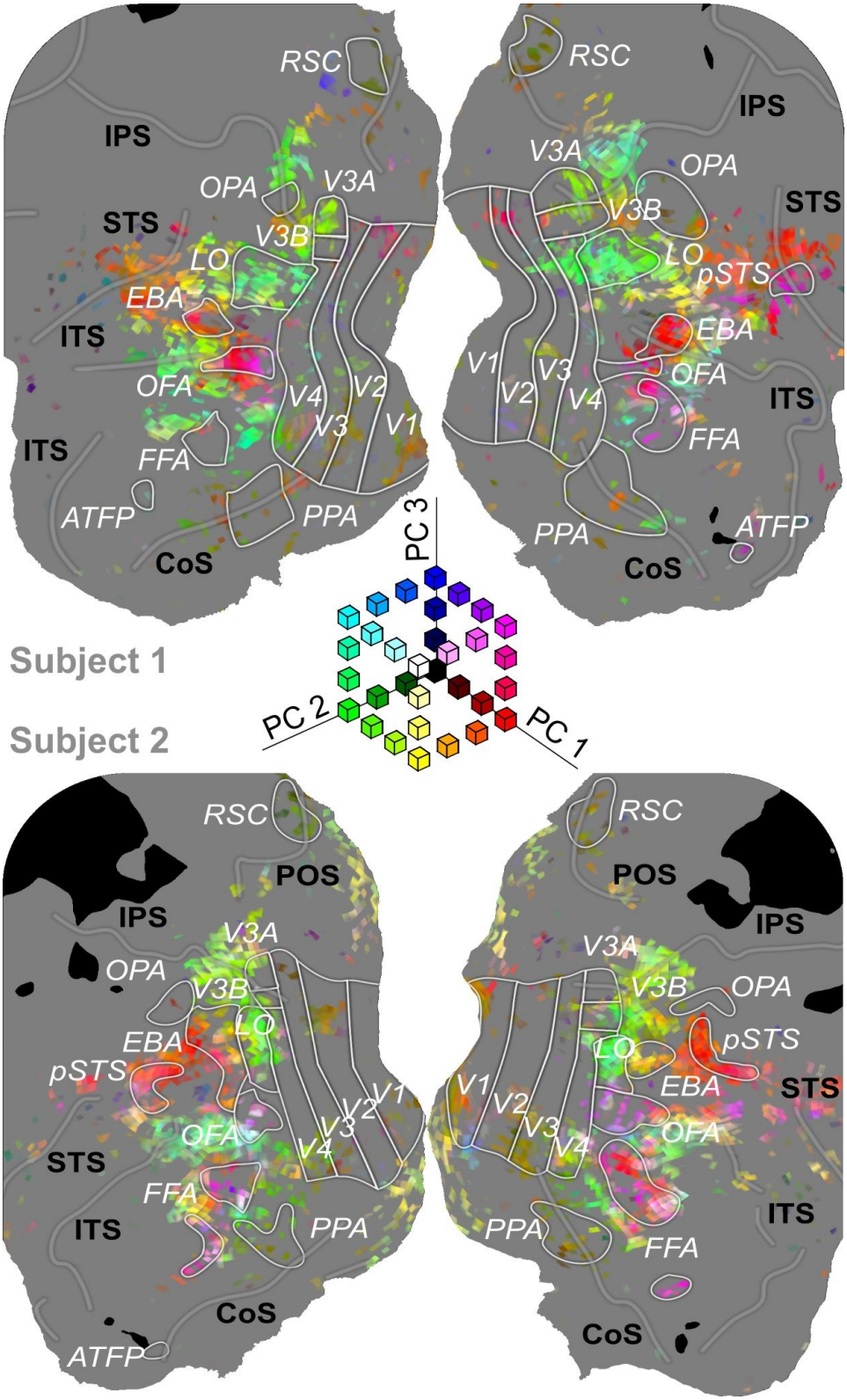


**Figure 2-5. Structure of OTC tuning as captured by PCA on CSVA model feature weights.** Mapping PC scores from the group-level PCA of CSVA model feature weights, across OTC voxels, reveals considerable spatial structure in tuning to the top three PCs. Maps for two representative subjects (S1 and S2) are

shown (see fig.2- S5 for maps from all subjects). Voxel-wise PC scores were calculated as the product of CSVA model feature weights for a given voxel by feature loadings for each PC. A RGB color space is used; red = scores on PC1, green = scores on PC2, blue = scores on PC3. PC scores are thresholded at 6 standard deviations above and below 0, with values beyond the threshold given the maximal (or minimal) color channel value. Areas where MRI data was not acquired are shown in black. Both voxels where the CSVA model did not fit significantly and those where the CSVA model fit significantly but did not outperform the semantic only model were excluded from the PCA (these voxels are shown in grey). PCA maps using CSVA model feature weights from all voxels where the CSVA model fit significantly are given in Figure 2-S6.

## OTC tuning to emotional natural images predicts behavioral responses.

A key question is whether OTC tuning to emotional natural stimuli might be able to guide selection of behavioral responses in a manner that goes beyond a simple approach - avoidance dichotomy. To address this, we asked a separate set of subjects recruited via Amazon's Mechanical Turk to select, from a fixed list, the behavioral responses appropriate to the content of each image viewed by our fMRI subjects (see Star Methods for details). The frequency with which each image was associated with each behavioral response was calculated (Fig. 2-6A). We examined the extent to which OTC tuning to emotional natural stimuli, as captured by CSVA model group PC scores, predicted these behavioral responses, across images. We varied the number of PCs used from 1 to 21. We capped the number of components extracted and used to predict behavior at 21 to facilitate comparison with the Semantic Only model (given this latter model's fewer features, this was the maximum number of extractable PCs). Each image was given a score for each PC that represented the product of its features by feature loadings for that PC. These scores were used to predict out of sample behavioral responses, across images (see Star Methods for details), Fig. 2-6B. The extent to which behavioral responses linked to each image were correctly predicted increased with the number of CSVA model PCs. The top 3 PCs alone accounted for 20% of explainable variance in behavioral responses across images, the top 10 PCs approximately 30%, and the top 20 roughly 40%. Notably, this performance was superior to that achieved using components from PCA conducted directly on CSVA images features, across stimuli, Fig. 2-6B. This indicates that OTC successfully extracts the information carried by CSVA model features that is pertinent to determining how to respond to a diverse range of emotional stimuli.

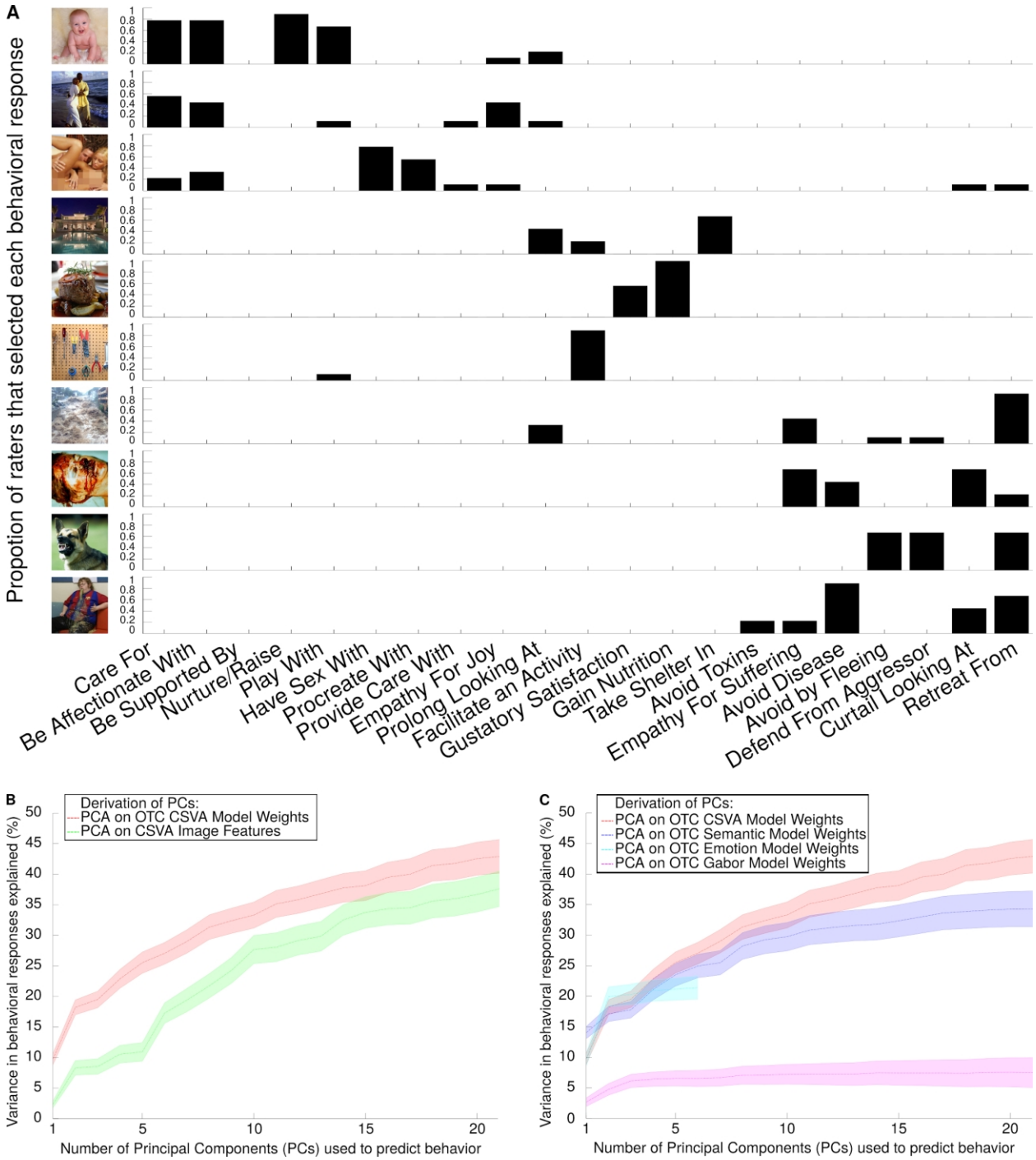


**Figure 2-6. OTC tuning patterns predict behavioral responses to emotional natural stimuli.**
(A) Subjects recruited through Amazon's Mechanical Turk platform were shown the emotional natural images used in this study and asked to pick one or more behavioral responses appropriate to the content of each image. For each image, we calculated the proportion of MTurk subjects that selected each behavioral response. This is illustrated here for 10 example images. The behaviors selected between are given on the x axis. Some terms have been abbreviated for illustration purposes, the full list is available in the Star Methods. In each row, the proportion of subjects selecting each behavior is plotted for a given image. (B) We examined the extent to which OTC tuning to emotional natural images, as captured by CSVA group-level PC scores (red line), predicted behavioral responses selected, across images. The percentage of out-of-sample variance in behavioral responses explained (y axis) is

plotted against the number of PCs included as predictors in an ordinary least squares regression analysis. The shaded area around the solid line represents the 95% confidence interval. We also calculated the variance in behavioral responses explained using PCs derived directly from PCA on CSVA image features, across images (green line). Across all levels of dimensionality considered (nu. of PCs=1 to 21), OTC tuning to CSVA features predicted behavioral responses significantly better than components from PCA conducted directly on the features themselves. This is consistent with OTC showing selective representation of image semantic and emotional features pertinent to behavior. (C). Here, we plot the variance in behavioral responses explained by PCs obtained from PCA on OTC feature weights for the CSVA model (red) versus (i) a low-level structural (Gabor) model (purple); (ii) the Semantic Only model (dark blue) and (iii) the Emotion Only model (light blue). Given the smaller feature space of the Emotion Only model, the maximum number of PCs that can be extracted for this model is six. The poor performance of Gabor model PCs in predicting behavioral responses suggests that OTC tuning to low-level image structural features is insufficient to guide behavior. Both the Semantic Only and Emotion Only models outperform the Gabor model in predicting behavior and perform particularly well at low levels of dimensionality. However, their maximal prediction of behavior (at n=21 and n=6 PCs respectively) is significantly less than that achieved by the CSVA model using an equivalent number of components.

We next sought to determine whether OTC tuning patterns captured by the CSVA model predicted behavioral responses to emotional natural images significantly better than OTC tuning to low-level image structure. To accomplish this, we fit a Gabor model to OTC BOLD time-courses (see Star Methods for details), conducted PCA on Gabor model weights across significantly predicted voxels and used the top 1-21 of these PCs to predict behavioral responses to each image. Figure 2-6C shows that OTC tuning to emotional images as captured by Gabor model PCs explained significantly less variance in behavioral responses than that captured by CSVA model PCs ($p < 0.05$). This indicates that OTC extraction of low-level structural image properties alone is unlikely to provide a basis for guiding behavior to emotional stimuli. Additional regression analyses revealed that the maximal performance of the Emotion Only model (at 6 PCs) and the Semantic Only model (at 21 PCs) was also significantly poorer than that of the CSVA model using an equivalent number of PCs, $ps < 0.05$, Fig. 2-6C. Here, as for the Gabor model, PCA was conducted on model weights across all OTC voxels where the model in question showed a significant fit.

## Discussion

It has been suggested that we are evolutionarily prepared to respond successfully to emotional stimuli, particularly those of animate origin. How we recognize and select responses to a wide range of emotional stimuli has remained unclear. Here, we suggest that an efficient mechanism for this might involve combined representation of stimulus emotional and semantic content. We investigated whether human OTC shows such combined representation and if this representation might suffice to guide behavioral responses to a wide range of emotional stimuli. Voxel-wise modeling of fMRI data collected while subjects viewed over 1600 emotional natural images revealed that many OTC voxels do indeed show combined representation of semantic and emotional image content, with this being most apparent for animate stimuli. A separate group of subjects selected behavioral responses that best matched image content. Regression analyses revealed that OTC tuning patterns, as captured by PCA on CSVA voxel-wise model weights, significantly predicted behavioral responses, across images. The amount of variance explained significantly exceeded that achieved when PCA was performed directly on stimulus

features, across images viewed. It also significantly exceeded that explained by OTC tuning to low-level image properties.

Within OTC, the presence of stretches of cortex with preferential tuning to different subtypes of animate and inanimate stimuli is now well established (Kanwisher, McDermott, and Chun, 1997; Gauthier et al., 2000; Downing et al., 2001; Epstein and Kanwisher, 1998). In contrast, work on the representation of emotional features of visual stimuli in OTC has been limited. Our findings suggest that there is integration of semantic and emotional information within OTC, with coding of both stimulus valence and arousal for animate stimuli and some more limited coding of stimulus arousal for inanimate stimuli. Voxels sensitive to this information are found in stretches of cortex adjacent to areas with well-characterized semantic selectivity, notably neighboring FFA, EBA and pSTS. Within these well characterized areas, semantic selectivity might have developed to facilitate navigation of our social environments. Our current results are consistent with there potentially being further differentiation of this tuning within adjacent stretches of cortex – with selective tuning to semantic features in these adjacent regions being combined with tuning to stimulus emotional attributes.

Our results further suggest that OTC co-representation of image semantic and emotional content has the potential to guide selection of diverse behavioral responses to a wide range of emotional stimuli. Indeed, across the different levels of dimensionality considered (1-21 PCs), OTC tuning to semantic and emotional image features better predicted behavioral responses to image content than direct use of an equivalent number of dimensions based directly on the stimulus feature themselves. This potentially indicates that OTC is effectively compressing and extracting stimulus semantic and emotional content of pertinence to behavior.

When a model of low-level visual structure (Gabor wavelet model) was fitted to OTC voxel time-courses, it did a significantly poorer job than the CSVA model in predicting behavioral responses. This further indicates that it is the encoding of higher-level features in this region of the brain, as opposed to lower-level structural features, that has the potential to be used to trigger appropriate behaviors. As information reaches OTC rapidly (Liu, 2009), representation of stimulus semantic and emotional content within this part of cortex might be especially well suited to drive rapid behavioral responses to emotional stimuli. We note that we are not arguing that the amygdala is not involved in the initial learning of emotional value, or indeed in signaling the presence of threat. Similarly, we are also not arguing against the presence of co-representation of semantic and emotional stimulus content within frontal cortex. We suggest, instead, that the OTC might be particularly well suited to the long-term storage and rapid activation of the representations needed to select between alternate emotional behaviors.

In summary, our findings reveal that stimulus semantic and emotional content are co-represented in OTC. This representation is seen in areas of cortex adjacent to, and partially overlapping with, areas with known semantic selectivity. This might reflect further differentiation of tuning patterns to additionally represent emotional value, allowing for more nuanced representation of our complex social environments. Our analyses further indicate that this co-representation of semantic and emotional image content within OTC carries information sufficient to drive rich behavioral responses to a diverse range of emotional stimuli. Indeed, the superior prediction of behavioral responses from OTC tuning to stimulus features than from stimulus features themselves possibly indicates that OTC is effectively compressing and extracting aspects of semantic and emotional image content that are of value to guiding behavior.

# Supplemental Methods

## Subject Details

Data were collected from 6 healthy adult human subjects (4 females, 2 males, mean age = 24, range = 21-26). All subjects had normal or corrected-to-normal vision. The study was approved by the University of California Berkeley committee for protection of human subjects. Informed consent was obtained from all subjects prior to participation.

## Experimental Stimuli

A total of 1620 stimuli were used in total, 1440 during model estimation runs and 180 during model validation runs. All stimuli were natural images selected from a combination of the International Affective Picture System (IAPS) set (Lang, Bradley and Cuthbert, 2008), the Lotus Hill image set (Yao, Yang and Wu, 2009), and internet searches. The images were labeled using 23 mutually exclusive semantic categories: faces, full bodies, body parts, couples, gatherings (2-9 people), crowds (10+ people), land mammals, water mammals, fish, reptiles, insects, birds, savory food, dessert food, household items, vehicles, other artifacts, indoor buildings, outdoor buildings, land scenery, sky scenery, water scenery, plants. These categories were based on those used by Naselaris and colleagues (Naselaris et al., 2009), with a number of modifications (e.g. differentiation between reptiles and insects, elimination of irrelevant categories (texture patterns)). Four raters independently categorized each image; their modal categorization was used in the models described below. In just one instance no modal category was obtained. This was resolved by discussion.

The small number of images labeled as plants (n = 36) or cars (n = 72) led to the possibility of empty feature vectors when stimuli were labeled using composite features for semantic category by valence by arousal (see Model Features below). Hence, for all except the first two subjects to complete the first estimation and validation fMRI session (subjects 1 and 3), these images were removed and replaced with images from the building category. Images were displayed at a visual angle of 12 x 12 degrees.

## Procedure

Subjects completed six fMRI sessions. Four 9.5 minute retinotopy scans (two with clockwise/counterclockwise wedges and two with expanding/contracting rings; Huth et al., 2012) were completed within session 1. Each of the subsequent five sessions comprised performance of the main task. In each of these sessions, participants completed six model estimation scans of 7.5 min duration and four model validation scans of 6 min duration. A structural scan was also acquired at the beginning of each session.

All stimuli were back-projected onto a translucent screen positioned in the bore of the magnet, visible via an angled mirror placed above the subject's head. Subjects fixated on a central white cross. Images were presented for 1 s with a 3 s inter-stimulus-interval during which the fixation cross was presented against a grey background with luminance matched to the mean luminance of the images in the stimulus set. During each estimation scan, forty-eight images were presented twice in a pseudo-random order. Null trials (no image presented) were also included, occurring once every eight trials. During validation scans, nine images were each presented nine times using a Type-1, Index-1 sequence (Aguirre, 2007) to control for order

effects. While viewing the images, subjects performed one of two tasks. Four subjects (1, 3, 5, and 6) were asked to categorize the valence of each image as negative, neutral, or positive. To control for effects of task, two subjects (2, 4) performed an alternate semantic categorization task, categorizing each image as human, animal, object, food, or building/scene. For these two subjects, valence categorizations were obtained from the post-scan behavioral sessions (see below).

## Post-Scan Behavioral Task

After the six fMRI sessions were completed, each subject returned to our lab to complete an additional behavioral task. Each image viewed within the fMRI sessions was re-presented to subjects with the same grey background and degree of visual angle as used in the fMRI sessions. Subjects were instructed to first rate the valence of each image as negative, neutral, or positive. Labels above the image indicated which button to press for each valence. Following valence categorization, the text above the image changed to indicate that image content should be rated for arousal using a nine-point scale. At the beginning of the task, subjects were instructed that 'positive' referred to "something you would want to look at, have, or be close to", while 'negative' referred to "something you wouldn't want to look at, have or be close to". Arousal was categorized on a continuum from not emotionally intense at all (1) to extremely emotionally intense (9). The task was self-paced and took between 3 and 6 sessions to complete, with each session lasting 1.5 hours.

## fMRI Data Acquisition

FMRI data were collected on a 3T Siemens TIM Trio scanner at the UC Berkeley Brain Imaging Center using a 32-channel head coil. An echo-planar T2*-weighted imaging (EPI) sequence was used with a decreasing slice series, repetition time (TR) = 2.0s, echo time (TE) = 34 ms, flip angle = 74, voxel size = 2.4 x 2.4 x 3.0 mm, inter-slice gap = 0.75mm, matrix size = 98 x 98, field of view = 224 x 224 mm. We prescribed 25 axial slices to cover all of temporal and occipital cortices, and as much of frontal and parietal cortices as possible. The first 5 volumes of each scan were discarded to allow for T1 equilibration effects. Anatomical data were collected using a T1-weighted MP-RAGE sequence with 1mm isotropic resolution. A separate T1 was acquired at the beginning of each scan session. Peripheral recordings: during fMRI data collection, pulse oximetry and respiration data were collected using a Biopac recording system (Biopac MP150 Data Acquisition Unit, Biopac UIM100C with Nonin 8600FO for pulse oximetry, and Biopac RSP100C with Biopac TSD221 for respiration). Eye tracking was performed using an Avotec camera and Arrington Viewpoint software.

## fMRI Data Preprocessing

FMRI data were preprocessed using Matlab version 8.0 (The MathWorks, Natick, MA) and SPM8 (Welcome Department of Imaging Neuroscience, London, UK). Blood oxygen level dependent (BOLD) images were first converted from DICOM to NIFTI format. Next, diagnostics were run on the BOLD time series from each scan. Following an approach similar to that adopted by Power and colleagues (Power at al., 2012), and Carp (Carp, 2013), bad volumes (with unusually high changes in mean whole-brain signal intensity) were identified using the SPM time-series diagnostic tool tsdiffana.m (http://imaging.mrc-

cbu.cam.ac.uk/imaging/DataDiagnostics). Among other indices, this tool calculates the mean square difference of voxel-wise signal intensities between each volume (n) and the previous volume (n-1) and divides this by the mean signal across the whole volume averaged over the whole time-series. Volumes (both n and n-1) were rejected using an absolute cutoff (the recommended default of 10) as this handles differences between subjects in the noisiness of data better than a within-subject percentile cut off. In line with findings by Power and colleagues (Power et al., 2012), bad volumes tended to correspond to those with notable spikes in movement. Bad volumes were replaced by the average of the volumes on either side. Subsequent to this initial data-cleaning step, image realignment was conducted to correct for within run head movement and to align images between runs. This was then followed by slice time correction. An 'image on/off' nuisance variable accounting for variance due simply to image presence was regressed out of the preprocessed BOLD data. This was coded with a 1 for volumes with a stimulus present and 0 otherwise and convolved with a 4 bin FIR function (FIR time bins: 2-4s, 4-6s, 6-8s and 8-10 s post stimulus onset). Low-frequency voxel response drift was identified using a Savitzky-Golay filter with a 120 s window and 3rd degree polynomials and this was subtracted from the signal. The mean response for each voxel within each run was then subtracted from the BOLD data to account for fluctuations in BOLD signal values across runs. Data were neither normalized to common space nor spatially smoothed in order to retain maximal resolution for our voxel-wise modeling.

## fMRI Data Modeling

**Models Constructed**

Each image viewed was labeled with a distinct set of features corresponding to the feature space for each of the following models:

(1) The 'Semantic Only' model. The feature space for this model comprised the following mutually exclusive semantic categories: Human-Face, Human-Body, Human-Body-Part, Human-Couple, Human-Gathering, Human-Crowd, Land-Mammal, Water-Mammal, Bird, Fish, Reptile, Insect, Savory-Food, Dessert-Food, Household-Object, Other-Artifact, Indoor-Building, Outdoor-Building, Water-Scenery, Sky-Scenery, Land-Scenery. Two subjects (1 and 3) also viewed a small number of images belonging to two additional categories: Plants and Cars (see the Stimuli, Task & Procedure section). Each image was given a '1' for the semantic category to which it was most often allocated, across raters, and a '0' for all other categories.

(2) The 'Combined Semantic, Valence and Arousal' (CSVA) model. Ratings from each subject were used to label each image for valence (negative, neutral or positive) and arousal (high or low). For subjects who did the valence categorization task in the scanner, each image was presented either twice (for images shown in model estimation runs) or nine times (for images shown in model validation runs). In both cases, the modal valence value was used. For cases in which there was no modal value (e.g. for images categorized once as negative and once as neutral), if one of the modes was neutral, the image was coded as neutral. All other images without modal responses (e.g. those with one negative and one positive categorization) were excluded from analysis (mean = 24 images excluded per subject). For the two subjects who did the semantic task within the scanner, post-scan valence categorizations were used. Image arousal level (high, low) was determined using a within-subject median split on the post-scan 9-point

ratings of image arousal; images rated below the median were categorized as low-arousal, and those equal to or above the median were categorized as high-arousal. These labels were used to subdivide each semantic category by valence and by arousal. In this manner, new composite features were created, each of which had a binary (1/0) value for semantic category, valence, and arousal (e.g. Human-Body, neutral valence, low arousal = '1'). In addition to these semantic by valence by arousal features, 18 additional binary (present ='1', absent = '0') features that carried information about both semantic and emotional image content were also included in the CSVA model. These comprised: Mutilated humans, Mutilated animals, Rotten food, Threat directed towards the viewer by a human aggressor, Threat-directed towards the viewer by an animal aggressor, Threat directed away from the viewer by a human aggressor, Threat directed away from the viewer by an animal aggressor, Romantic couples portrait (i.e. face only), Romantic couples full bodies, Human babies, Animal babies, Human social interaction portrait, Human social interaction single human (e.g. person playing golf), Human social interaction couples, Human social interaction gatherings, Erotica portrait, Erotica single human, and Erotica couples.

(3) The 'Semantic with Valence by Arousal for Animate Stimuli' model. This model included composite semantic by valence by arousal features (as described in the CSVA model above) for images belonging to animate semantic categories. Only semantic features were included for images belonging to inanimate semantic categories (as in the Semantic Model). In other words, information about image valence or arousal was only included for stimuli belonging to animate semantic categories.

(4) The 'Semantic with Valence by Arousal for Inanimate Stimuli' model This model included composite semantic by valence by arousal features (as described in the CSVA model above) for images belonging to inanimate semantic categories. Only semantic features were included for images belonging to animate semantic categories (as in the Semantic Model). In other words, information about image valence or arousal was only included for stimuli belonging to inanimate semantic categories.

(5) The 'Emotion Only' model. This model included six composite features indicating the valence (negative, positive, neutral) and arousal (high, low) values for each image. The semantic category of each image was not included in the model. This resulted in each image being assigned a single composite feature that was a combination of subject defined valence and arousal (e.g. negative, low-arousal).

(6) The 'Gabor' Model. This model was used to assess the representation of low-level image structure, specifically variation in luminance contrast across the image. Model features comprised the results of filtering each image, after it was grayscaled and zero-meaned, with a set of 474 Gabor filters that spanned 4 orientations (0, 45, 90, and 135 degrees) and 5 spatial frequencies (1.5, 3, 6,12 and 24 cycles/image) across a square grid of spatial locations covering the image (500 x 500 px). The filters were spaced using a grid determined separately for filters at each spatial frequency such that adjacent Gabor filters were separated by 3 standard deviations of the spatial Gaussian envelope. Feature weights were z-scored across all images to normalize differences in energy magnitude due to Gabor filter size. Image processing was conducted using the STRFlab toolbox for Matlab (strflab.berkeley.edu).

**Model Estimation**

Model estimation was performed using custom software written in Matlab. For each subject, a design matrix was created for each model with regressors that indicated the presence (1) or absence (0) of each of the model's features for each stimulus. Each of these feature regressors was convolved with a finite impulse response (FIR) filter, resulting in 4 new regressors for each feature, each one representing a time delay of 2-4, 4-6, 6-8 and 8-10 seconds from stimulus onset, respectively. Taking the dot product of these regressors with a set of linear weights is functionally equivalent to convolution of the original feature with a linear temporal kernel that has nonzero entries for 2-4, 4-6, 6-8, and 8-10 second delays. Six movement parameters, as estimated by SPM 8, were also included in the model as nuisance regressors.

For each subject, fMRI data from the model estimation runs were concatenated. L2-penalized (ridge) linear least square regression was used to find feature regressor weights which mapped the model features onto the BOLD time-series for each voxel. L2-penalized regression requires specification of a hyper-parameter lambda, which determines the amount of penalization applied during feature weight estimation (i.e. how much the feature weights are shrunk towards a Gaussian distribution). A range of 10 lambda values logarithmically scaled from 10-9 to 105 were tested. K-folds cross-validation was used to determine the optimal value of lambda. Specifically, for each value of lambda, each model was fit on 9/10ths of the estimation data by selecting 27 of the 30 estimation runs without replacement. Using the weights estimated, voxel-wise BOLD time-series were predicted for the remaining 1/10th of the data. This was repeated until all runs had been included once in the held-out data. Concatenating the 10 predicted sections of the data resulted in a predicted time-series for the entire estimation dataset for each voxel. This complete predicted time-series was correlated with the actual recorded BOLD time-series and the single lambda value which produced the highest mean correlation value across all voxels was selected. Although selecting lambda individually for each voxel would almost certainly result in higher model performance, we opted to use a single value across all voxels in order to keep feature weights on the same scale and allow for subsequent principal components analysis of feature weights across voxels (see below). The final feature weights used were created by re-estimating the weights across all the estimation data using the selected best lambda value. This fitting procedure was repeated for each model and each subject.

**Model Validation**

Model validation was performed using custom software written in Matlab. For each subject, the voxel-wise prediction accuracy of each model was assessed using the concatenated BOLD time-series from the validation runs. Completely new images were viewed within these runs. As in model estimation, a design matrix was constructed for each model by creating regressors to indicate the presence or absence of each model feature for each of the images viewed during the validation scans. This design matrix was then convolved with four FIR bins (2-4s, 4-6, 6-8s and 8-10s post stimulus onset). The feature weights obtained during model estimation were multiplied with the FIR-convolved validation design matrix to produce voxel-wise predicted BOLD time-series for the 6760 seconds of validation data. We correlated these predicted time-series with the observed validation BOLD time-series to obtain estimates of model prediction accuracy for each voxel, which provide a metric of model fit that controls for over-fitting. Permutation testing was used to determine the significance of this correlation for each voxel for each model. Please see the Quantification and Statistical Analysis section for more details.

We note that correlation values reported here are not scaled by each voxel's noise ceiling, and hence are lower than those reported in papers where such scaling is used. Noise ceiling calculations require exact repetition of validation stimuli in the same order. This has the advantage of allowing a measurement of the amount of explainable variance, assuming that responses to the same stimuli in the same order are constant. Given potential issues of habituation across multiple presentations for emotional stimuli, we chose to minimize order effects by presenting images in a pseudo-randomized Type-1, Index-1 order (Aguirre, 2007). A consequence of this is that we separately predict voxel-wise responses to each individual presentation of a given stimulus. Inevitably this is noisier than predicting averaged responses across several stimulus presentations leading to lower raw prediction values but similar power to detect prediction significance given the increased number of data points available without averaging.

**Model Comparison**

We compared the voxel-wise prediction performance of the CSVA model to that of the Semantic only and Emotion Only models. Voxels whose activity were significantly predicted by any one of the three models were included in these comparisons. Correlations between predicted validation BOLD time-series and actual validation BOLD time-series were computed for each voxel for each model as described above. We then calculated the proportion of these correlations that were greater for the CSVA model compared with the Semantic Only and Emotion Only models. We used a bootstrap test to determine whether this proportion was significant. Please see the Quantification and Statistical Analysis section for more details. The same method, and cortical mask was used to compare the Semantic Only model against the Semantic with Valence by Arousal for Animate Stimuli (SVAA) and the Semantic with Valence by Arousal for Inanimate Stimuli (SVAI) models.

## PCA of CSVA model feature weights

Principal components analysis (PCA) of CSVA model feature weights, across voxels, allows us to identify consistent patterns of co-tuning to image features, i.e. to identify image features to which voxels tend to show similar responses. We conducted a PCA of CSVA model feature weights across all voxels within OTC where the CSVA model fit significantly (as assessed by significant prediction of the BOLD time-series for the validation dataset) and outperformed the Semantic Only model. The weights from the 4-6 s and 6-8 s FIR bins were averaged together for each feature, as these time points correspond to the peak of the BOLD hemodynamic response function (HRF). Non-centered PCA was applied to these weights. The feature weights associated with vehicle and plant semantic categories were excluded as only 2 of the 6 subjects viewed stimuli from these categories. Both group-level and subject-wise PCA analyses were conducted.

For the group-level PCA, voxels were concatenated across subjects. Some of the structure in feature tuning across voxels may merely reflect co-variance between stimulus features. Hence, we sought to identify the top group PCs for which the amount of variance explained was more than achieved by consideration of stimulus features alone. To address this, we used a jackknifing procedure. Please see the Quantification and Statistical Analysis section for more details. This resulted in the retention of the top three PCs.

We next calculated the similarity between the top three group PCs and the top three PCs from each subject-wise PCA. This was achieved by correlating the group and single subject PC vectors of feature loadings (fig. S4B). Since the ordering of the PCs is by explained variance, the ordering of components at the group level may not always be the same as that at the single subject level. To ensure we were comparing single subject PCs to group PCs that were capturing similar dimensions, the top three single-subject PCs were re-ordered so that each single-subject PC was matched to the group PC with which it had the highest correlation. If there were conflicts (where the highest correlation for 2 or more single subject PCs was with the same group PC), we resolved them by selecting the single subject PC that was closest to the group PC in its ordering. This was done recursively where necessary.

**Interpreting the top 3 group-level PCs**

By correlating theoretically informative dimensions with each of the group PCs it is possible to investigate the aspects of stimulus content encoded by each PC. Each theoretical dimension of interest was formalized as a vector comprised of values for each of the CSVA model features. We used three theoretical dimensions to explore the representation of animacy. The first of these comprised a four-level scale of animacy (Connolly et al., 2012), with inanimate objects at the bottom of the 0-3 point scale, followed by invertebrates and non-mammalian vertebrates, then non-human mammals, with humans at the top of the scale. We also used a simpler binary animacy dimension where features indicating that the stimulus was animate were given a '1' and features indicating that the stimulus was inanimate were given a '0'. The third dimension specifically coded for presence (1) or absence (0) of humans. The remaining dimensions encoded the perceived valence (positive =+1; negative=-1, neutral =0) and arousal (high = 1; low =0) of stimuli from either animate or inanimate semantic categories. As an example, the fourth dimension encoded the perceived valence of all animate stimuli. Here, all CSVA composite features indicating an image was perceived as positively valenced and belonged to one of the 12 animate semantic categories were given a '1', all CSVA composite features indicating an image was perceived as negatively valenced and belonged to one of the 12 animate semantic categories were given a '-1', and all CSVA composite features indicating an image was perceived as neutrally valenced and belonged to one of the 12 animate semantic categories were given a '0'. To determine the valence for each of the 18 semantic-emotion features in the CSVA model, we took the mode of the valence categorizations for stimuli possessing that feature across subjects. Finally, inanimate stimuli were given the value that equated to the mean of animate stimulus values, in order to exclude their influence on the correlation with the PCs. In a parallel fashion, we also created dimensions that coded inanimate stimulus valence, animate stimulus arousal and inanimate stimulus arousal.

The vector for each theoretical dimension of interest was correlated with the CSVA feature loadings for each of the PCs from the group-level PCA on CSVA model feature weights, across OTC voxels. We used a bootstrap procedure to assess the significance of these correlations. Please see the Quantification and Statistical Analysis section for more details.

## Prediction of behavioral responses

We next sought to test whether patterns of tuning to semantic and emotional content in OTC could predict behavioral responses to the emotional stimuli. To investigate this, Amazon Mechanical Turk (AMT) workers were presented with 25 potential behavioral responses and

asked to select the response(s) that would be appropriate to take if confronted with the content of a given image from our stimulus set. The 25 behavioral responses included were informed by recent theoretical work (LeDoux, 2012) and extended to fit our stimulus set. They were as follows: Bond with, Take Care of Young with, Nurture/Raise, Procreate with, Be Affectionate with, Have Sex with, Play with, Gain Social Support from, Gain Nutrition from, Gain Gustatory Satisfaction from, Relieve Thirst with, Use to Facilitate Activity, Take Shelter in, Warm Oneself with, Empathic Response to Suffering, Empathic Response to Joy, Prolong Looking at, Curtail Looking at, Retreat from, Defend Self from Aggressor, Defend Self from Unwanted Sexual Attention, Avoid Predation by Fleeing, Avoid Predation by Freezing, Avoid Disease by Not Touching, Avoid Toxins by Not Consuming. In total 49 workers allocated a total of 27516 behavioral responses with a mean of 2 behavioral responses per worker per image, with 9 workers evaluating the behavioral responses suited to the content of each image. Four of these behavioral responses (Relieve Thirst with, Warm Oneself with, Defend Self from Unwanted Sexual Attention, Avoid Predation by Freezing) were rarely assigned to our images (less than 1% of total responses) and hence excluded from further analyses.

In order to compare models' prediction of behavioral responses, we conducted additional PCAs on voxel feature weights for the Gabor model, the Semantic Only model and the Emotion Only model, respectively. We included all OTC voxels where the model in question significantly fit the validation BOLD time-series. For both the CSVA model and these additional comparison models, we projected the vector of feature values for each image into the PCA space of each model. This was achieved by calculating the inner product of the images' feature vector with each of the PC loading vectors. For models incorporating emotion features (the CSVA and Emotion only models), we used the modal values of valence and arousal, across subjects, for each image. For the Emotion Only model, this procedure resulted in 6 PC scores per image, for the Semantic Only model, this procedure resulted in 21 PC scores per image. For the remaining models, we retained the top 21 PC scores per image to match the number of PCs in the Semantic Only model.

In addition, we conducted PCA on CSVA model image features, across images, concatenating design matrices across subjects. We also projected the vector of feature values for each image into this PCA space, retaining the top 21 PC scores per image.

The final step entailed using these PC scores to predict the behavioral responses allocated to each image. This effectively allows us to determine how well the tuning captured by each model predicts behavioral responses to each image. We conducted a series of regression analyses using PCs from each of the PCAs described above. We varied the number of PCs entered as regressors; this number was increased in steps of 1 from 1 to n (n = 21 except for the Emotion Only model, where n=6), see Fig. 2-6B. In order to control for over-fitting we used leave-one-out cross-validation (LOOCV) to calculate the amount of variance explained in the behavioral responses (R2). Please see the Quantification and Statistical Analysis section for more details.

## Supplementary PCA across all cortical voxels

The PCA on OTC CSVA model feature weights, described above, was conducted across voxels limited to OTC where the CSVA model showed a significant fit and performed better than the Semantic Only model. We anticipated this would best capture the structure of tuning to emotional natural stimuli within OTC. To compare the structure of tuning to emotional stimuli captured in this manner to that achieved using all cortical voxels whose time-courses were

significantly fit by the CSVA model, we repeated the PCA described above relaxing the selection criteria – simply including all voxels whose activity was significantly predicted by the CSVA model. Figure 2-S5A shows the amount of variance explained by the top 10 group-level PCs of this expanded-voxel-selection PCA, and figure 2-S5B shows the correlation between the top three PCs from this analysis and those from the OTC-restricted PCA reported in the main text. Figure 2-S5C shows the projection of PC scores from this supplementary PCA onto cortical flatmaps.

## Visualization of fMRI results

**Flatmap Construction and ROI labeling**

Cortical flatmap construction was conducted using PyCortex (Gao et al., 2015). This python tool makes use of the Freesurfer image analysis suite for cortical reconstruction and volumetric segmentation (http://surfer.nmr.mgh.harvard.edu/). Following initial automatic segmentation, white matter and pial surface maps were hand edited to remove any remaining artifacts and the surface was regenerated. In order to flatten the cortical surface, five relaxation cuts were made into the surface of each hemisphere and the surface crossing the corpus callosum was removed. The calcarine sulcus cut was made at the horizontal meridian in V1 using retinotopic mapping data from Session 1 as a guide.

Early visual regions V1-V4 were defined using the retinotopic mapping data. In addition, a simple semantic model with 8 categories (Faces, Bodies, Body Parts, Multiple People, Animals, Food, Objects, Scenes) was fit to the estimation data and used to identify the following functional landmarks on each subject's flat map: RSC, Retrosplenial Complex; OPA, Occipital Place area; LO, Lateral Occipital cortex; pSTS, Posterior Superior Temporal Sulcus; EBA, Extrastriate Body Area; OFA, Occipital Face Area; FFA, Fusiform Face Area; PPA, Parahippocampal Place Area; ATFP, Anterior Temporal Face Patch. We also label the following sulci: IPS, Intraparietal Sulcus; STS, Superior Temporal Sulcus; ITS – Inferior Temporal Sulcus; CoS, Collateral Sulcus; Post-CS, Post Central Sulcus; CS, Central Sulcus; SF, Sylvian Fissure. Note, these ROIs are only for orientation of the viewer and were not used to constrain any of our analyses.

**Display of PC scores on the cortical maps**

A RGB color key was used to project voxel-wise PC scores for the top three group PCs onto individual subjects' cortical maps (red = loading on PC1, green = loading on PC2, blue = loading on PC3). Voxel-wise PC scores were calculated as the product of CSVA model feature weights for a given voxel by feature loadings for each PC. PC scores were thresholded at 6 standard deviations above and below 0, with values beyond the threshold given the maximal (or minimal) value. Thus, a value of 0 for a given color channel is 6 s.d.'s below a PC score of 0, a value of 128 for a given color channel has a PC score of 0, and a value of 255 is 6 s.d.'s above a PC score of 0.

## Quantification and Statistical Analysis

**Model prediction significance testing**

We calculated voxel-wise prediction accuracies for each model for each subject. As described further in the Methods section above, the feature weights obtained during model estimation were multiplied with the FIR-convolved validation design matrix to produce voxel-

wise predicted BOLD time-series for the validation data-set. We calculated voxel-wise prediction accuracy scores by correlating these predicted time-series with the observed validation BOLD time-series. Permutation testing was used to determine the significance of model prediction for each voxel. The following procedure was used. We randomly shuffled validation data-set images without replacement, and convolved the resulting feature regressors with the FIR filters to create a randomized FIR-convolved validation design matrix. This randomized design matrix was then multiplied by the feature weights for each voxel from model estimation. This created a predicted validation BOLD time-series for each voxel. This predicted time-series was correlated with the observed validation BOLD time-series for the same voxel. This was repeated 5000 times for each voxel, producing a null distribution of prediction accuracy scores (i.e. correlations) for each voxel. Significance was determined on a voxel by voxel basis by evaluating the proportion of values in this null distribution that fell below the actual prediction accuracy score for that voxel for the model in question. We corrected for multiple comparisons by using the Benjamini-Hochberg procedure to apply a false discovery rate (FDR) correction across all cortical voxels within the subjects' flat-map ($q < 0.05$).

**Model Comparison**

In order to compare model fit between two models a bootstrap procedure was used to estimate a distribution of prediction accuracy scores for each voxel for each model by resampling across the 20 validation runs with replacement, 1000 times. Pairs of models were compared by examining the relative proportion of voxels better predicted by one model than the other, for each of these 1000 iterations. We calculated a z-value from the mean and standard error across these 1000 ratio values, after subtracting 0.5 (chance level, both models perform equally). A p-value indicating the extent to which one model was superior to the other was then computed from the resulting z-value.

**Principal component significance testing**

We conducted group-level and subject-wise PCAs of CSVA model feature weights, across voxels (see Methods for voxel selection, time-bin selection, and other details.) For the group-level PCA, voxels were concatenated across subjects. In order to determine which group PCs explained significantly more variance than stimulus features alone we conducted a jackknife test. Thirty sets of voxel-wise feature weights were created by holding out each estimation run in turn and estimating the feature weights for each voxel using the remaining twenty-nine runs. PCA was conducted on each of the resultant thirty sets of feature weights. A standard jackknife test was used to determine whether the amount of variance explained by the top PC from each of these jack-knifed PCAs was significantly greater than that explained by the top stimulus feature PC (i.e. the first component from a PCA conducted on the presence or absence of stimulus features across images). This procedure was repeated for subsequent PCs until the difference in variance explained by the PC from the group PCA of CSVA model feature weights, across voxels, and the corresponding stimulus PC was no longer significant. This procedure was used to determine the group PCs retained for further consideration.

**Relating PCs to hypothetical dimensions**

Three PCs were retained from the group PCA of CSVA model feature weights (see section above). We investigated the aspects of stimulus content encoded by each of these PCs by correlating each of them with hypothetical dimensions of interest. The hypothetical dimensions

were formalized as vectors comprised of values for each of the CSVA model features (see Methods for further details). To assess the significance of correlations between the PCs and the hypothetical dimensions, we created a bootstrapped distribution for each correlation using randomized sampling with replacement. Specifically, we resampled CSVA model features with replacement 5000 times and re-estimated the correlations between the group PCs and hypothetical dimensions across the sampled features. In this manner, we obtained a distribution of 5000 values for each correlation. This distribution was then used to conduct a bootstrap test for significance and to obtain the 95% confidence interval. An alpha threshold of 0.05 was used to determine significance.

**Predicting behavioral responses from OTC tuning**

We investigated how well OTC tuning, as captured by the CSVA model, could predict behavioral response to the images viewed. We compared this prediction performance with that of three alternative models of OTC tuning and with that achieved using image features (as labeled by the CSVA model) as opposed to brain responses to these features. In each case, feature dimensionality was reduced using PCA.

More specifically, PCA was conducted on feature weights for each of the four models considered (see Methods), across OTC voxels, and the inner product of each images' feature vector with each of the PC loading vectors calculated. This gave a score for each image on each PC for each model. PC scores for each image were also obtained using PCA on the CSVA model image features themselves.

These PC scores were entered into regression analyses to predict behavioral responses for each image (see Methods). We used leave-one-out cross-validation (LOOCV) to calculate the amount of variance in behavioral responses explained. This form of cross-validation controls for over-fitting of the behavioral responses. For each model, and each behavioral response, each of the images was left out one at a time. The PC scores for the other images were then used to predict the proportion of raters that allocated the behavioral response to the left out image. This was conducted in turn for all of the images. The amount of variance explained was summed across iterations, giving the total amount of variance in workers' selection of this one behavior that could be explained by a given set of PC scores. This was repeated across behaviors. This LOOCV R2 value was then scaled by the total explainable variance in behavioral mappings (this reflecting the consistency in behaviors selected for each image across MTurk workers.) Bootstrapping across images was used to determine confidence intervals for the scaled LOOCV R2. Specifically, across 1000 iterations, images were randomly sampled with replacement, and the above procedure repeated. The resultant 1000 scaled LOOCV R2 values were used to calculate confidence intervals.

**Transforming prediction accuracies to z-values**

We transformed model prediction accuracies to z-values for display on each subject's cortical map (Fig. 2-2 and figs. 2-S1 – 2-S3). To do so, we first converted the correlation of the predicted and actual validation BOLD time-series to a t-statistic using the following equation:

$$t = r*sqrt((n-2)/1-r^2)$$

where r is the Pearson correlation coefficient and n is the number of validation volumes (3380). We then converted that t-statistic to a z-value by first finding the probability of the t-

value using the student's-t cumulative distribution function, and then using the normal probability density function to find the z-value associated with that probability value.

# Supplemental Figures

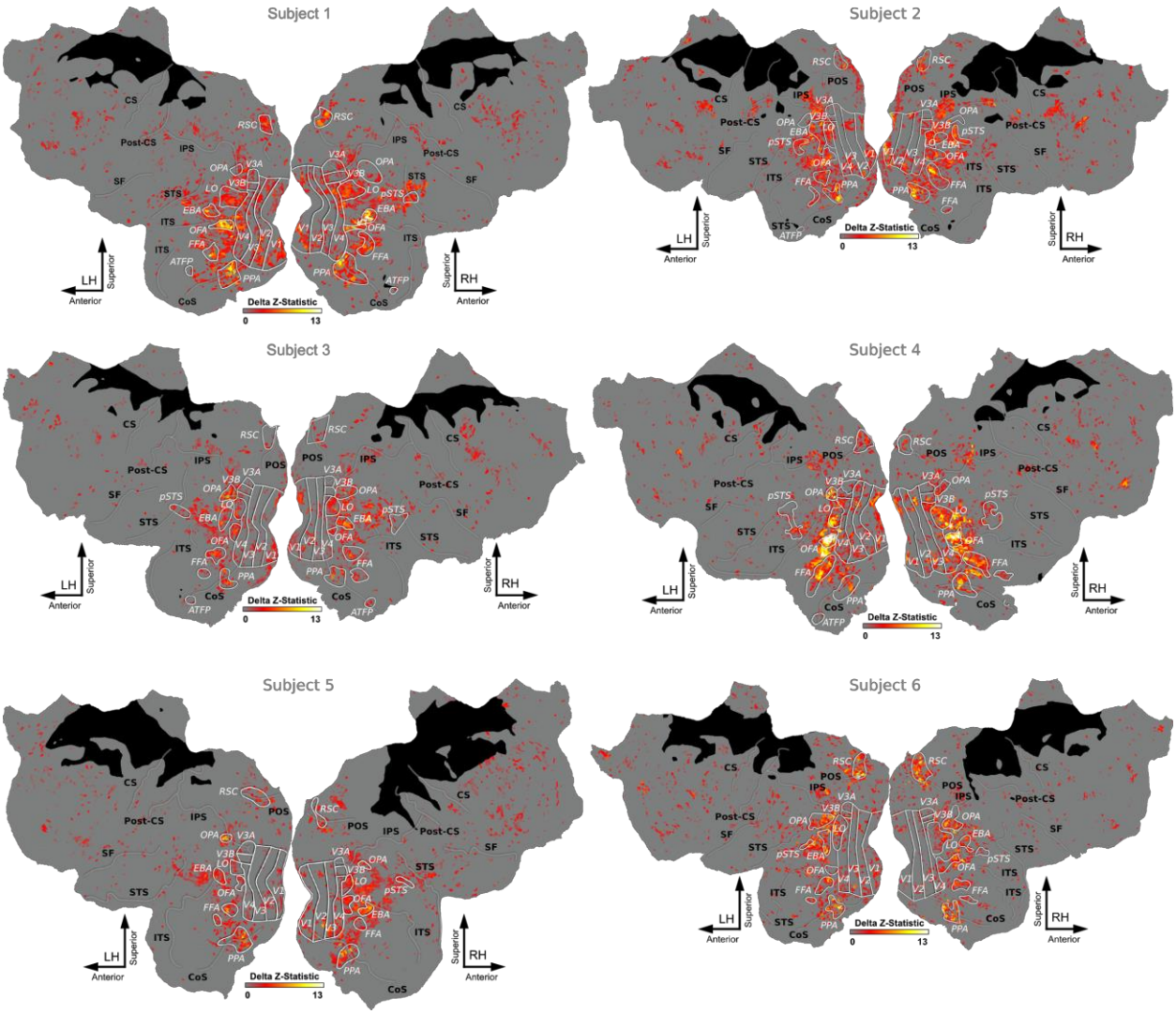


**Figure 2-S1. Improvement in voxel-wise prediction accuracies for the CSVA model relative to the Emotion Only Model.**

Cortical maps for each subject show voxels where prediction accuracy was greater for the CSVA model than the Emotion Only model. Voxel-wise prediction accuracy values for each model were z-transformed and subtracted (see the Quantification and Statistical Analysis section of the Star Methods for full details). Only voxels whose activity was significantly predicted by any one of the following three models were included in these comparisons: CSVA, Emotion Only and Semantic Only.

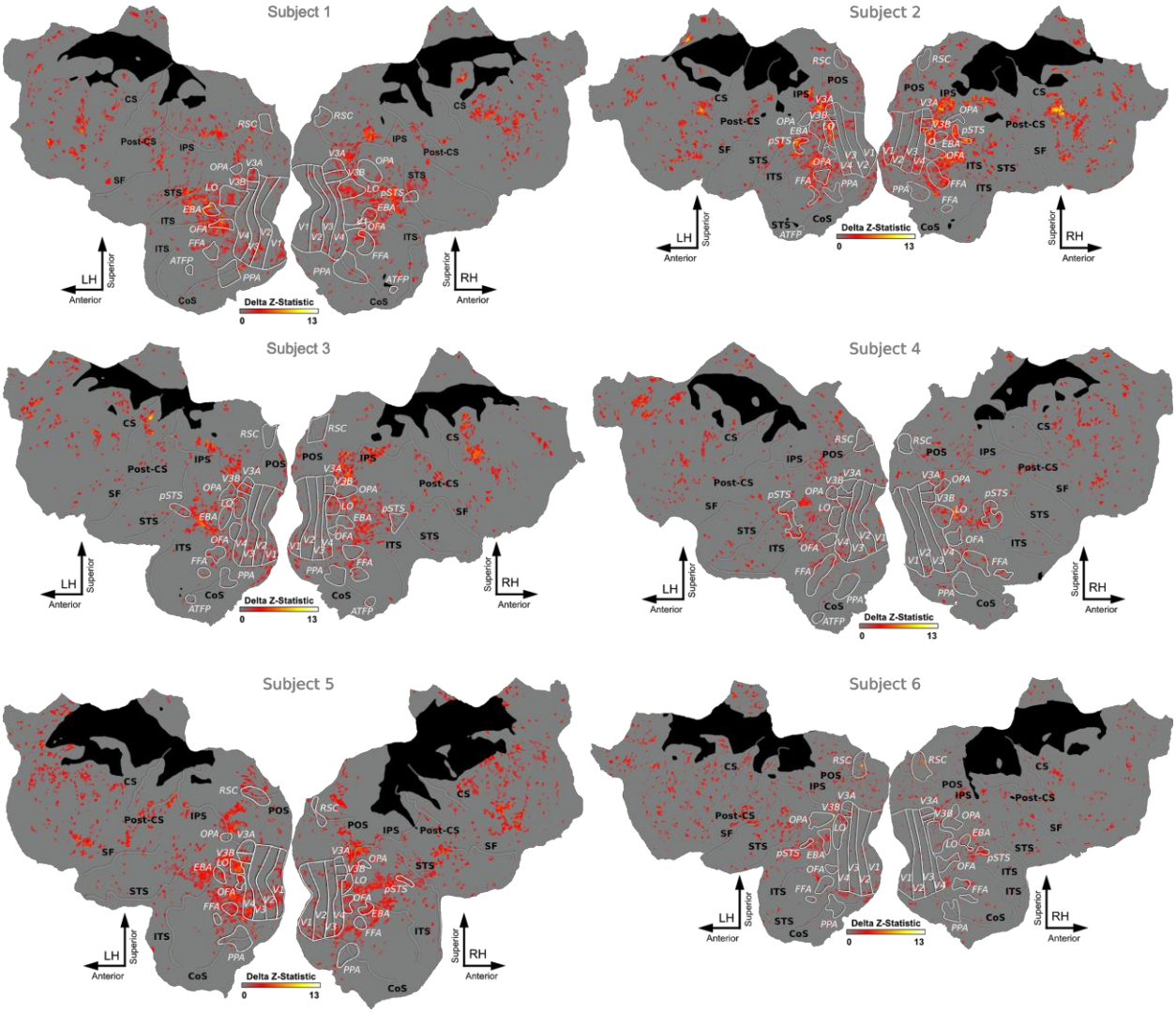


**Figure 2-S2. Improvement in voxel-wise prediction accuracies for the CSVA model relative to the Semantic Only Model.**

Cortical maps for each subject show voxels where prediction accuracy was greater for the CSVA model than the Semantic Only model. Voxel-wise prediction accuracy values for each model were z-transformed and subtracted (see the Quantification and Statistical Analysis section for full details). Only voxels whose activity was significantly predicted by any one of the following three models were included in these comparisons: CSVA, Emotion Only and Semantic Only.

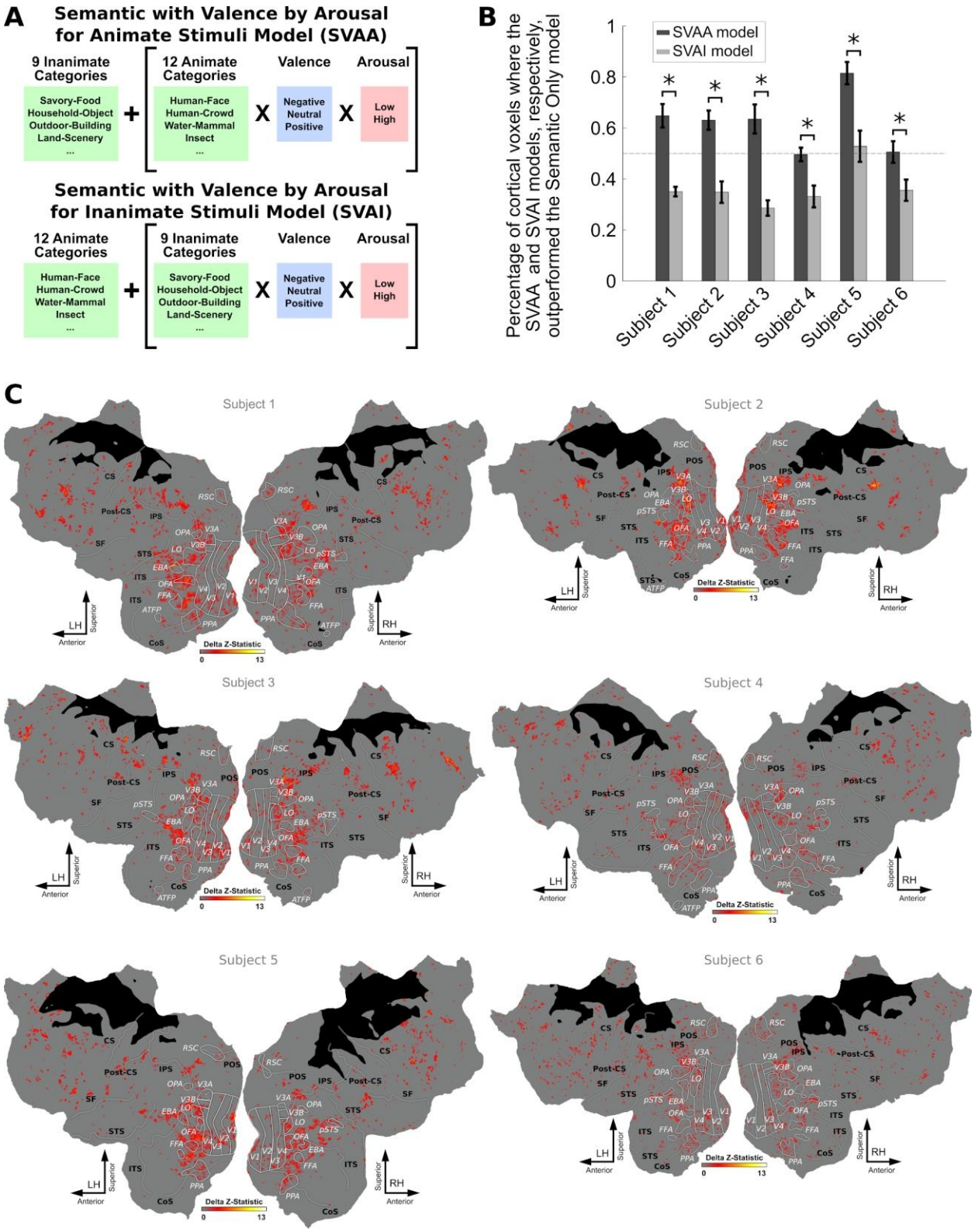


**Figure 2-S3. Cortical tuning to stimulus emotional content is greatest for animate stimuli.** (A). We examined the extent to which including the valence and arousal of either animate or inanimate stimuli improved model fit over and above modeling semantic category information alone. To investigate this, we

constructed two additional models. The Semantic with Valence by Arousal for Animate Stimuli (SVAA) model includes features for each semantic category, but only stimuli belonging to animate semantic categories are also labeled for valence and arousal. The Semantic with Valence by Arousal for Inanimate Stimuli (SVAI) model includes features for each semantic category, but here only stimuli belonging to inanimate semantic categories are labeled for valence and arousal. (B) This plot shows the percentage of cortical voxels where the SVAA and SVAI models, respectively, outperformed the Semantic Only model, for each individual subject (see the Model Comparison section of the Star Methods for details). For all six subjects, labelling valence and arousal features for animate images improved model fit to a greater extent than labelling valence and arousal features for inanimate images (* = $p<.005$). Significance was calculated using a bootstrap test (see the Quantification and Statistical Analysis section of the Star Methods for details). (C) Here we show relative prediction accuracies for the SVAA model versus the SVAI model projected onto cortical maps for each subject. This subtraction effectively indicates the extent to which baseline prediction accuracies achieved by modeling image semantic category are improved to a greater extent by including valence and arousal features for animate than for inanimate stimuli.

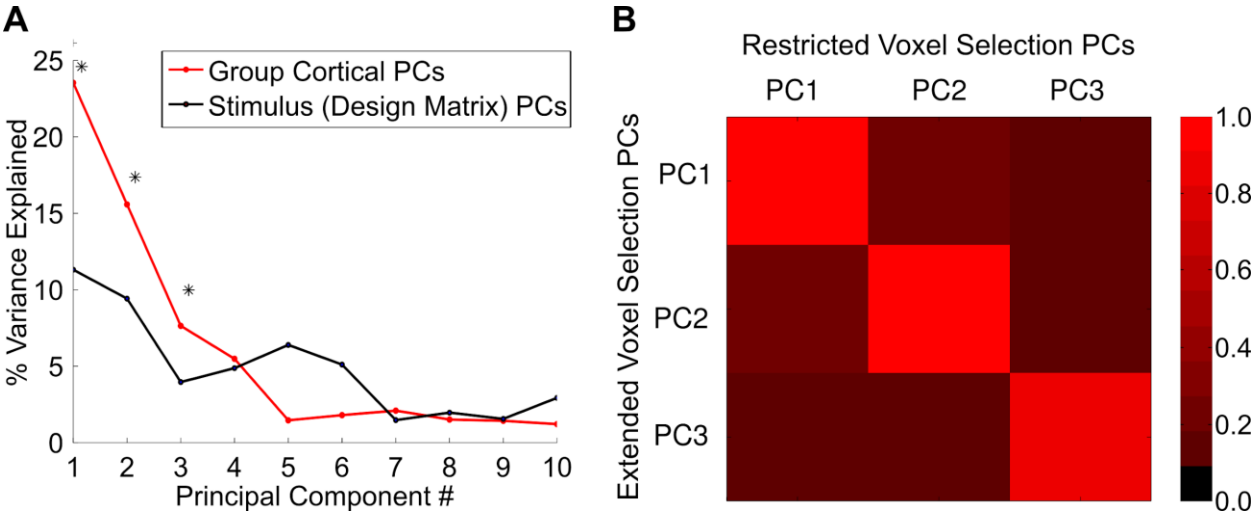


**Figure 2-S4. Results of group-level PCA on CSVA model feature weights when voxel selection is expanded to include all cortical voxels where the CSVA model showed a significant fit.**

(A) The top three PCs from this analysis explained significantly more variance than the top three stimulus PCs (jackknife test, $p < 10\text{-}8$;, stimulus PCs derived from PCA on the design matrix concatenated across subjects). This parallels the finding from the PCA reported in the main text where two additional voxel selection criteria were used (voxels within OTC only; CSVA model fit > Semantic Only model fit), see Fig. 2-4. (B) Correlation matrix shows correlations between the top three Group PCs from this 'expanded' voxel-selection PCA and the 'restricted' voxel selection PCA reported in the main text. Each corresponding PC shows a correlation of r>=0.95.

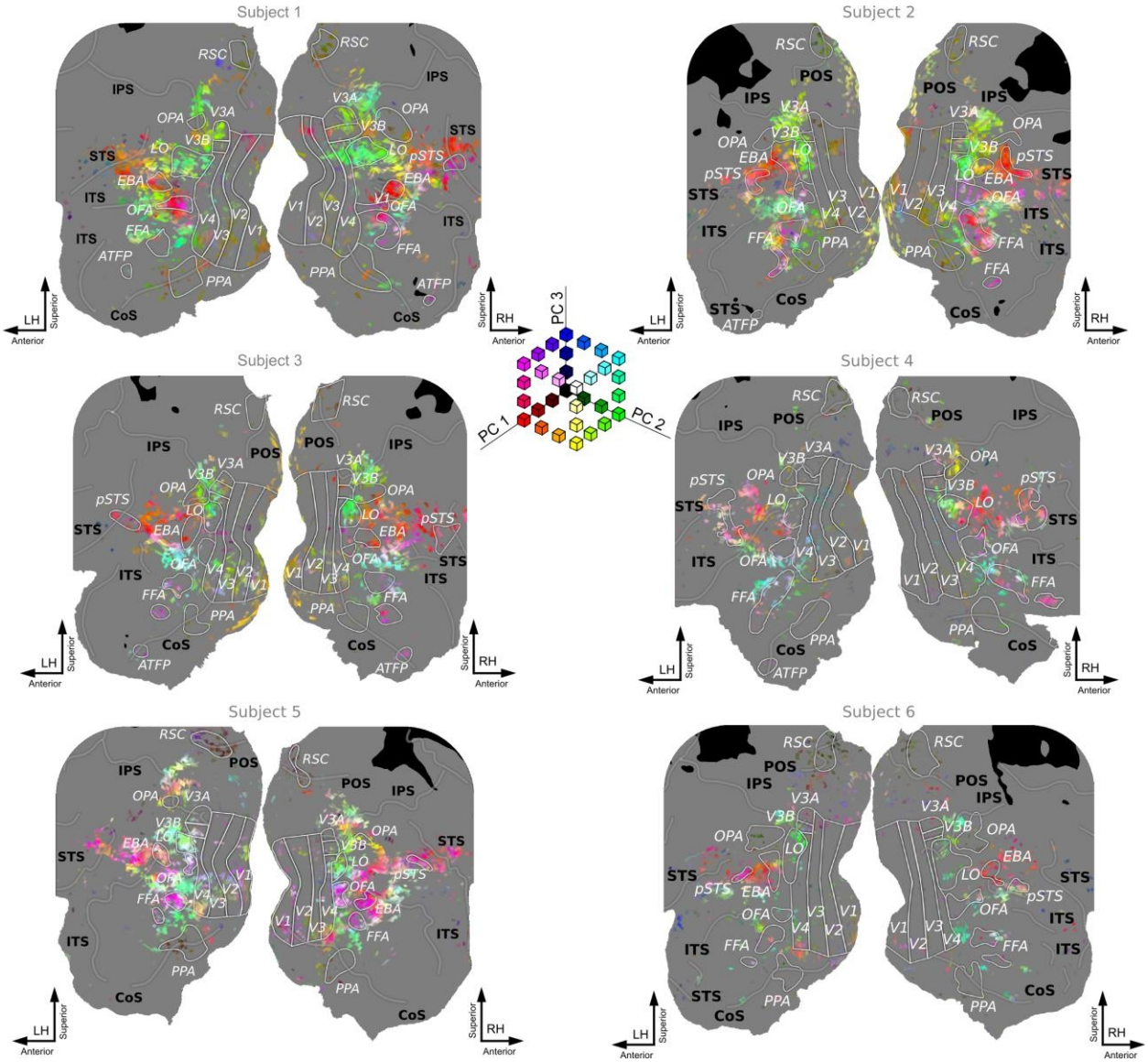


**Figure 2-S5. PC scores from the CSVA model are projected onto OTC flat maps for each subject.**

A principal components analysis (PCA) was conducted on CSVA model feature weights for all OTC voxels where CSVA model fit was significant and superior to that of the Semantic Only model. PC scores were calculated as the product of CSVA feature weights for a given voxel by feature loadings for each PC. Here, a RGB color space is used to map PC scores onto cortex (red = scores on PC1, green = scores on PC2, blue = scores on PC3). PC scores are thresholded at 6 standard deviations above and below 0 with values beyond the threshold given the maximal (or minimal) color channel value. Consistent spatial structure of voxel-wise tuning to the top three group PCs is observed across subjects. Note. Areas where MRI data was not acquired are shown in black. Both voxels where the CSVA model did not fit significantly and those where the CSVA model fit significantly but did not outperform the semantic only model were excluded from the PCA (these voxels are shown in grey). PCA maps using CSVA model feature weights from all voxels where the CSVA model fit significantly are given in fig. 2-S6.

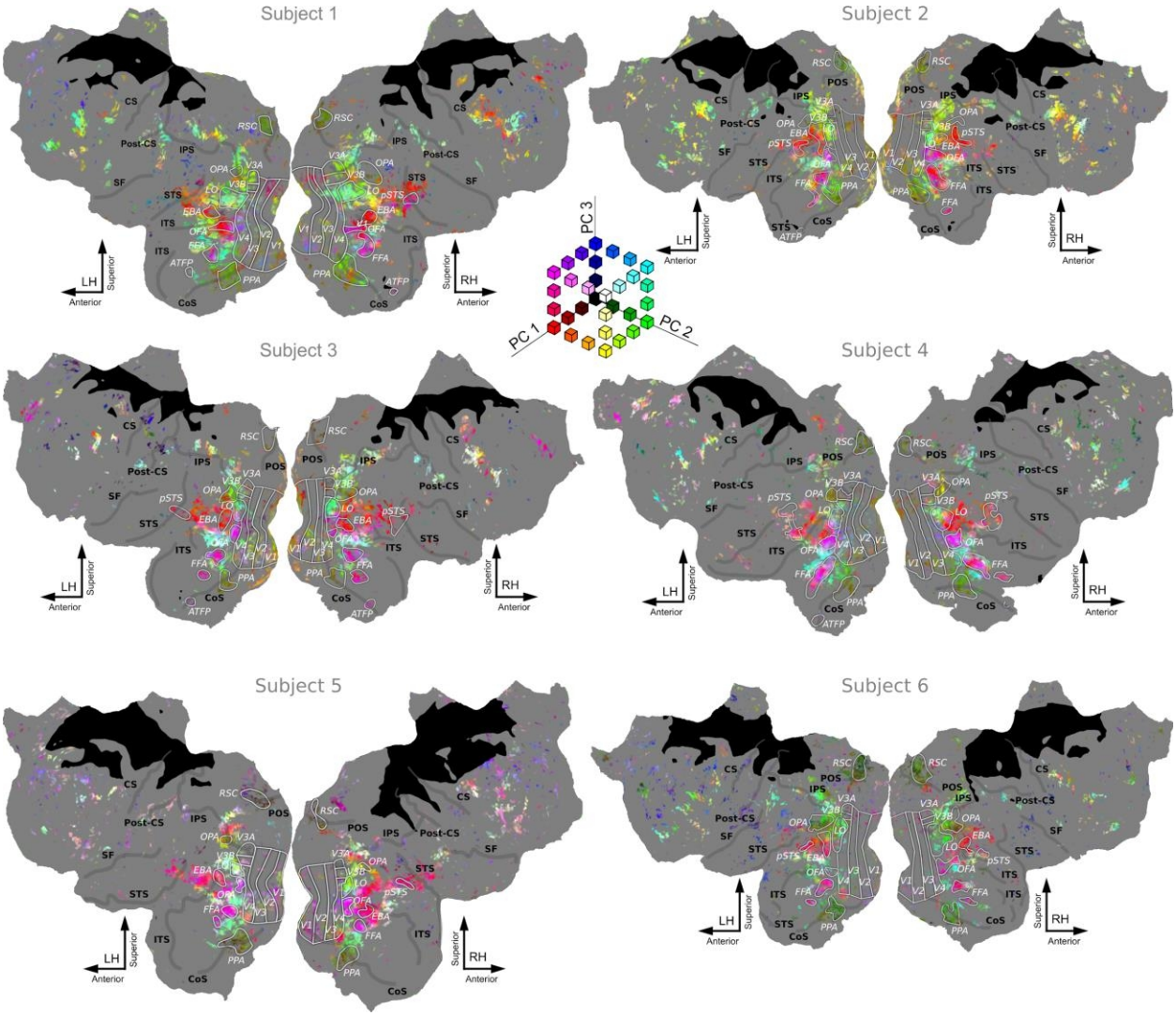


**Figure 2-S6. PC scores from PCA on CSVA model weights, across all cortical voxels where the CSVA model fits, are projected onto cortical flat maps for each subject.**

Here, we map PC scores onto cortex using the top three PCs from PCA on CSVA model weights across all cortical voxels where the CSVA model fit was significant. As illustrated in fig. 2-S4, when voxel selection is expanded in this manner, the top three PCs are highly correlated with those from the OTC analysis reported in the main text (rs>.95). As in Fig. 2-5 and fig. 2-S5, a RGB color space is used to map PC scores onto cortex (red = scores on PC1, green = scores on PC2, blue = scores on PC3). PC scores are thresholded at 6 standard deviations above and below 0.

# Chapter 3. Addressing Limitations of Chapter 2

## Introduction

In the previous chapter we reported findings from a voxel-wise modeling (VWM) fMRI study that explored representation of naturalistic emotional stimuli within occipital-temporal cortex (OTC). We found that subjective emotional valence and arousal, in addition to semantic category information, were represented within OTC, that this representation of emotional content was predominantly present for animate as opposed to inanimate stimuli, and that tuning to combined emotional and semantic content within OTC better predicted appropriate behavioral responses than did the image features themselves, OTC tuning to low-level structural image information, or OTC tuning to semantic or emotional content alone. Here, I address potential limitations of this study at both a theoretical and empirical level.

First, I will address potential theoretical limitations of the study through two literature reviews. In chapter 2 we claim that we've found evidence for representation of emotional content in addition to semantic content. One potential criticism is that we have not clearly specified how we define 'semantics'. Of pertinence to this is an active debate as to whether OTC simply differentiates visual object categories (Goodale & Milner, 1992), or represents more abstract semantic information. If OTC representations can be used to determine appropriate behavioral responses to emotional stimuli, as we suggest in chapter 2, then it is important that OTC can represent the abstract semantic content of those emotional stimuli, and not just visual features of them. Thus, the first literature review (section 1) will report on a fairly new body of work which addresses the question of whether OTC representation subserves visual object category information, or whether it represents supramodal (i.e. a combination of visual, auditory, gustatory, or tactile stimuli) semantic information.

The second theoretical issue to be addressed concerns whether we can distinguish representation of stimulus emotional content from influences of emotional salience on attention. While studies of how emotional visual stimuli are represented within OTC are few, there is a large literature which has studied how emotional visual stimuli influence attention. The behavioral effects, and neural underpinnings, of this type of attentional modulation will be reported in the second literature review (section 2).

In addition to the above theoretical issues, this second chapter will also address some empirical limitations of the study reported within chapter 2. In total, seven quantitative follow-up analyses are reported, conducted e using the same fMRI data set as used in the analyses of chapter 2. A recent study (Chikazoe et al., 2014) exploring representation of semantic and valence information within ventral temporal cortex (VTC) found representation for both using a form of multivoxel pattern analysis (MVPA) called representational similarity analysis (RSA), as well as a traditional univariate SPM analysis. In chapter 2 we argued that using VWM, which is a powerful new approach to neuroimaging, allowed us to discover the combined representation of semantic and emotional tuning within OTC that we found. In order to demonstrate whether the

findings we've reported on in chapter 2 could have been found using other methods, both RSA and univariate SPM analyses were conducted. We report those in sections 3 (RSA) and 4 (univariate SPM analysis) of follow-up analyses.

Figure 2-2 (and figs 2-S1 - 2-S3) shows the results of a model comparison between the full combined semantic, valence, and arousal (CSVA) model and semantic only and emotion only models. This revealed that the CSVA model outperformed both semantic only and emotion only models within OTC. Variance partitioning is a technique that allows the total variance explained by a model to be partitioned into unique and combined partitions for several subsets of model features. In section 5 we report findings from a variance partitioning of CSVA model explained variance across cortex, which is a more precise way of determining whether the full CSVA model outperforms models containing either only semantic or emotional features.

In chapter 2, a PCA was done across voxel feature weights within OTC to determine the primary dimensions of tuning within that region, and those dimensions were then used to predict appropriate behavioral responses to all stimuli used in the study. It is known that representations within regions of early-visual cortex (EVC) such as V1 are well modeled using low-level structural features of visual stimuli (such as Gabor wavelets), while regions of OTC not within EVC are better modeled by semantic categories than low-level features (Naselaris et al., 2009). In order to determine whether the results of the OTC PCA done in chapter 2 could be driven by voxels within EVC, two separate PCAs were done on voxels within EVC, and those within OTC regions outside of EVC. The results from these analyses were then used to predict appropriate behavioral responses to the study images, as was done in chapter 2. Those results are reported in section 6.

Chikazoe and colleagues (2014) and others (Ongur & Price, 2000; Shenhav, Feldman-Barrett, & Bar 2013) have reported representation of emotional valence within the part of frontal lobe known as orbitofrontal cortex (OFC). In order to compare our findings from OTC with OFC, and other frontal regions, additional PCAs were conducted on voxel feature weights from OFC voxels, as well as all frontal voxels not within OFC (motor cortex was also excluded) and reported in section 7.

While OTC is better modeled by semantic categories than low-level structural features (Naselaris et al., 2009), regions within OTC are still retinotopically organized (Arcaro et al., 2009; Grill-Spector & Weiner, 2014) and thus any correlations between CSVA model stimulus condition and location of that stimulus within the visual field could result in type II error that is not about the CSVA conditions of interest, but rather our choice of stimuli. Another key issue is that fMRI coding may reflect axonal inputs into a region rather than the region's representational tuning itself (Logothetis, 2001; Logothetis et al., 2008). Hence controlling for gabor representation allows us to control for such influences from early visual cortex upon the bold activation in later OTC regions. In order to determine whether the results of the PCA on voxel feature weights conducted within OTC were not a result of retinotopic organization correlated with semantic category, we controlled for the effects of a Gabor wavelet model on the BOLD

signal and re-fit the PCA. Results from this new PCA, and their comparison with the original PCA results from chapter 2, are reported in section 8.

Finally, physiological arousal is known to modulate the fMRI BOLD signal (Verstynen & Deshpande, 2011). It has been argued that emotional stimuli may give rise to physiological arousal and hence activation thought to reflect the emotional content of stimuli might merely reflect associated physiological arousal. To control for this, we included nuisance regressors in our VWM CSVA model created from respiration rate and pulse oximetry data collected at scan time. The results of this control analysis are reported in section 9.

# 1. Literature Review: Supramodal Representations within OTC

As part of the visual cortex, the ventral occipital temporal cortex (VOTC) has traditionally been thought of as playing "the major role in the perceptual identification of objects" (Goodale & Milner, 1992). In that seminal 1992 work, Goodale & Milner suggested that the visual cortex can roughly be divided into two pathways. The ventral pathway, consisting of regions within VOTC and labelled the "what" pathway, was thought to be responsible for object recognition. The dorsal pathway, consisting of regions of the posterior parietal cortex and labelled the "where" pathway, was thought to "mediate the required sensorimotor transformations for visually guided actions". A large body of neuroscientific research exists which spans decades and which has examined the functional properties of neural processing within visual cortex (for a review see Grill-Spector & Weiner, 2014). Recently however, the field has seen a renewed interest in VOTC processing following the huge successes of artificial deep neural networks to classify image object categories, starting with AlexNet in 2012 (Krizhevsky, Sutskever & Hinton, 2012). This renewed interest has raised some interesting new questions concerning the nature of processing within VOTC. Does VOTC simply extract object categories for use by other brain regions responsible for cognitive processes such as memory, decision making and cognitive control as the traditional view holds? Or are its representations idiosyncratically specialized to subserve the appropriate behavioral outcomes, such as navigation, social cognition and tool use, that are appropriate to the various types of objects it is selective to? Is the observed VOTC category selectivity supramodal in nature, integrating information for auditory, tactile, and other sensory modalities? And are the representations found within VOTC semantic in nature, or is the object selectivity observed simply a function of groupings of complex visual features? While these questions are still largely unanswered, several recent and compelling review articles make arguments in attempts to answer them.

The first of these articles, by Ricciardi et al (2014), reviews a growing body of literature investigating both VOTC processing of object categories and posterior parietal processing of motion discrimination and spatial localization, all within the congenitally blind. Previous work in sighted individuals has found evidence for multi-modal object representations (including auditory, tactile, olfactory, and gustatory stimuli) within VOTC regions (Amedi et al., 2005a, 2005c; Peelen et al., 2010; Ricciardi et al., 2006; Lacey et al., 2007; Pascual-Leone and Hamilton, 2001). These findings suggest that VOTC may represent more than simply visual object categories, but rather represent supramodal semantic information. As many authors have pointed out however, these findings could simply be VOTC activity that results from the recall of visual imagery, which is known to elicit similar VOTC responses to those elicited by visual stimulation of the same stimulus (Ishai, 2010; Ishai et al., 2000). Studies of supramodal VOTC object category selectivity within the congenitally blind overcome the limitations of these studies, as congenitally blind individuals have no visual memories to recall. Nearly 40 such studies outlined by Ricciardi and colleagues (2014) have shown similar patterns of activity within VOTC both between sighted and congenitally blind subjects, and across sensory

modalities (for example: De Volder et al., 2001; Lambert et al., 2004; Pietrini et al., 2004; Poirier et al., 2006; Ptito et al., 2009; Ricciardi et al., 2007; Vanlierde et al., 2003; Weeks et al., 2000). These studies use stimuli such as: visual images; auditory sounds elicited by the objects in question; written, spoken or braille words; haptic stimulation; and sensory substitution devices, which create non-visual "sensory-scapes" through training. For example, Mahon et al. (2009) found that the well studied animate-inanimate gradient found from medial-to-lateral VOTC is present in congenitally blind subjects using auditory stimuli, and sighted subjects using both auditory and visual stimuli. Ricciardi (2014) also present evidence from numerous studies on cross-modal cortical plasticity in both the congenital/early blind and patients with later onset blindness. Here a differentiation is made between cross-modal plasticity, defined as the "potential of the human brain to reorganize itself" under sensory deprivation (blindness in this case), and supramodal representation, defined as an abstract structural or semantic representation not attributable to cortical plasticity. These studies highlight differences between congenitally/early blind subjects and sighted subjects. Task difficulty differentially affects the recruitment of occipital cortex between congenitally blind subjects and sighted subjects, where higher cognitive, perceptual, and simple sensorimotor tasks activate occipital cortex in sighted subjects but only higher cognitive and perceptual tasks do so in congenitally blind subjects (Noppeney, 2007). Additionally, cross-modal functional activations are smaller in those with later onset blindness than in those with congenital/early blindness, suggesting that cortical plasticity decreases as a function of the age at which blindness occurs. Ricciardi and colleagues argue that, while cross-modal cortical plasticity is a real phenomenon in which regions of sensory cortex not utilized due to lack of afferent innervation are repurposed, this phenomenon cannot account for the similar supramodal activation observed in both sighted and blind subjects such as that seen in the Mahon et al. 2009 study. While these findings provide compelling evidence of supramodal semantic representation of object category within VOTC, in a recent paper, van den Hurk et al. (2017) point out that these studies utilize congenitally/early blind subjects, most of whom have had some minimal visual stimulation, if not visual experience. To address this concern, van den Hurk and colleagues studied congenitally blind subjects where a portion had anophthalmia, a disorder characterized by an absence of oculi (eyes). Using face, body, scene and object related natural sounds they found that information contained within VOTC responses in the congenitally blind could discriminate these stimulus categories. Furthermore, they found that MVPA models built using that the patterns of activity in blind subjects could successfully be applied to the brain activity of sighted subjects in order to predict the visual object categories seen when that brain activity was recorded. This is perhaps the strongest evidence to date that VOTC represents supramodal semantic information concerning object categories.

In a more nuanced reading of the literature regarding supramodal representation within VOTC, Bi, Wang & Carmazzo (2014) concluded that supramodal (or multi-modal) representations within VOTC exist in a heterogeneous fashion dependant on the object category, or domain, of the stimulus. Bi argues for a tripartite distinction in multi-modal representations

within VOTC across object domains related to (i) spatial navigation, (ii) animate (non body-part) items, and (iii) body-parts and manipulable objects. Regions within medial fusiform gyrus and parahippocampal gyrus are well known to show selectivity to pictures of scenes and buildings (Epstein & Kanwisher, 1998), and have been implicated in spatial navigation tasks (Kamps, Lall & Dilks, 2016; Park & Chun, 2009). Across several studies of the congenitally blind, supramodal representations were found within these regions, elicited by various domains including haptic touch (Wolbers et al., 2011) and auditory words (He at al., 2013). Studies of sighted individuals showed similar results, eliciting multi-modal VOTC activity to scene and building stimuli (Fairhall et al., 2014; Adam & Noppeney, 2010). Animate items more strongly activate regions of lateral posterior fusiform cortex. In the congenitally blind, various forms of stimuli depicting either human faces or non-human animals elicited no lateral posterior fusiform selectivity when presented as auditory words (Mahon et al., 2009), mental imagery (Lambert et al, 2004), or haptic stimulation (Kitada et al., 2010). Numerous studies in sighted studies showed conflicting results for supramodal representation within lateral posterior fusiform cortex. Lateral occipital temporal cortex (LOTC) shows strong selectivity to images of body parts and small, manipulable objects. Across multiple modalities, selectivity to tools in LOTC was seen in both sighted (Lewis et al., 2005; Noppeney et al., 2006; Tranel et al. 2005) and congenitally blind subjects (Peelen et al., 2013). Studies that tested activity related to body-parts also found supramodal representation within LOTC, including verbal or haptic stimuli in sighted subjects (Costantini et al, 2011) and haptic and sensory substitution devices in congenitally blind subjects (Striem-Amit & Amedi, 2014; Kitada et al., 2014). Selectivity to words (Le et al. 2000; Rogers et al., 2006) and visual imagery (Kitada et al. 2009) in sighted subjects did not not show supramodal representation to tools or body-parts within LOTC, however. Thus, a closer reading of the empirical findings regarding supramodal representation within VOTC reveals evidence for supramodal representation of some object categories (i.e. scenes/buildings and body-parts/tools), but not other object categories (i.e. animate human faces, bodies and non-human animals). This conclusion is in contrast to the conclusions made by Ricciardi and colleagues that supramodal representation exists within VOTC without regard to object category. Interpreting these findings as a whole, Bi and colleagues further speculate that the relationship between the visual shape and functional relevance of the objects in this tripartite domain classification drives these observed differences in VOTC supramodal representations. They argue that visual shape strongly constrains our ability to physically interact with inanimate objects, and to a much smaller degree, and in a less articulated manner, for animate objects. This differentiation suggests that visual features of inanimate artifacts may be defined both by their visual characteristics and their potential action value, and thus their representations may be accessible through different modalities, making them supramodal. For example, a long thin object (e.g. a knife) affords the act of cutting while a flat solid object (e.g. mallet) affords the act of pounding. Whereas the affordances animate objects confer are much less dependent upon their size and shape than are inanimate artifacts. For example, a dog and wolf can have a very similar size and shape, but the dog often affords approach and petting while the wolf afords avoidance. Or a ladybug may afford

cultivation as they are considered cute and beneficial to gardens, but a black widow spider affords an avoidance response as it is deadly.

Extending the arguments made by Bi et al. that some object category representations within VOTC combine visual shape and action potential information, Peelen and Downing (2017) recently argued that object recognition of all object categories is unlikely to fully account for the category selectivity observed in VOTC. Rather, they argue that VOTC representations are likely tuned in an idiosyncratic fashion in support of various behaviors relevant to the different object domains, such as spatial navigation for scenes, social cognition for human faces and bodies and tool use for objects and body parts. They cite several studies which show that VOTC category selectivity can be dissociated from high-level visual object representations. For example, Weisberg et al, (2007) showed that LOTC activity to images of complex novel objects increased after training subjects to use those objects as tools, suggesting that LOTC represents how objects can be used as tools, not just the (perhaps view-invariant) shape representation of those objects. Pointing to the previously mentioned literature on the congenitally blind, Peelen & Downing argue that the supramodal nature of object representation within VOTC supports the hypothesis that these regions are tuned to allow for domain-appropriate behavioral responses. Additionally, they cite increasing evidence for selective connectivity between category-selective VOTC regions and other brain regions known to serve domain-related functions (Bracci et al., 2012; Hutchison et al., 2014; Simmons and Martin, 2012). For example, a recent functional connectivity analysis (Simmons and Martin, 2012) using the posterior superior temporal sulcus (implicated in biological motion) as a seed region found connectivity to multiple regions involved in social cognition, including fusiform face area, posterior cingulate/precuneus (involved in theory of mind), insula (associated with visceral-emotive responses), anterior temporal pole (representations of social information), medial prefrontal cortex (underlies social reasoning) and amygdala. Taken together, this combination of functional, as opposed to visual object, representation, supramodal representation, and connectivity with functionally related brain regions suggests that VOTC "category-selective responses are closely aligned with knowledge of what a thing means to the observer; that is, the cognitions and behaviors an object is associated with."

If semantic representation is defined as the combination of supramodal representation and tuning to object function, and not just shape, these recent empirical and theoretical advancements towards understanding representations within VOTC suggest that semantic information is indeed represented within VOTC. Using this definition of semantic information does not, however, eliminate the possibility that semantic information is also represented in other regions. For example, a large literature on semantic representations within the temporal pole (see Binder & Desai, 2011 for a review), as they relate to memory, indicates that indeed supramodal representations exist there as well. One possible explanation for this could be that supramodal semantic information is represented in different regions for use in different cognitive or behavioral processes. OTC could store supramodal semantic information relevant to appropriate behavioral action, while temporal pole represents supramodal semantic information relative to

memory formation and retrieval. For example, Simmons and Martin (2012) and Hutchison et al. (2014) showed that temporal regions (pMTG) implicated in tool use are functionally connected with ventral premotor regions. There is no a priori reason to believe that the brain contains only one region to store abstract semantic information about the world.

In chapter 2 we reported findings indicating that a combined semantic valence and arousal model (CSVA) predicts bold activity in much of OTC, and that it does better than either a semantic or valence by arousal model alone. A principal components analysis of voxel feature turning revealed that primary dimensions of representation within OTC include not only semantic information (animacy), but arousal for both animate and inanimate stimuli, and valence for animate stimuli. We then showed that at low dimensionality, tuning from OTC to the CSVA model better predicted appropriate behavioral responses to the emotional stimuli used to fit the model, than did stimulus features alone, low-level structural Gabor features, or semantics and valence by arousal alone. These findings are in line with the hypothesis that OTC represents semantic information tuned to appropriate behavioral response. Furthermore, it extends this theoretical position by suggesting that valence and arousal are important dimensions of representation within OTC that contribute to determining what the appropriate behavioral response should be.

# 2. Literature Review: Perception and Attention for Emotional Images

Within the attention and motivation literatures, a substantial number of studies have investigated the differential effects of negative (versus neutral or positive) stimuli on behavior, and to a much less degree of emotional (versus neutral) stimuli. Across numerous visual tasks, increases in attention to negative, relative to neutral or positive, stimuli have been observed (for reviews see: Vuilleumier, 2005; Pourtois, Schettino, & Vuilleumier, 2013). This increased attention confers "prioritized processing" upon stimuli of high survival value (i.e. emotional stimuli). For example, visual search tasks are done faster when the target is either a negative stimulus compared with a neutral stimulus (Ohman et al., 2001) or positive and negative stimulus relative to emotional stimulus (Eastwood et al., 2001), and this effect cannot be explained by low-level visual features such as contrast or luminance. In an emotional stroop task, with words classically conditioned to negative stimuli, subjects takes longer to read the words for emotional (negative) stimuli than they do for neutral stimuli (Richards & Blanchette, 2004). Tasks invoking the attentional blink are less often missed when the preceding stimulus is emotional (negative or positive) versus neutral (Anderson, 2005). Using binocular rivalry tasks, Alpers and colleagues have found that emotional (positive and negative) images are the first to be perceived, and occupy consciousness for more time, than do neutral images (Alpers & Pauli, 2006; Alpers & Gerdes, 2007). Additionally, fearful faces break into consciousness more quickly than do happy faces or neutral faces in a continuous flash suppression paradigm (Yang, Zald, & Blake, 2007). Finally, in tasks utilizing exogenous cueing (dot probe task) to elicit spatial orienting behaviors, invalid negative cues are more distracting that neutral invalid cues (Brosch et al, 2011; Mogg & Bradley, 1999; Pourtois, et al., 2005). This confluence of evidence for the prioritized processing of emotional stimuli raises the question, what brain mechanisms are responsible for this prioritized processing?

Increased activity to emotional stimuli, relative to neutral stimuli, has been repeatedly observed within sensory cortices across numerous fMRI, PET, EEG & MEG neuroimaging experiments. This effect has been found in visual cortex using varied stimuli types such as images of faces depicting both positive and negative emotional expressions (using PET: Morris et al., 1998; and using fMRI: Pourtois et al., 2010; Phan et al., 2002), negative threat words (using fMRI: Tabert et al., 2001), complex positive and negative emotional scenes (using PET: Lane et al., 1999; and using fMRI: Sabatinelli et al., 2005, 2007, 2010), and fear-conditioned stimuli (using fMRI: Armony & Dolan, 2002). Primary visual cortex (area V1) as well as several regions of extra-striate cortex (e.g. FFA, EBA) have shown this increased activity for emotional rather than neutral stimuli. Additionally, these effects seem to occur regardless of whether the subject is attending to the emotional stimulus or not (Morris et al., 1998; Vuilleumier et al., 2001). Enhanced activation to emotional stimuli has also been observed within auditory cortex, in response to angry and happy voices (Grandjean et al., 2005; Ethofer et al., 2012, 2009), as well as nonverbal vocalizations such as screams (Fecteau et al., 2007). Vuilleumier and

colleagues (2005, 2013) have argued that this increased activity in response to emotional versus neutral stimuli within sensory cortices is a plausible mechanism for the observed prioritized processing of such emotional stimuli, and may be mediated by the amygdala as will be discussed later on. Other forms of selective attention such as bottom-up (or exogenous) and top-down (or endogenous) attention are known to similarly increase activity within regions of sensory cortices relevant to the emotional stimulus. Vuilleumier and colleagues argue that emotional, or "motivated", attention should be considered as a separate form of attentional modulation. Several studies have shown a triple dissociation in terms of behavior and neural activity between bottom-up, top-down, and emotional attention (Brosch et al., 2011; Kiel et al., 2005). These three forms of attention modulation can all increase sensory cortical activity in an additive fashion, suggesting that three separate neural mechanisms are responsible for their observed behavioral effects. If emotional attentional is a dissociable process from endogenous and exogenous attention, that what brain regions subserve the sensory cortex modulation observed when perceiving emotional versus neutral stimuli?

The amygdala, a subcortical region located in the medial temporal lobe, has been implicated in a large number of studies as being critically involved in sensory cortical modulation to emotional stimuli. Vuilleumier and others (Vuilleumier, 2005; Lang and Davis, 2006; Amaral et al., 2003) have argued that, due to its strong bidirectional connections with sensory cortices (Amaral et al., 2003; Catani et al., 2003; Gschwind et al., 2012), it is a likely source of emotional gain control to those regions. Additionally, amygdala activity has shown correlations with visual cortical activity (i.e. functional connectivity) in a number of studies (Sabatinelli et al., 2005), in both primary visual cortex (Morris et al., 1998) and FFA (Pessoa et al., 2002). Additional evidence in support of this hypothesis has come from neuropsychological lesion studies. Examining epileptic patients with either unilateral hippocampal or amygdalar lesions using a spatial attention task and fMRI, Vuilleumier and colleagues (2004) found that while both groups of patients (and healthy controls) showed amplified FFA activity to task-relevant vs. task-irrelevant faces, only those patients with amygdalar lesions showed a lack of FFA amplification to fearful vs. neutral faces. Similar findings were reported for patients with epilepsy and medial temporal lobe sclerosis (and lesions), with lesion patients showing reduced activity in several OTC regions to fearful versus neutral faces relative to healthy controls (Amaral et al., 2003). Finally, patients with amygdala lesions do not show the same decrease in the emotional attentional blink task relative to healthy controls (Anderson, 2005), again suggesting amygdala's functional role in the modulation of emotional attention. While it seems likely that the amygdala plays a causal role in the amplification of sensory cortices which subserves emotional attention, there is still debate as to how signals reach the amygdala to serve this function in the first place.

Several theories hypothesize different roles that the amygdala may play in emotional visual processing. The traditional view, termed the two-route hypothesis, postulates that a dedicated subcortical "low-road" from the retina, to the superior colliculus, to the pulvinar and finally to the amygdala processes unconscious emotional (mostly threatening) visual stimuli, and

a “high road” from the retina, to the lateral geniculate nucleus (LGN) of the thalamus, and on to the visual cortex processes conscious emotional visual stimuli (Ohman & Mineka, 2001; Tamietto & de Gelder, 2010). Alternatively, two related theories (the “two-stage” theory: Vuilleumier, 2005; Pourtois, Schettino, & Vuilleumier, 2013 and the multiple-waves theory: Pessoa & Adolphs, 2010) both hypothesize that fast, feed-forward processing through the visual cortex propagates signals to the amygdala, which then sends re-entrant signals back to the visual cortex. The survival imperative that threatening stimuli be processed quickly to avoid bodily injury has led proponents of the two-route hypothesis to argue that cortical processing of visual stimuli could not explain the observed prioritized processing of emotional stimuli. However, new research regarding connectivity between sub-cortical visual regions and visual cortex has shown that numerous “short-cut” projections exist from LGN, superior colliculus and pulvinar to many regions of striate and extrastriate occipital, as well as ventral-temporal regions of visual cortex (Pessoa & Adolphs, 2010; Pourtois, Schettino & Vuilleumier, 2013). Latencies of information processing to visual cortex through these “short-cuts” are thus on par with processing latencies to the amygdala via the “low road” (Pessoa & Adolphs, 2010; Pourtois, Schettino & Vuilleumier, 2013). As Pourtois and colleagues argue (2013), there is thus

“no reason to assume that preattentive or unconscious emotion processing might be ‘magic’ and benefit from an exclusive ‘low’ route, just like evidence for subliminal perception fo words or numbers (Dehaene et al., 1998, 2001; Kouider and Dehaene, 2007) or unconscious priming of motor processes (Eimer and Schlaghecken, 2002, 2003) does not require the existence of dedicated subcortical pathways to the corresponding cortical areas.”

Whether or not the amygdala’s modulatory role in the processing of visual emotional stimuli is subserved by dual subcortical and cortical routes, or simply feedforward cortical processing, does not shed light on whether the amygdala is also necessary for the discrimination of emotional stimuli. Conclusions from early studies of epileptic patients with amygdala lesions indicated that indeed these patients showed deficits in the recognition of fearful faces and other signals of threat (Adolphs et al., 1994, 1995; Adolphs, 1999; LaBar et al., 1995). More recent studies have shown that this observed deficit in a patient (SM) with bilateral amygdala lesions is one of attention, and not actually a deficit of emotional discrimination (Adolphs et al., 2005; Kennedy and Adolphs, 2010). Perceiving the shape of the eyes is crucial to the discrimination of fearful facial expressions. As shown using eye tracking, SM was not attending to the eye regions of the faces with fearful expressions. Once given explicit instruction to attend to the eye regions, SM’s observed deficit in discrimination disappeared. This neuropsychological evidence strongly suggests that the deficits from amygdala lesions are attentional, and not perceptual per se, in nature. Thus, at least in patient SM, processing done within the amygdala is not necessary for discrimination of different emotions.

The experiment described in chapter 2 reported finding representations of arousal and valence for primarily animate stimuli across regions of OTC. These findings are consistent with the “two-stage” and “multi-wave” theories of emotional attention, which hypothesize that fast, feed-forward information processing within the ventral stream (OTC) of the visual cortex

projects to the amygdala, which in turn amplifies the representations of the emotional stimuli within visual cortex. We found that the primary dimensions of representation within OTC contain elements of arousal for both animate and inanimate stimuli and that this representation was much higher for animate than inanimate stimuli. Additionally, representation of valence within the top 3 dimensions of OTC was only found for animate stimuli.

Given the evidence that activity within OTC increases for emotional vs. non-emotional (i.e. high vs. low arousal) images, could this "gain function" for emotional stimuli explain our findings from chapter 2? While it is almost certain that the emotional stimuli used for the VWM study of chapter 2 elicited emotional attention mechanisms, this explanation cannot account for all of our findings for several reasons. First, a simple "gain" should affect all OTC voxels in a similar fashion, but we saw different regions of OTC more impacted by arousal, as represented in PC1 scores, than others. Second, if a gain function was responsible for our findings we would expect to see the same amount of representation of arousal for animate and inanimate stimuli, but we saw the opposite. PC 1 was much more highly correlated with arousal for animate stimuli than it was for arousal of inanimate stimuli. Finally, we found differential representation of negative and positive animate stimuli, as evidenced by a correlation of PC 3 with valence for animate stimuli. Taken together, these findings paint a picture of differential representation for emotional stimuli, not a simple gain function for arousing stimuli.

# 3. Representational Similarity Analysis (RSA)

## Introduction

Representational similarity analysis (RSA) is a form of multi-voxel pattern analysis (MVPA) that quantifies stimulus similarity across voxels in a brain region with stimulus similarity of one or more theoretical models of interest. In 2014 Chikazoe and colleagues used RSA to investigate representation of low-level structural features, animacy and valence of static images within early-visual cortex (EVC), ventral temporal cortex (VTC), and orbitofrontal cortex (OFC). Their findings suggested that low-level features were best represented within EVC, animacy was best represented within VTC, and valence was best represented within OFC. Additionally, they found significant representation of valence within VTC, although this representation was significantly less than that of animacy. Likewise, they found significant representation of animacy within OFC which was significantly less than representation of valence was within OFC. A searchlight analysis further revealed that representation of valence and animcy within VTC were not within the same regions. In contrast with these findings, in chapter 2 we argued for a combined representation of semantics, valence and arousal within occipital-temporal cortex (OTC). One possibility for this difference in findings is that RSA is not well suited to revealing co-representation of emotional content (e.g. valence) and animacy. Another possibility is that insufficient data was collected to provide the power needed to observe such representation. Here, we address this by conducting a RSA using the same well-powered dataset as was used for the VWM analysis of chapter 2. In addition, we separately model both valence and arousal, in contrast to Chikazoe et al., 2014 where only valence was modeled. We used full Spearman rank correlations to conduct the RSA due to its relaxed constraints (nonlinear monotonic relationship as opposed to the linear relationship of a Pearson correlation) on the relationship between the brain and model dissimilarity matrices, as well as being standard in the field (see Table 3-1 for review of RSA techniques used in recent relevant studies). We also used partial Spearman correlations to control for possible effects of low-level structural image features, as well as animacy and emotion.

To explore the multi-voxel pattern of activity within the regions of interest (ROIs) from both our chapter 2 study and Chikazoe's and colleagues' 2014 study, we created seven ROIs (see ROI section below). The first ROI is simply the occipital temporal cortex (OTC) ROI used in chapter 2. The second ROI contains voxels from regions V1-V4 in early visual cortex (EVC). Third, we created an OTC ROI with all the EVC voxels removed, to allow for an analysis of voxels within visual cortex that are not best characterized by low-level structural features (Naselaris et al., 2009). In order to compare these findings with those from the Chikazoe et al. (2014) paper, we also defined the same VTC and OFC ROIs as those authors had used. Additionally we created a second VTC ROI where the few EVC voxels contained within the VTC ROI were removed to again remove voxels best modeled by low-level structural features. Finally, as our results from chapter 2 indicated, the CSVA model we used fit within regions of

frontal cortex not contained within OFC. To better understand the representation within those frontal regions we defined a second frontal cortex ROI containing all the voxels in frontal lobe except for those within the OFC ROI and motor cortex.

## Methods

### Region of Interest (ROI) Definitions

The following seven Regions of Interest (ROIs) were defined in the following ways:

1. Occipital-Temporal Cortex (OTC): This ROI was defined anatomically and comprised all regions within the Occipital lobe and the Temporal lobe. The ROI was drawn onto the cortical surface, and began at the anterior portion of the Sylvian Fissure (lateral sulcus). It extends all the way up the Sylvian fissure continuing straight to the parieto-occipital sulcus on the lateral surface. It ends just above the retrosplenial cortex (RSC) on the medial surface. See Figure 3-1 below for flatmaps showing the voxels included in this ROI.

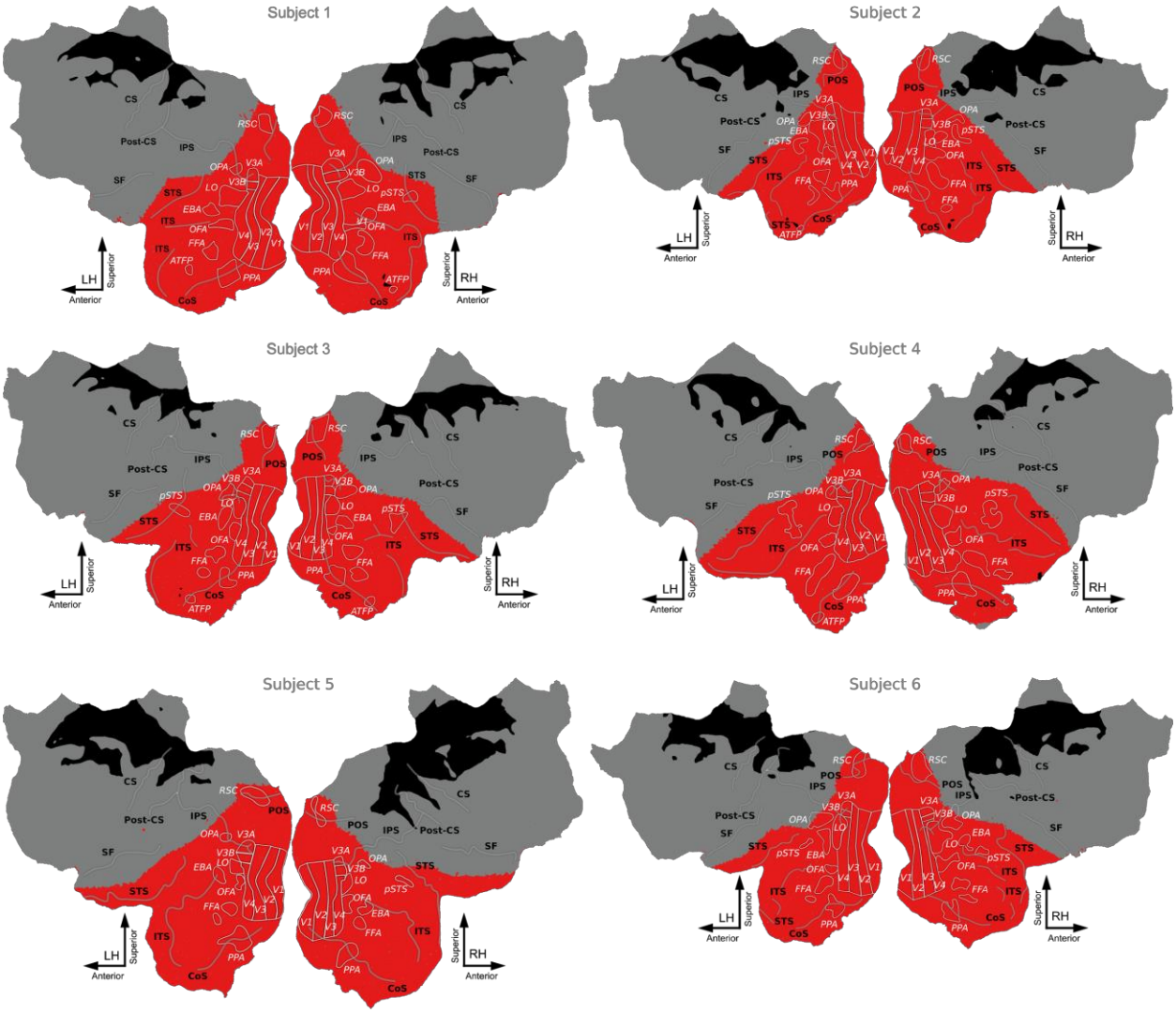


**Figure 3-1. Occipital-Temporal Cortex (OTC) ROI Projected onto the Cortical Surface.**
Flatmaps showing voxels included in the OTC ROI in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

2. Early-Visual Cortex (EVC): This ROI comprised functionally localized areas V1, V2, V3 & V4, defined using retinotopic mapping data collected across four scan runs, two with clockwise/counterclockwise wedges and two with expanding/contracting rings. (for details see chapter 2 methods and Huth et al., 2012). See Figure 3-2 below for flatmaps showing the voxels included in this ROI.

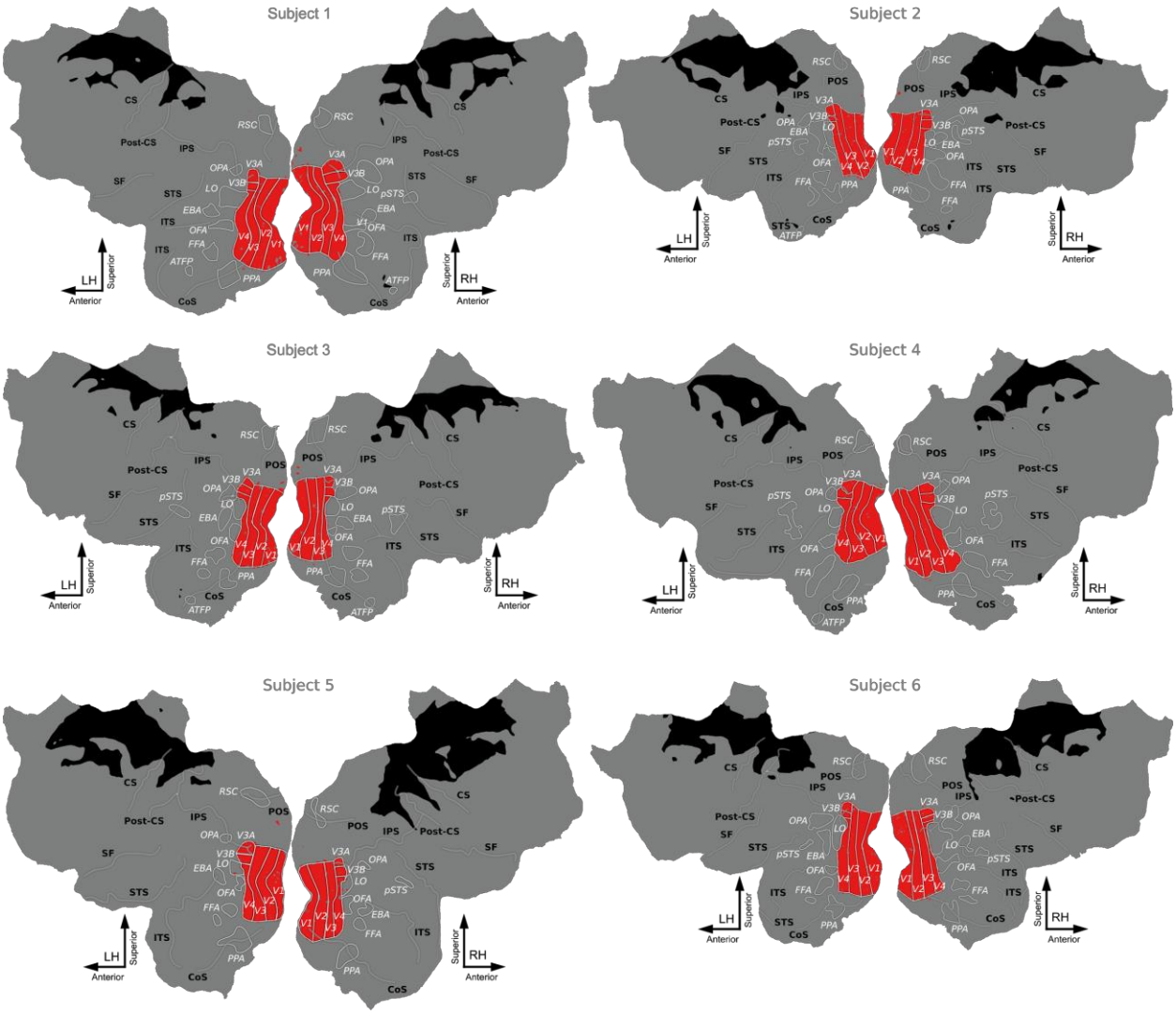


**Figure 3-2. Early Visual Cortex (EVC) ROI Projected onto the Cortical Surface.**
Flatmaps showing voxels included in the EVC ROI in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

3. non-EVC Occipital-Temporal Cortex (non-EVC OTC): This ROI included voxels from the entire Occipital and Temporal lobes, as defined in the OTC ROI, excluding voxels contained within early visual areas V1, V2, V3 & V4. See Figure 3-3 below for flatmaps showing the voxels included in this ROI.

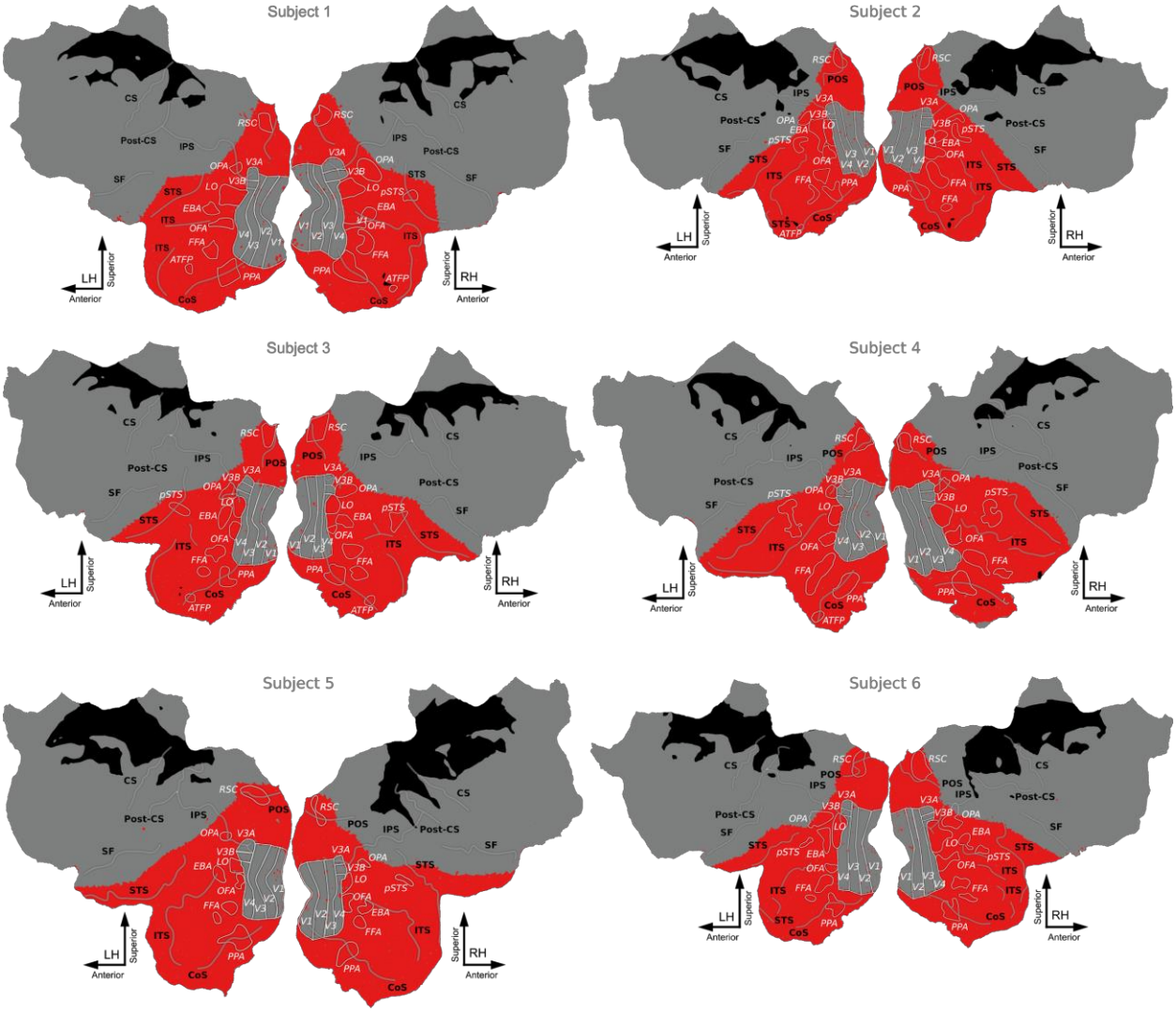


**Figure 3-3. non-EVC OTC ROI Projected onto the Cortical Surface.**
A flatmap showing voxels included in the OTC ROI, that were not within the EVC ROI, in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

4. Ventral Temporal Cortex (VTC): This ROI was defined just as the VTC ROI was defined in Chikazoe et al., 2014 using structural ROIs defined in the AAL template (Tzourio-Mazoyer et al., 2002; Rolls, Joliot, & Tzourio-Mazoyer, 2015). The following AAL regions were combined together: lingual gyrus, parahippocampal gyrus, fusiform gyrus and inferior temporal cortices in the bilateral hemispheres. These AAL ROIs were back-projected from MNI space to subject anatomical space using a non-linear transformation (inverse spatial normalization from SPM8), and then into subject functional space using a linear 12-dimensional affine transformation (spatial co-registration from SPM8). See Figure 3-4 below for flatmaps showing the voxels included in this ROI.

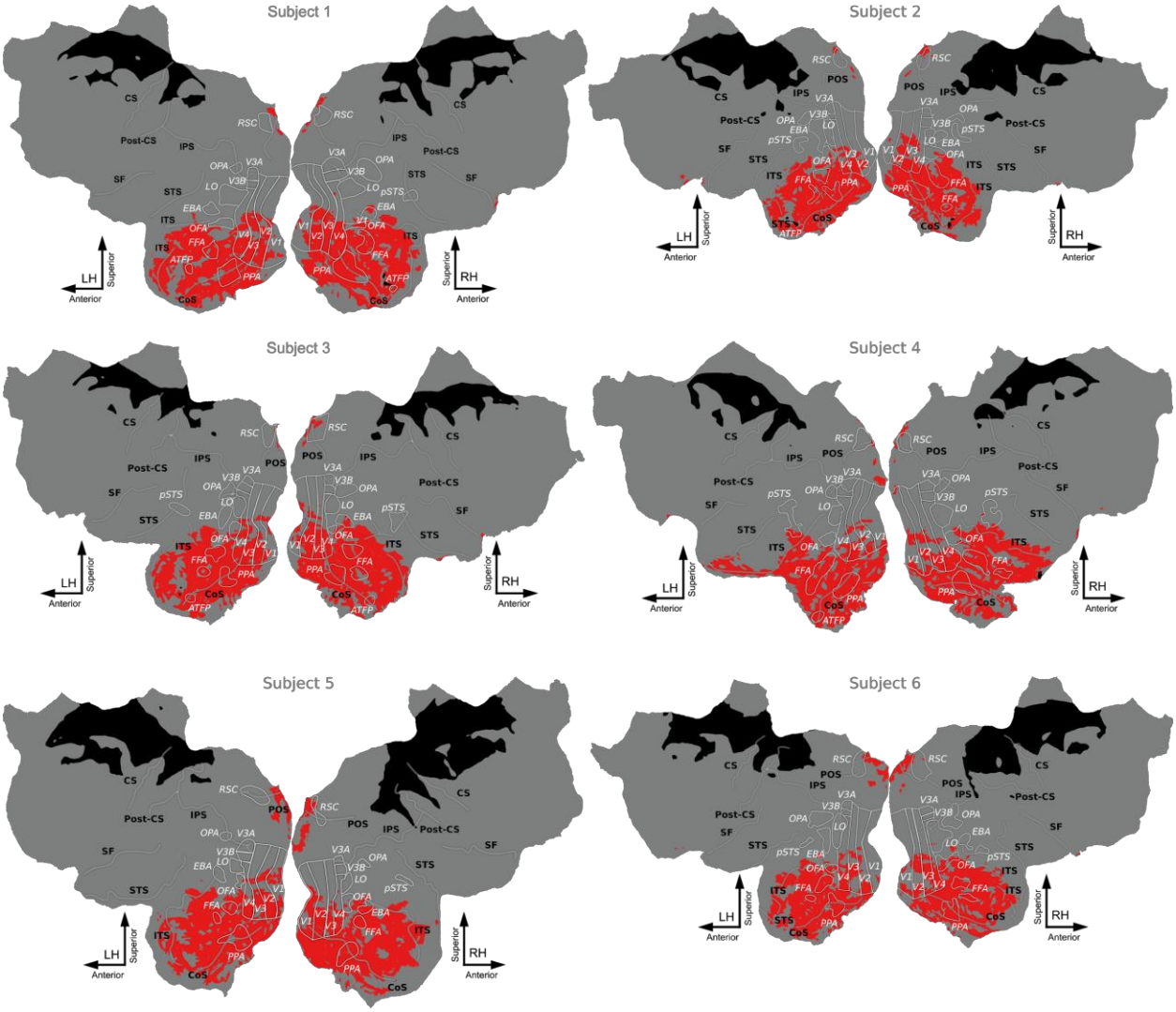


**Figure 3-4. Ventral Temporal Cortex (VTC) ROI Projected onto the Cortical Surface.**
A flatmap showing voxels included in the VTC ROI in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

5. non-EVC VTC: This ROI consisted of all VTC ROI voxels, as defined above, except those contained within functional localized regions V1, V2, V3, & V4 (EVC ROI). See Figure 3-5 below for flatmaps showing the voxels included in this ROI.

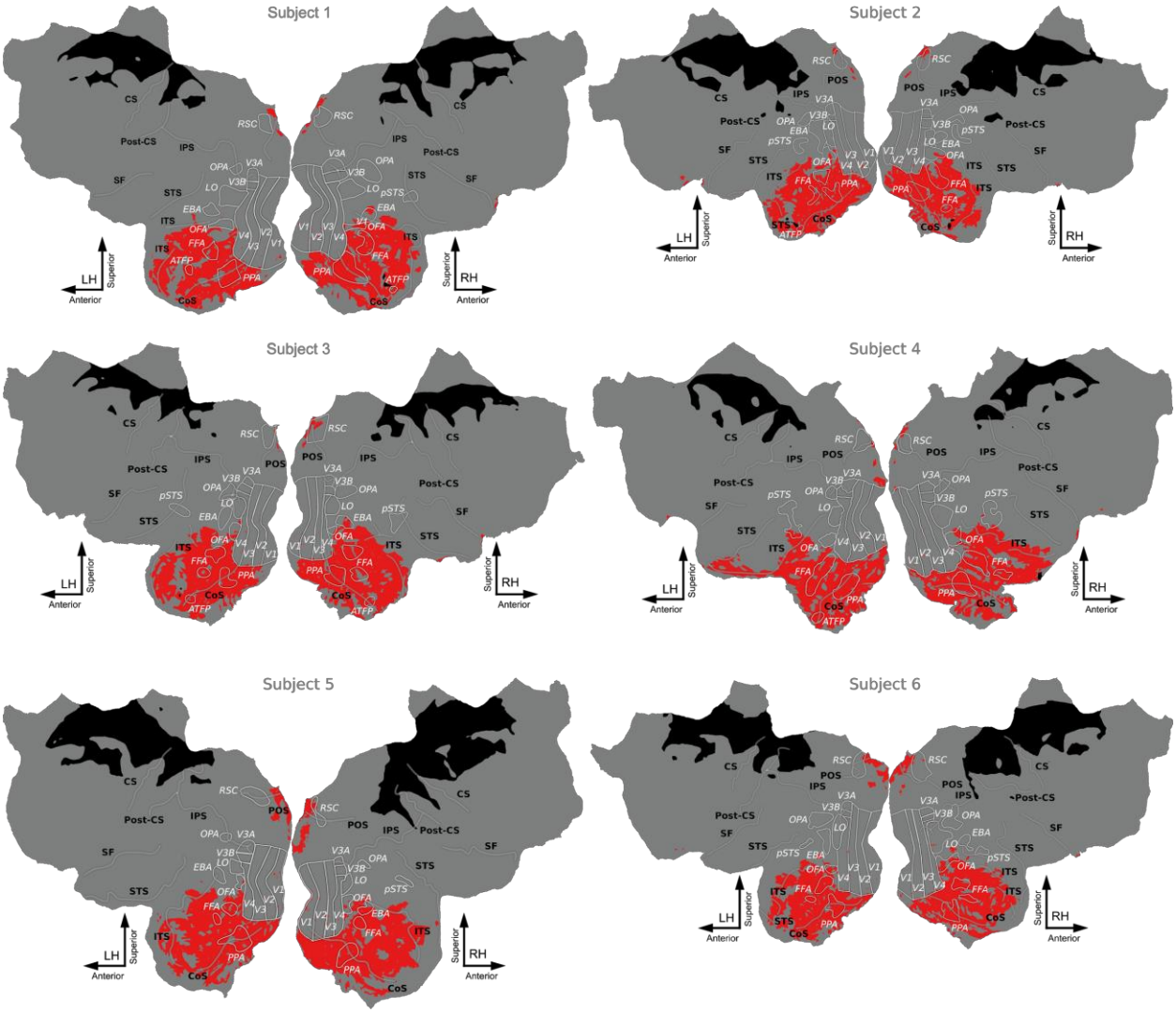


**Figure 3-5. non-EVC VTC ROI Projected onto the Cortical Surface.**

A flatmap showing voxels included in the VTC ROI, that were not within the EVC ROI, in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

6. Orbitofrontal Cortex (OFC): This ROI was defined using 4 structural ROIs defined in the AAL template (Tzourio-Mazoyer et al., 2002; Rolls, Joliot, & Tzourio-Mazoyer, 2015), namely the superior, middle, inferior and medial OFC in the bilateral hemispheres. These AAL ROIs were back-projected from MNI space to subject anatomical space using a non-linear transformation (spatial normalization from SPM8), and then into subject functional space using a linear 12-dimensional affine transformation (spatial co-registration from SPM8). See Figure 3-6 below for flatmaps showing the voxels included in this ROI.

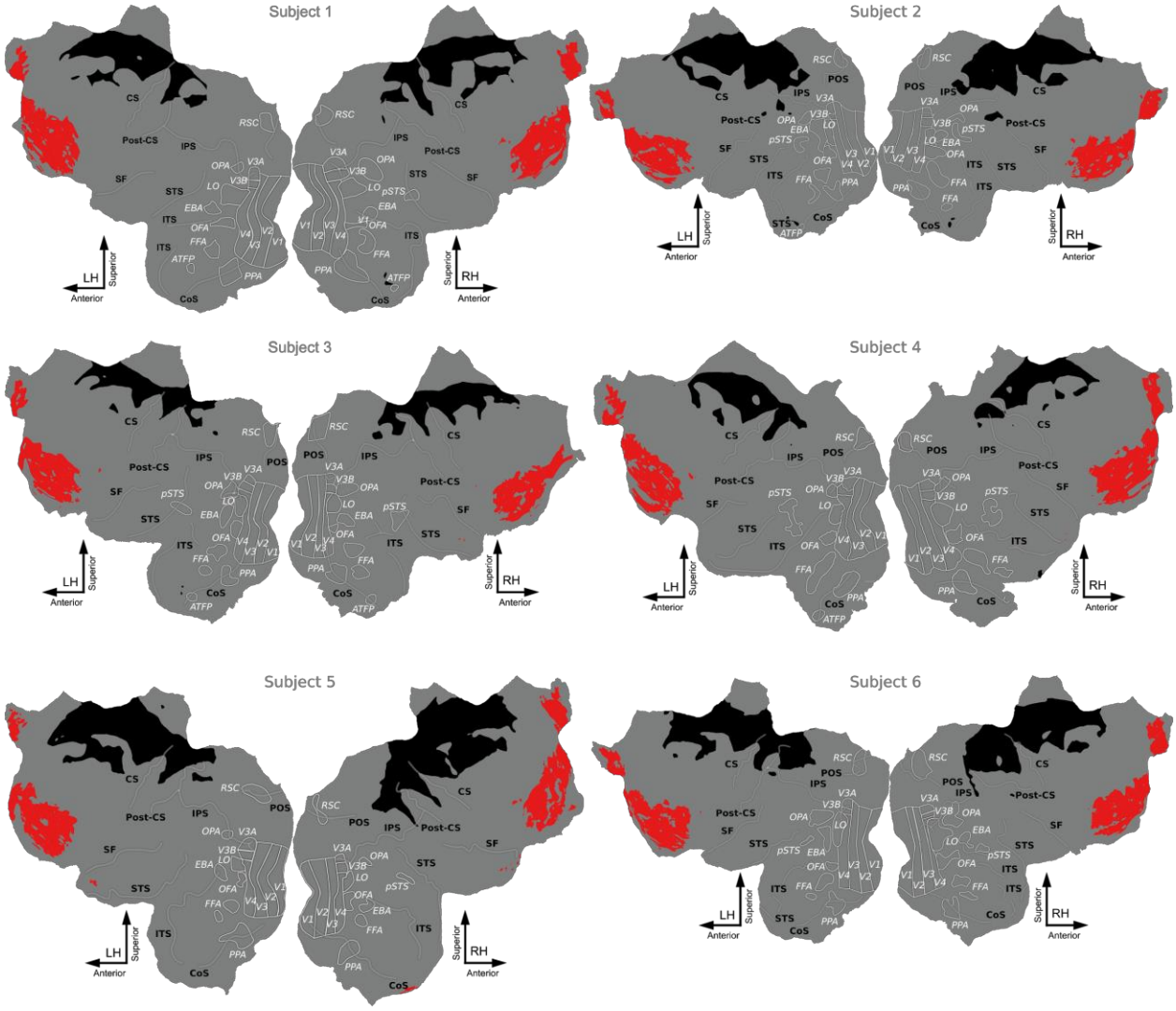


**Figure 3-6. Orbitofrontal Cortex (OFC) ROI Projected onto the Cortical Surface.**
A flatmap showing voxels included in the OFC ROI in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

7. non-OFC Frontal Cortex: This ROI was defined using the following 5 AAL template structural ROIs: superior, middle, and superior medial frontal regions, as well as frontal inferior operculum and frontal inferior triangularis. These ROIs were back-projected into subject functional space in the same manner as the OFC ROI. See Figure 3-7 below for flatmaps showing the voxels included in this ROI.

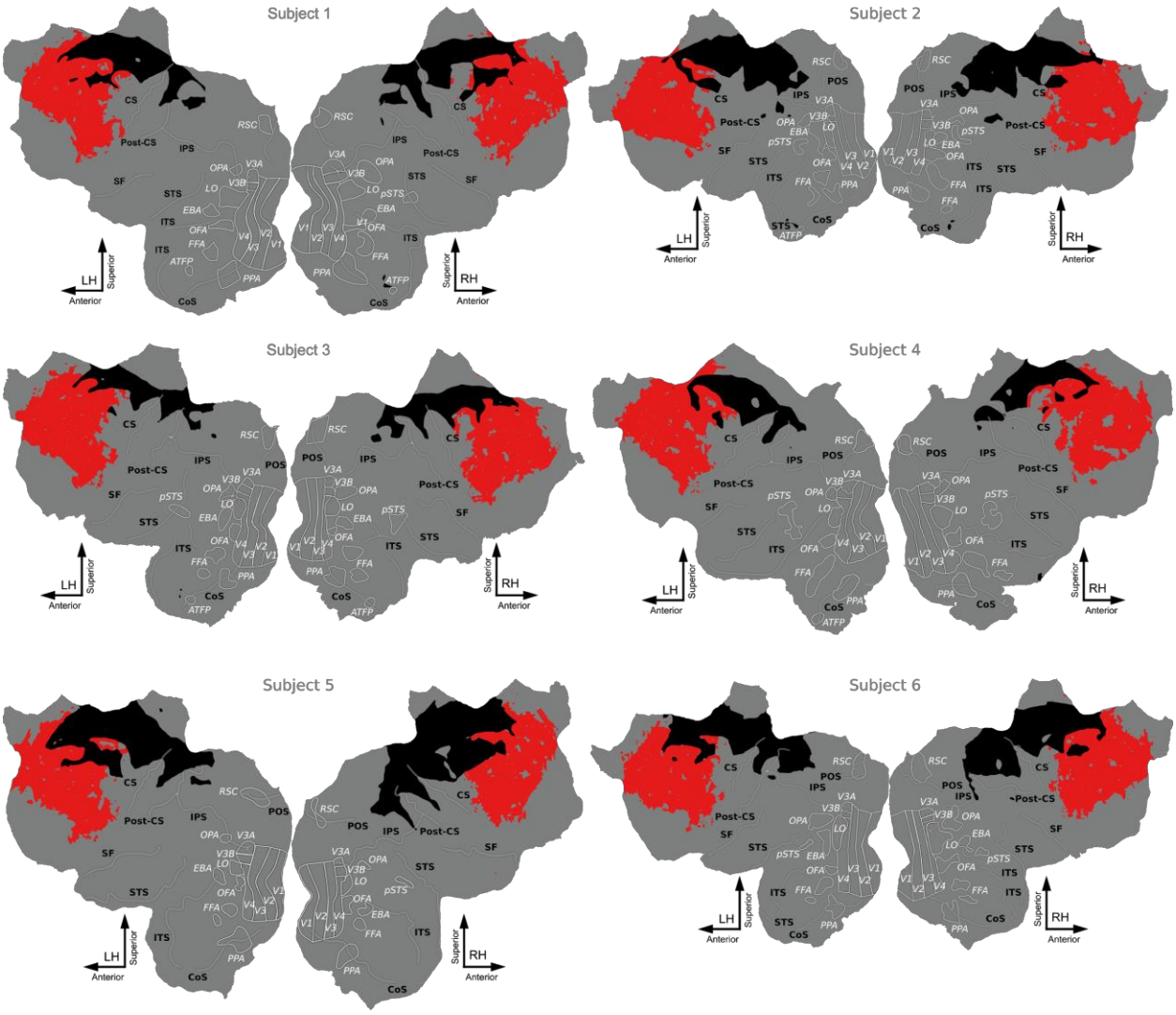


**Figure 3-7. non-OFC Frontal Cortex ROI Projected onto the Cortical Surface.**
A flatmap showing voxels included in the Frontal Lobe ROI, that were not within the OFC ROI or within motor cortex, in red for all six subjects. Voxels in grey were excluded from the ROI, and voxels in black were outside of the MRI slice prescription, and not collected.

## Representational Similarity Analysis (RSA)

A representational similarity analysis (RSA) was conducted separately for each subject using BOLD data collected during the thirty scans (7.5 minute each) collected for model estimation for the study outlined in chapter 2. Across these thirty runs, a large set of naturalistic

emotional images (n=1440) spanning 21 semantic categories and six valence (negative, neutral, positive) by arousal (high, low) conditions were presented for one second, with a three second inter-stimulus interval (ISI). RSA is done by creating multiple representational dissimilarity matrices (RDMs) of both brain activity (called brain RDMs) and stimulus feature model predictions (called model RDMs), and then quantifying the pairwise similarity between a single brain RDM and a single model RDM, possibly controlling for the variance explained by other model RDMs (Kriegeskorte, Mur & Bandettini, 2008; Chikazoe et al., 2014; Diedrichsen & Kriegeskorte, 2017). To create the brain RDMs, preprocessing of the BOLD data was first conducted using SPM 8, which comprised slice time correction, realignment, and linear detrending. The data was also coregistered to the T1 anatomical image for the searchlight analyses using spatial coregistration in SPM 8, as this analysis was done in anatomical space on the cortical surface. High-pass filtering was not done (beyond linear detrending) to avoid the accidental removal of low-frequency stimulus signals resulting from only 2 presentations of each stimulus. Spatial smoothing was also not done as is customary with RSA (Diedrichsen & Kriegeskorte, 2017). SPM 8 was then used to estimate a t-contrast across all brain voxels for each of the 1440 estimation stimuli, which quantified each voxel's average response to each stimulus relative to baseline. Each unique image was assigned an event, and the resulting t-contrast for that event was used as the response amplitudes for the brain RDMs. Nuisance regressors accounting for movement were also included in the model, as well as a constant bias term per run to account for differences in baseline BOLD activity between runs. Voxel selection within SPM was done by providing an explicit mask which was the same voxel-selection mask as was used for the VWM analyses in chapter 2. Additionally, implicit masking was disabled by setting the threshold to negative infinity (the SPM default is 80%), which tells SPM not to discard any voxels based on mean signal relative to the global mean signal. All RSAs were done using custom Matlab scripts that utilized the CosMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016).

RSA was then conducted both on specific ROIs, and across all grey-matter cortical voxels using a searchlight procedure. The ROI brain RDMs were created using the above defined ROIs for voxel selection. For each ROI, stimulus pairwise Pearson's correlations between the response amplitudes (t-contrast values) of all selected voxels were estimated between all 1440 estimation stimuli, resulting in a single brain RDM per ROI. As this experiment was not designed with RSA in mind, we checked the magnitude of correlations between adjacent images within the design matrix. The mean correlation of design matrix columns across all images was 0.01, while that for adjacent images was 0.2 (s.d.=0.34, min=0, max=0.5). This showed that the response amplitudes would be estimable and decently powered. All ROI brain RDMs were created using the CosMoMVPA function "cosmo_target_dsm_corr_measure". The stimulus set used in this experiment consisted of 1440 images, which resulted in a brain RDM that is larger than most monitors can display. Thus, to visualize the brain RDM we averaged the brain RDMs per ROI using both the animate by valence by arousal features used for the model RDMs, which is more easily interpretable, and the semantic by valence by arousal features used in the CSVA

model described in chapter 2. For each ROI, and the whole brain, a brain RDMs was calculated by first averaging response amplitudes of all stimuli within the same animate (or semantic) by valence by arousal condition for each subject individually. Brain RDMs were then created from these averaged response amplitudes, the RDM values were then Fisher transformed, a group averaged brain RDM was created by averaging across subjects, and the value were reverse Fisher transformed for visualization. The resulting matrix was then ordered by the rows/columns of the RDM first by whether the stimulus associated with the row/column was animate or inanimate (or the semantic category), then by negative, neutral or positive (valence), and finally by low and high arousal.

In addition to RSA within our 7 ROIs, a searchlight procedure was used that selected the 100 closest grey-matter voxels to each cortical voxel, and created a "searchlight" brain RDM using these 100 voxels. The 100 closest voxels were determined on the cortical surface, using FreeSurfer generated cortical surfaces. To determine the 100 closest voxels in cortical surface space the "cosmo_surficial_neighborhood" function was used. The relevant neighborhood function just described was then passed in to "cosmo_searchlight" function, using the "cosmo_target_dsm_corr_measure" function as the measure of similarity between model and brain RDMs. This measure function can be used to either calculate full correlations between brain and model RDMs (Spearman rank-order correlation was used) or to estimate partial correlations (partialling out variance in the brain RDM accounted for by other model RDMs before calculating a pairwise Spearman correlation between brain and model RDMs). The result of the RSA for each "searchlight" brain RDM is a correlation value for pairwise partial correlations [e.g. with variance explainable by Gabor wavelet model covaried out], assigned to the central cortical voxel of the current "searchlight", and the analysis repeated across all cortical voxels.

Creation of the model RDMs was done using custom Matlab scripts only, as no RSA toolboxes the author is aware of offer this functionality. The models for these RDMs consisted of the animacy, valence, arousal and gabor features used to fit the voxel-wise models in chapter 2. The following model RDMs were created, using the distance metrics specified:

1. Gabors - Cosine distance in the 426 feature gabor feature space
2. Animacy 4 level - Euclidean in the 4 feature space
3. Animacy 2 level - Binary
4. Semantics - Binary across the 21 semantic categories used in the CSVA model.
5. Valence 1-Dimensional - Euclidean in the 3 dimensional positive, neutral, negative space.
6. Arousal - Binary between high & low
7. Valence by Arousal 2D - Euclidean in the two dimensional Valence by Arousal space. This results in 6 points in the 2-D space: high-negative, low-negative, high-neutral, low-neutral, high-positive, and low-positive. There is a distance of 1 between each valence level (thus a difference of 2 between negative and positive), and a distance of 1 between both arousal levels.

8. Animacy by Valence by Arousal - Euclidean distance in 3-D animacy (2-Level), valence and arousal spaces, as defined above in the Animacy 2-Level and Valence, and Arousal models. This results in 12 points in the 3-D space: animate-high-negative, animate-low-negative, animate-high-neutral, animate-low-neutral, animate-high-positive, animate-low-positive, inanimate-high-negative, inanimate-low-negative, inanimate-high-neutral, inanimate-low-neutral, inanimate-high-positive, and inanimate-low-positive. There is a distance of 1 between animate and inanimate, a distance of 1 between each valence level (thus a difference of 2 between negative and positive), and a distance of 1 between both arousal levels.
9. Animacy by Valence - Euclidean distance in 2-D animacy (2-level) by valence (3-level) space. This results in 6 points in the 2-D space: animate-negative, animate-neutral, animate-positive, inanimate-negative, inanimate-neutral, and inanimate-positive. There is a distance of 1 between animate and inanimate and a distance of 1 between each valence level (thus a difference of 2 between negative and positive).
10. Animacy by Arousal - Euclidean distance in 2-D animacy (2-level) by arousal (2-level) space. This results in 4 points in the 2-D space: animate-high, animate-low, inanimate-high, and inanimate-low. There is a distance of 1 between animate and inanimate, and a distance of 1 between high and low arousal.
11. Animate Only Valence by Arousal - Custom Distance: 1 between all 6 Animate by Valence by Arousal categories and the single inanimate category. Then either the 1D or 2D Valence by Arousal Distances specified above.
12. Animate Only Valence - Custom Distance: 1 between all 3 animate x Valence categories and the single inanimate category. Then 1 between animate negative and neutral and animate neutral and positive, and 2 between animate negative and positive.
13. Animate Only Arousal - Binary
14. Inanimate Only Valence by Arousal - Custom Distance: 1 between all 6 Inanimate by Valence by Arousal categories and the single animate category. Then either the 1D or 2D Valence by Arousal Distances specified above.
15. Inanimate Only Valence - Custom Distance: 1 between all 3 Inanimate by Valence categories and the single animate category. Then 1 between inanimate negative and neutral and inanimate neutral and positive, and 2 between inanimate negative and positive.
16. Inanimate Only Arousal - Binary

The ROI RSAs were then calculated by estimating Spearman rank-order correlations between each ROI brain RDM and each model RDM. **This resulted in a *r* x *m* matrix of full correlations between ROI brain RDMs and model RDMs,** where ***r* is the number of ROIs** and ***m* is the number of models tested**. Correlations between model feature spaces were low (Fig. 3-8). However, to be confident in our interpretation of model to brain RDM relationships, we also conducted partial Spearman correlations to control for shared variance among the feature

spaces. Partial Spearman correlation was done using the "cosmo_target_dsm_corr_measure" function from the CosMoMVPA toolbox. This function implements partial Spearman correlation by taking as inputs: (a) the list of voxels (ROI) to create a ROI brain RDM, (b) the final model RDM, and (c) a list of control model RDMs that will be partialed out. As Spearman correlation (and partial correlation) is simply Pearson's correlation on the rank-ordering of two identical length vectors, all of these RDMs are first converted into rank-orders, and then separate multiple linear regressions are fit to the rank-ordered ROI brain RDM and the rank-ordered final model RDM, using the rank-ordered control model RDMs as the independent variables in both regressions. The residuals from both regressions, which are now in rank-order, are then subjected to a Pearson correlation, resulting in the partial Spearman correlation. For the ROI RSA, partial correlations were calculated for each model using two sets of control model RDMs. First, only the Gabors model RDM was partialed out of all models (except the Gabors model itself). Second, the below list details the "complete" set of control model RDMs used in the partial Spearman correlation between each of the final model RDMs and each ROI brain RDM.

1. Animacy 4-level - Gabors, Valence 1D, & Arousal
2. Animacy 2-level - Gabors, Valence 1D, & Arousal
3. Semantics - Gabors, Valence 1D, & Arousal
4. Valence 1D - Gabors, Animacy 2-level, & Arousal
5. Arousal - Gabors, Animacy 2-level, & Valence
6. Valence by Arousal 2D - Gabors, & Animacy 2-level
7. Animacy by Valence by Arousal - Gabors
8. Animacy by Valence - Gabors, & Arousal
9. Animacy x Arousal - Gabors, & Valence
10. Animate Only Valence by Arousal - Gabors, Animacy 2-level
11. Animate Only Valence - Gabors, Animacy 2-level, & Arousal
12. Animate Only Arousal - Gabors, Animacy 2-level, & Valence
13. Inanimate Only Valence by Arousal - Gabors, Animacy 2-level
14. Inanimate Only Valence - Gabors, Animacy 2-level, & Arousal
15. Inanimate Only Arousal - Gabors, Animacy 2-level, & Valence

The single subject RSA results for all RSAs by models were then Fisher transformed, averaged across all subjects, and reverse Fisher transformed for visualization of group results. This was done on the full correlations, the partial correlations of the Gabors model, and the "complete" partial correlations described in the above list. To calculate significance of these group RSA results, the subject-specific correlation values were Fisher transformed and subjected to one-tailed t-tests across subjects. These results were corrected for multiple comparisons using a Bonferroni correction across all r ROIs x m models.

The searchlight RSAs were done in an analogous manner to the ROI RSAs, on a subset of the models tested in the ROI RSA, just across all cortical voxels, as explained above. Partial

Spearman correlations were calculated for the following 5 model RDMs, using the "complete" set of control model RDMs described in the list above:

1. Animacy 2-level
2. Animate Only Valence
3. Animate Only Arousal
4. Inanimate Only Valence
5. Inanimate Only Arousal

Additionally, a partial Spearman correlation searchlight RSA was done on the Animacy 2-level model RDM using only the Gabors model RDM. These searchlight analyses resulted in 6 cortical maps per subject, one for every RSA done. To determine significance of these searchlight correlation values, the correlation coefficients were convert to t-statics using the equation: $t = t * tttt \left(\frac{(t-2)}{(1-t^2)}\right)$, where t is the t-statistic calculated, r is the correlation coefficient, and n is the number of samples in the vectors being correlated. The Students-t cumulative distribution function (cdf) was then used to determine the p-values for all voxels. A Bonferroni family-wise error correction for multiple comparisons was then done across all cortical values using a threshold of p=.05. Only those voxels that survived the Bonferroni correction had their correlation coefficient values displayed on the flatmaps. We note however, that with 1440 stimuli used in this analysis, the resulting RDMs had 1,036,080 entries. As such, a correlation value of just 0.0017 was significant ($p<.05$, uncorrected), which resulted in most of cortex being significant for our 5 models tested. Additionally, permutation tests of several sample correlations were done using 1000 permutations, which resulted in a maximal null-distribution value of ~0.002 which indicates that a p-value of 0.001 has an r-value of ~0.002. Results from the permutation tests were in line with the results from the t-statistic conversion method to calculate significance, thus the t-statistic conversion was used to calculate significance as it is much more computationally efficient. In order to show more meaningful maps, we also showed flatmaps where the r-values were thresholded to only display those with a correlation above 0.05. This number was determined by finding the smallest r-value that would be significant if the calculation of significance had used an n=1440, the number of independent stimuli in the experiment.

**Table 3-1: Description of methods used in selection of higher visual neuroscience studies using RSA.**

| Paper | Models Tested | Brain RDM Distance Metric | Model RDM Distance Metric | Pairwise Correlation Metric | Partial Correlation /Stepwise Regression | GLM Models Used |
|---|---|---|---|---|---|---|
| 1 | 18 structural, animacy, or face models | Pearson | Pearson | Spearman | n/a | n/a |
| 2 | Low-level, Animacy, Valence | Pearson | Low-level - Euclidean, Animacy - Non-subject 1-17 ratings, Valence - Subject 1-7 positive & 1-7 negative ratings | n/a | Stepwise Regression - Two models fit in single GLM, then Spearman Correlation on residuals using remaining model. Done on all 3 models. | Yes, with all 3 models. In ROIs and searchlight. |
| 3 | Bodies vs. Faces vs. Places vs. Objects | Pearson | Subject image similarity judgments | Pearson | n/a | n/a |
| 4 | Low-level (LL), 6 Object Category (OC) models, Semantic Features (SF) | Pearson | LL - Pearson, OC - Binary, SF - Cosine angle | Spearman | Partial Spearman's Correlation , partial out 8 models then find partial correlation on the remaining 9th model. Done for all models. | n/a |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 | Faces vs. bodies vs. Natural Inanimate vs. Artificial Inanimate | Pearson | Pearson (for MEG brain RDM) | Spearman to compare fMRI and MEG RDMs | n/a | n/a |
| 6 | Object Size, Object Category, Spatial GIST Descriptors (Gabor wavelets), Spectral GIST Descriptors | Leave-one-participant-out (LOPO) cross-validation of Correlation | Binary | n/a | n/a | GLM on Object Size & Category, GLM on GIST features, |

Numbers in column 1 refer to the following papers:

1. Kriegeskorte et al., 2008
2. Chikazoe et al., 2014
3. Charest et al., 2014
4. Clarke & Tyler, (2014)
5. Cichy, Pantazis, & Oliva, (2014)
6. Watson, Young, & Andrews, (2016).

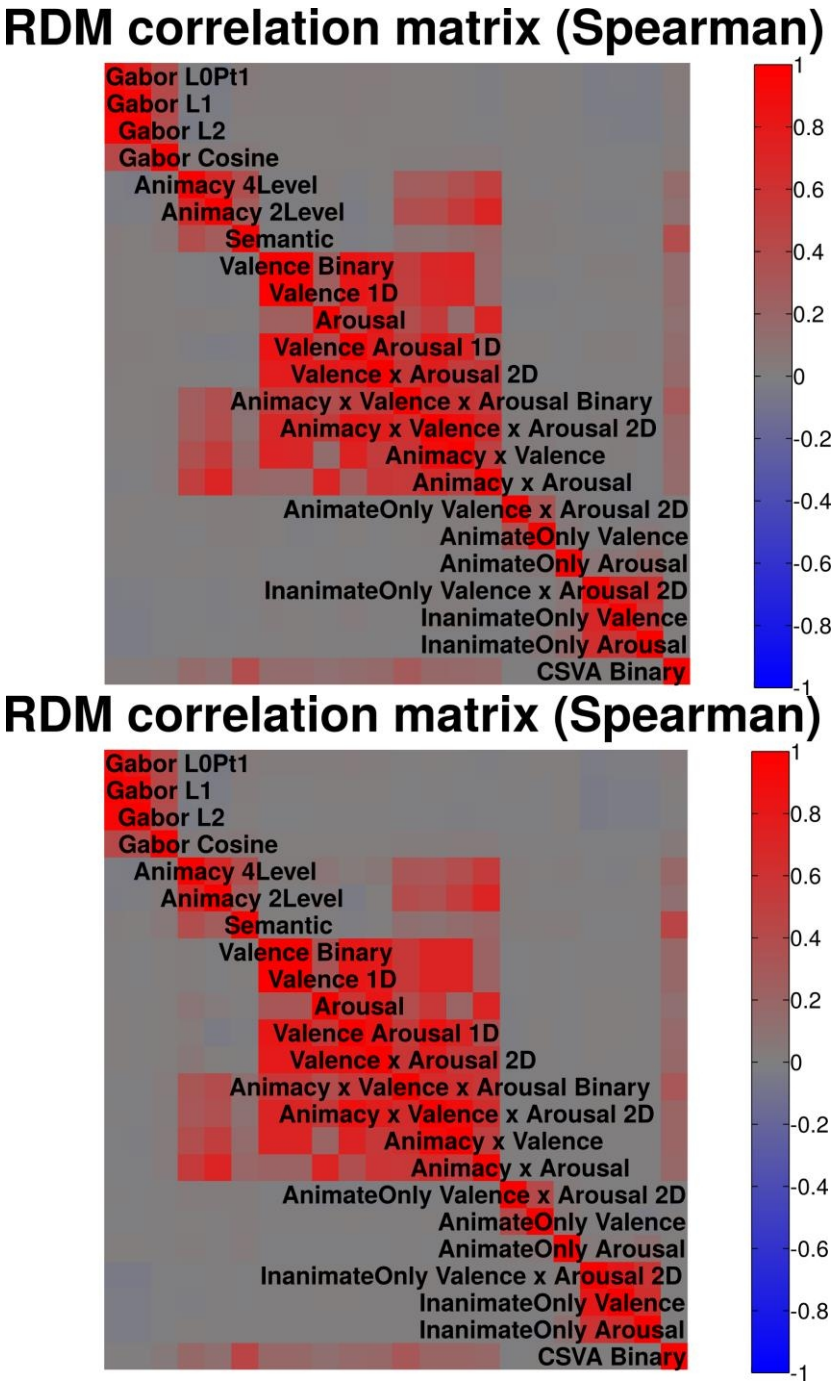


**Figure 3-8. Spearman Correlations between all model RDMs for Two Sample Subjects.**
Correlation matrix showing Spearman rank correlations between the representational dissimilarity matrices (RDMs) for all 16 models evaluated in this analysis for subject 1 (top) and subject 2 (bottom). Correlations between the Gabor Model and all other models was very low. Correlation between the 2 and 4 level animacy, and semantics models were high. Models containing valence, arousal, or both across all stimuli all were moderately to highly correlated, but were not highly correlated with the valence, arousal and valence by arousal models for only animate or inanimate stimuli.

## Results

Representational similarity across stimulus conditions of animacy by valence by arousal was quantified for each ROI by correlating activity patterns to these stimulus conditions across all voxels from each region, creating a representational dissimilarity matrix (RDM) per ROI. Figure 3-9 presents group-averaged RDM plots of all seven ROIs, plus an RDM for the whole brain for reference. Across all RDMs plotted, an animate vs. inanimate divide can be seen, as has been reported previously for VTC and OFC (Kriegeskorte et al, 2008; Chikazoe et al, 2014). This divide is more striking within OTC and VTC ROIs however, and even slightly more so when excluding EVC voxels from both OTC and VTC. Furthermore, animate high arousal stimuli are more dissimilar to other high arousal animate stimuli than they are to low arousal stimuli within the two VTC and two OTC ROIs, as are inanimate stimuli, although perhaps to a lesser extent. This effect is also present within OFC, the non-OFC Frontal ROI and EVC, although to a lesser extent. Additionally, positive low and high arousal stimuli show decreased dissimilarity to each other relative to other valence categories within the two VTC and OTC ROIs. Figures 3-10 - 3-17 show brain RDMs averaged, and ordered by, the 126 feature semantic by valence by arousal model (as described in chapter 2) for all seven ROIs and the whole brain. While a similar pattern emerges here as was seen in the animacy by valence by arousal RDMs (Fig. 3-9), a more nuanced differentiation can be seen between the semantic categories. For example, looking at dissimilarity between the six valence by arousal face categories within the non-EVC VTC ROI, we can see clear similarity between negative and positive high arousal stimuli, as well as between low and high arousal positive stimuli. This pattern was not present for body-parts however. In order to quantify these effects, RSA was conducted on each ROI by comparing theoretical model RDMs with brain RDMs from each region.

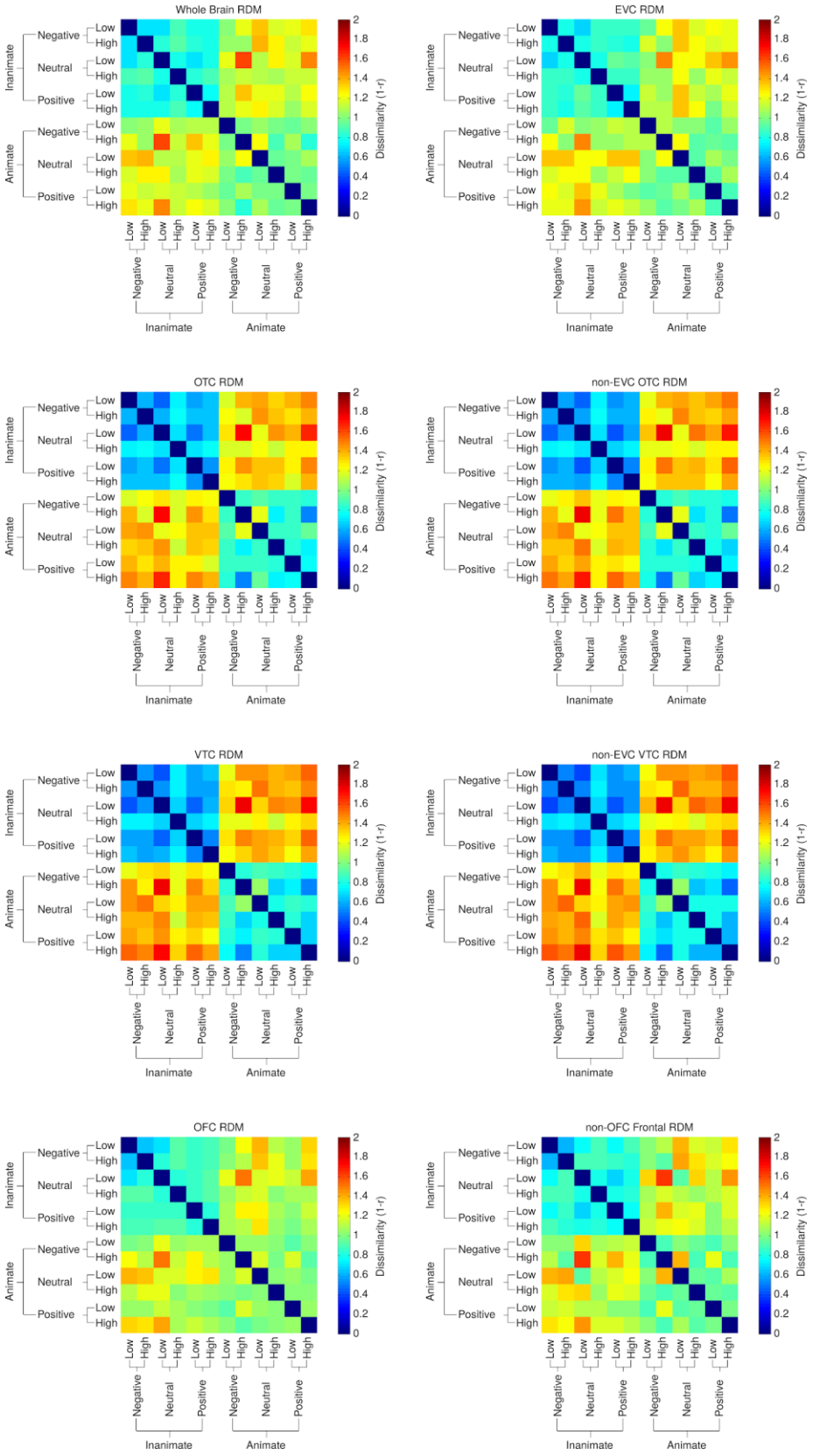
Whole Brain RDM
EVC RDM
OTC RDM
non-EVC OTC RDM
VTC RDM
non-EVC VTC RDM
OFC RDM
non-OFC Frontal RDM
Inanimate
Animate
Negative
Neutral
Positive
Low
High
Dissimilarity (1-r)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

**Figure 3-9. Group-Averaged Brain RDMs showing Similarity of Stimulus Representation Within ROIs.**

Matrices plotting group-averaged dissimilarity of stimulus representation across all voxels within seven ROIs, and the whole brain. Voxel stimulus representations were averaged across animacy by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. As has been reported previously (Kriegeskorte et al., 2008), across all seven ROIs, and the whole brain, an animate vs. inanimate divide can be seen, although it is much more prominent within the two VTC and two OTC ROIs, as would be expected from previous findings. Additionally, within the two VTC and two OTC ROIs, high arousal stimuli are less dissimilar to each other than they are to low arousal stimuli. This pattern is also present within OFC and non-OFC Frontal ROIs, although to a less extent. Finally, within the two VTC and OTC ROIs, animate positive low and high arousal stimuli show decreased dissimilarity to each other relative to other valence categories, suggesting animate positive stimuli may have a common similarity structure within these regions.

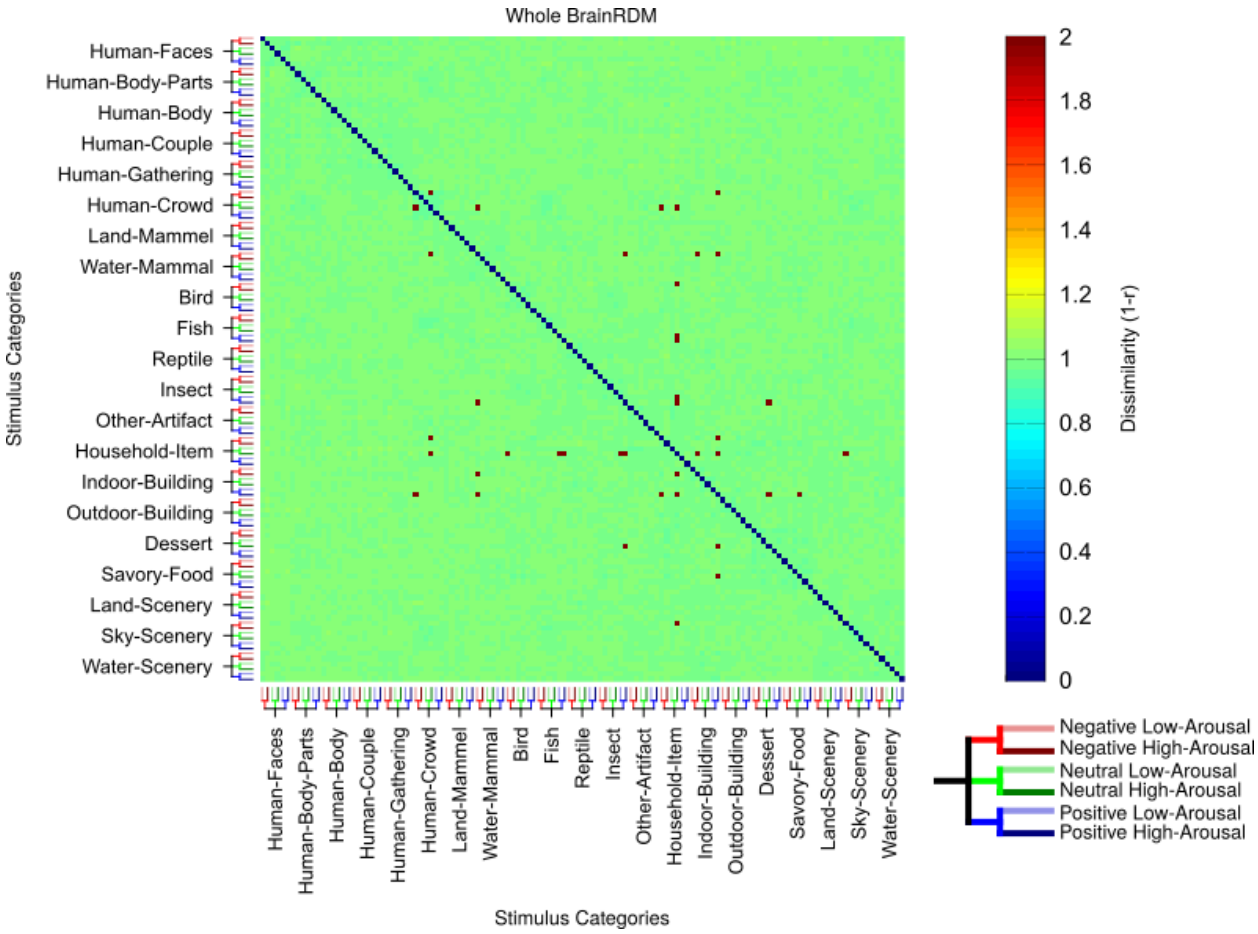


**Figure 3-10. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation in the Whole Brain.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the whole brain. Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. No clear patterns of similarity can be seen across the whole brain.

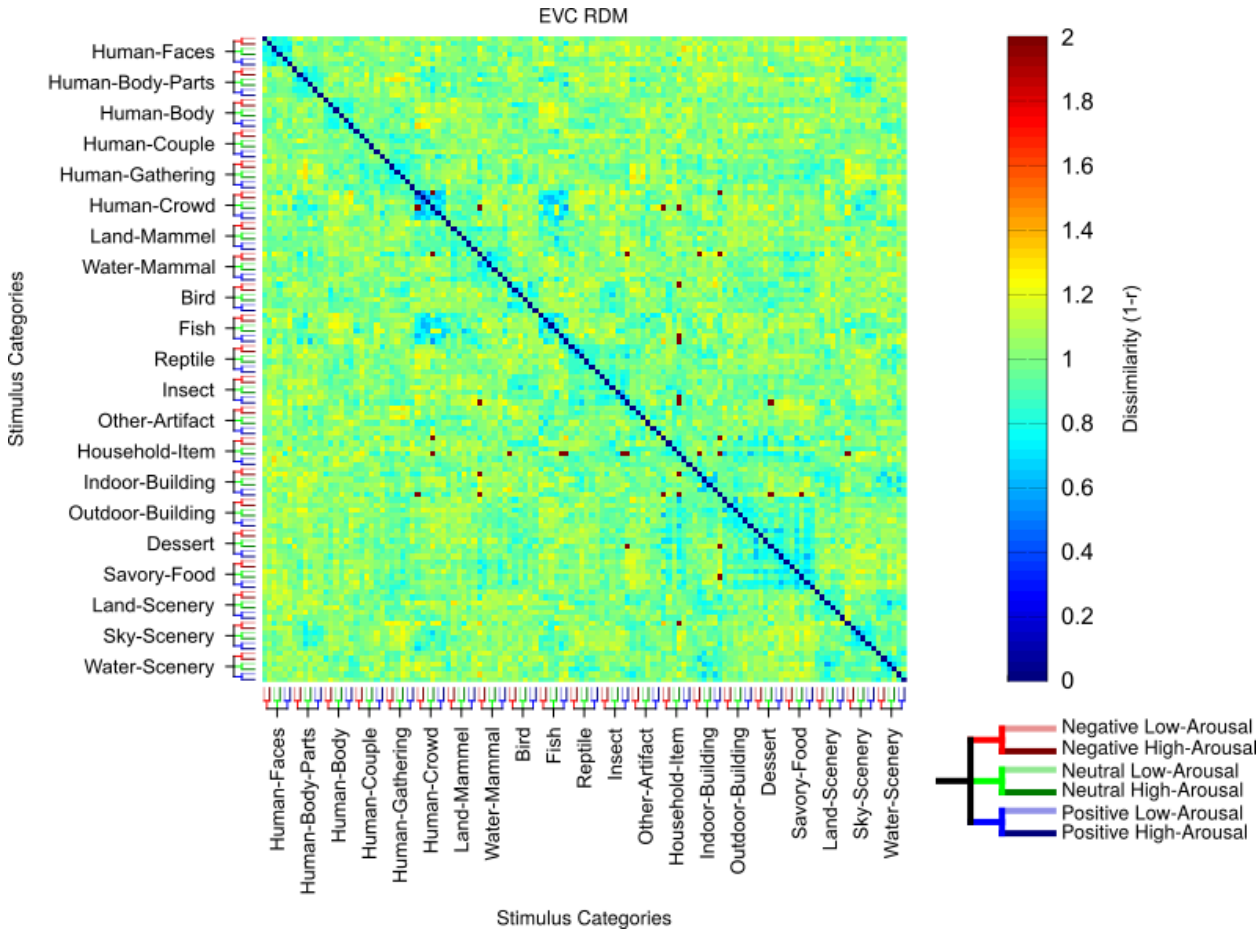


**Figure 3-11. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the EVC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the early visual cortex (EVC) ROI (see Fig. 3-1 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. While clear similarity exists within similar semantic categories (e.g. all face stimuli are similar), there is not a high level of differentiation between valence by arousal categories.

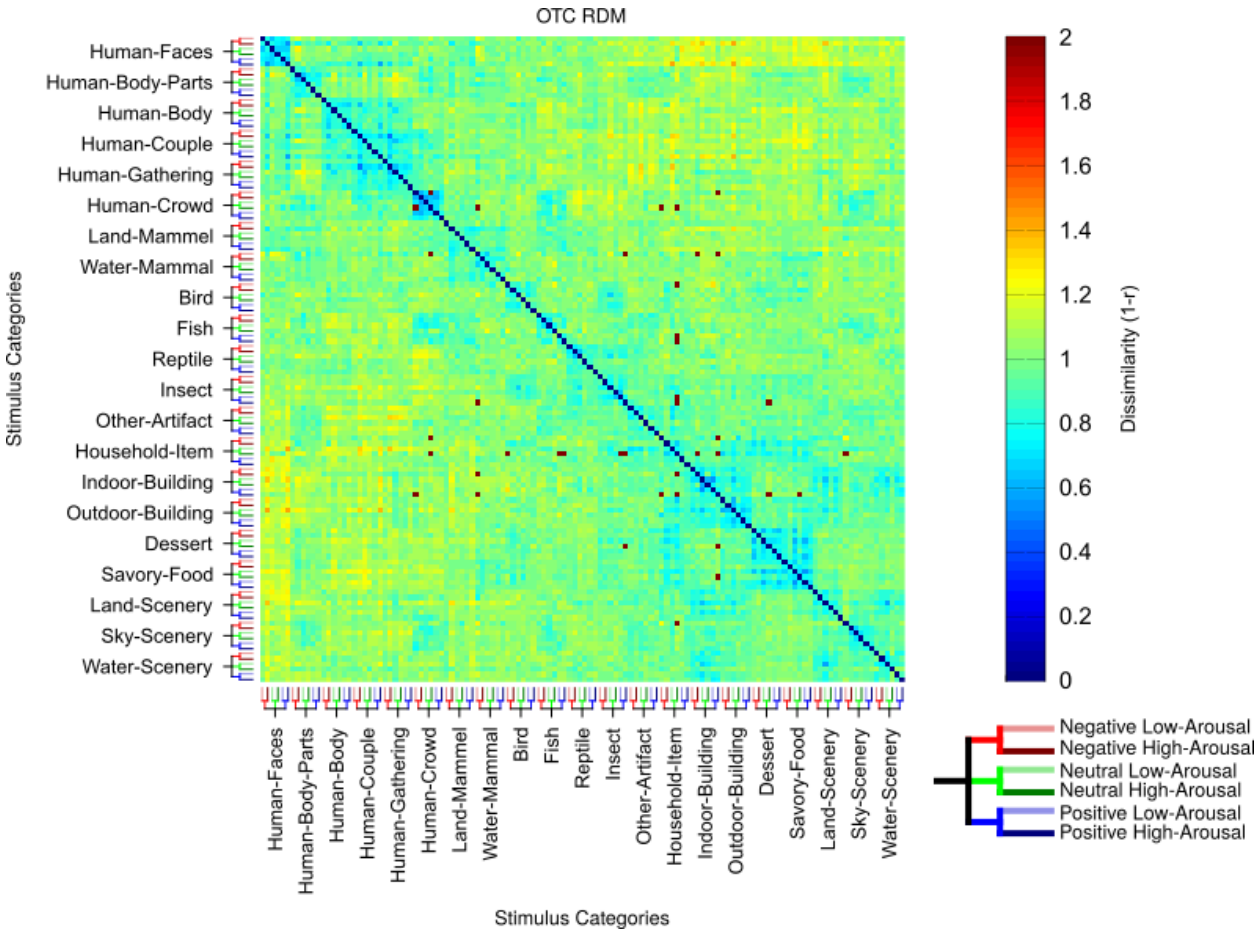


**Figure 3-12. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the OTC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the occipital-temporal cortex (OTC) ROI (see Fig. 3-2 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. Clear similarity exists within similar semantic categories (e.g. all face stimuli are similar). Additionally, interesting patterns of similarity between high arousal negative and positive stimuli exist within most of the human semantic categories as well as buildings and food.

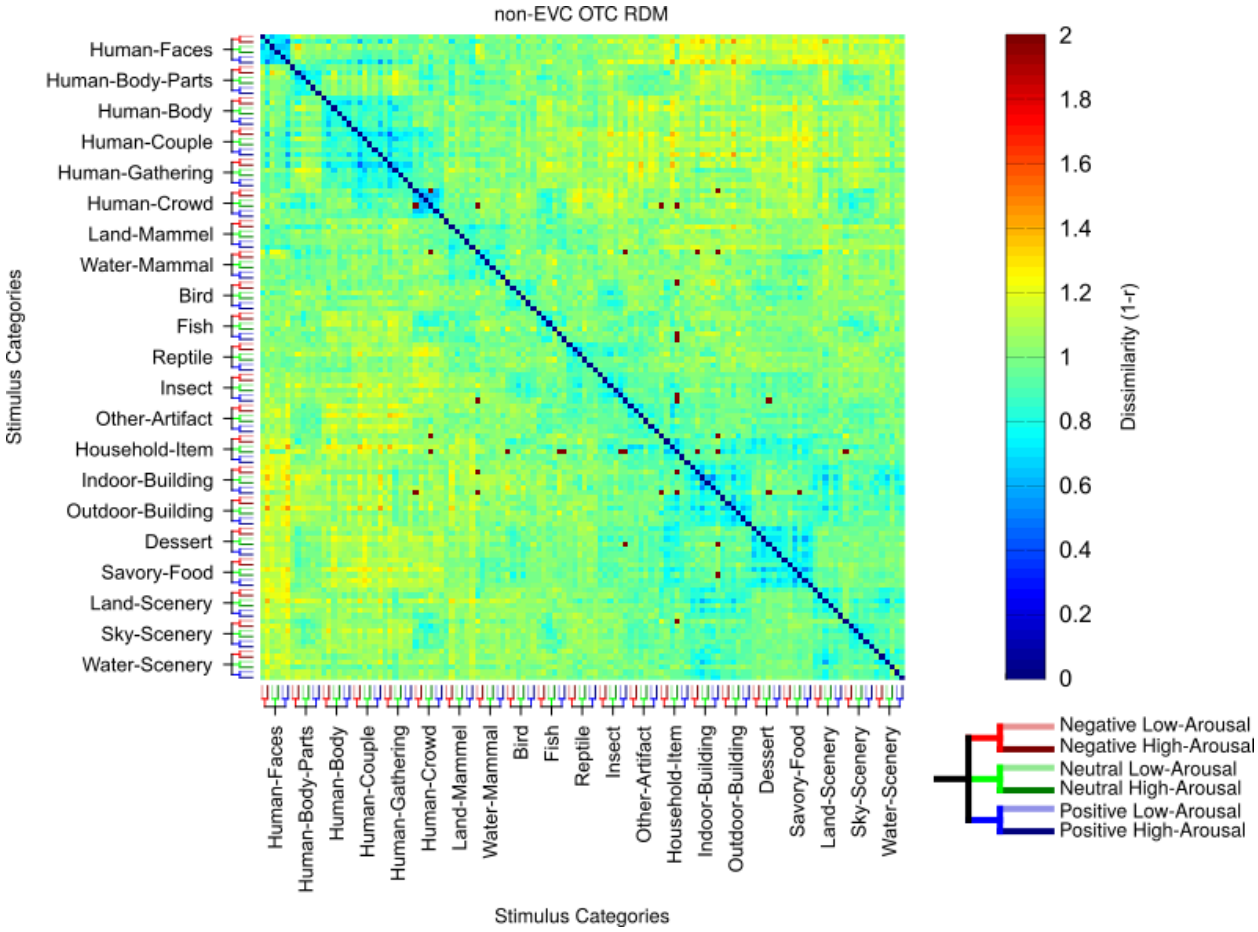


**Figure 3-13. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the non-EVC OTC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the non-EVC OTC ROI (see Fig. 3-3 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. Clear similarity exists within similar semantic categories (e.g. all face stimuli are similar). Additionally, interesting patterns of similarity between high arousal negative and positive stimuli exist within most of the human semantic categories as well as buildings and food. Furthermore, similarity between positive faces is observed.

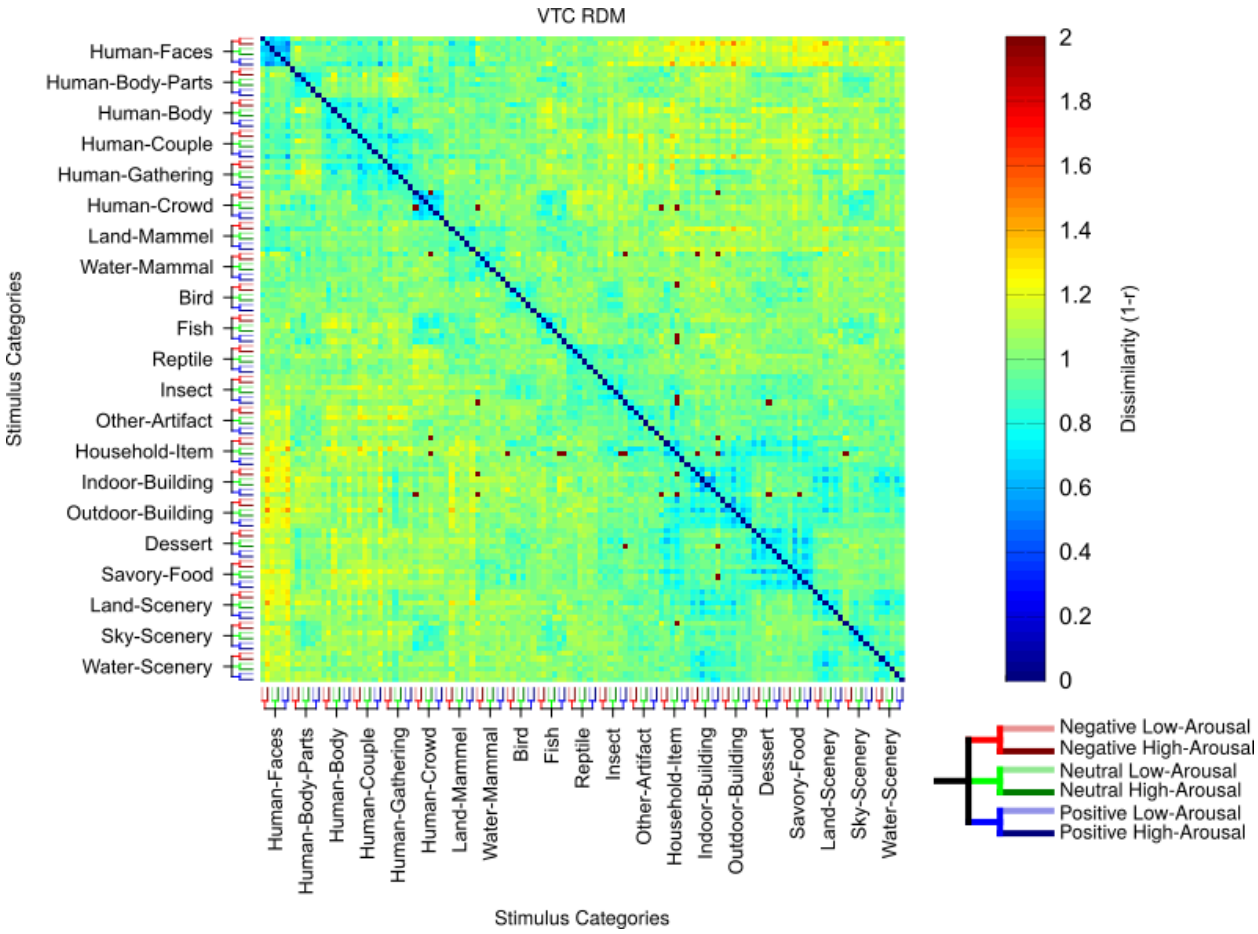


**Figure 3-14. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the VTC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the ventral-temporal cortex (VTC) ROI (see Fig. 3-4 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. Clear similarity exists within similar semantic categories (e.g. all face stimuli are similar). Additionally, interesting patterns of similarity between high arousal negative and positive stimuli exist within most of the human semantic categories as well as buildings and food. Furthermore, similarity between positive faces is observed.

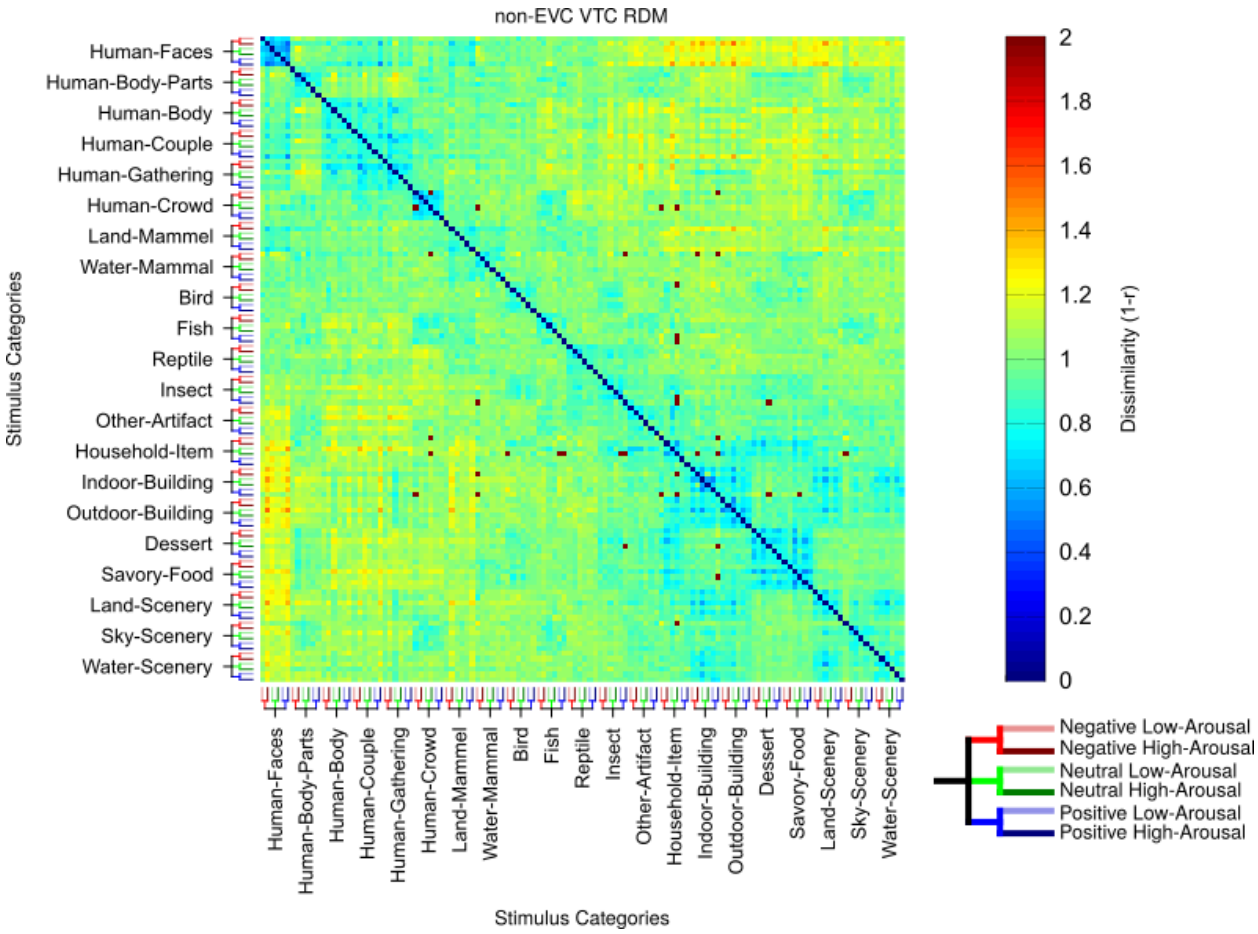


**Figure 3-15. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the non-EVC VTC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the non-VTC OTC ROI (see Fig. 3-5 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. Clear similarity exists within similar semantic categories (e.g. all face stimuli are similar). Additionally, interesting patterns of similarity between high arousal negative and positive stimuli exist within most of the human semantic categories as well as buildings and food. Furthermore, similarity between positive faces is observed.

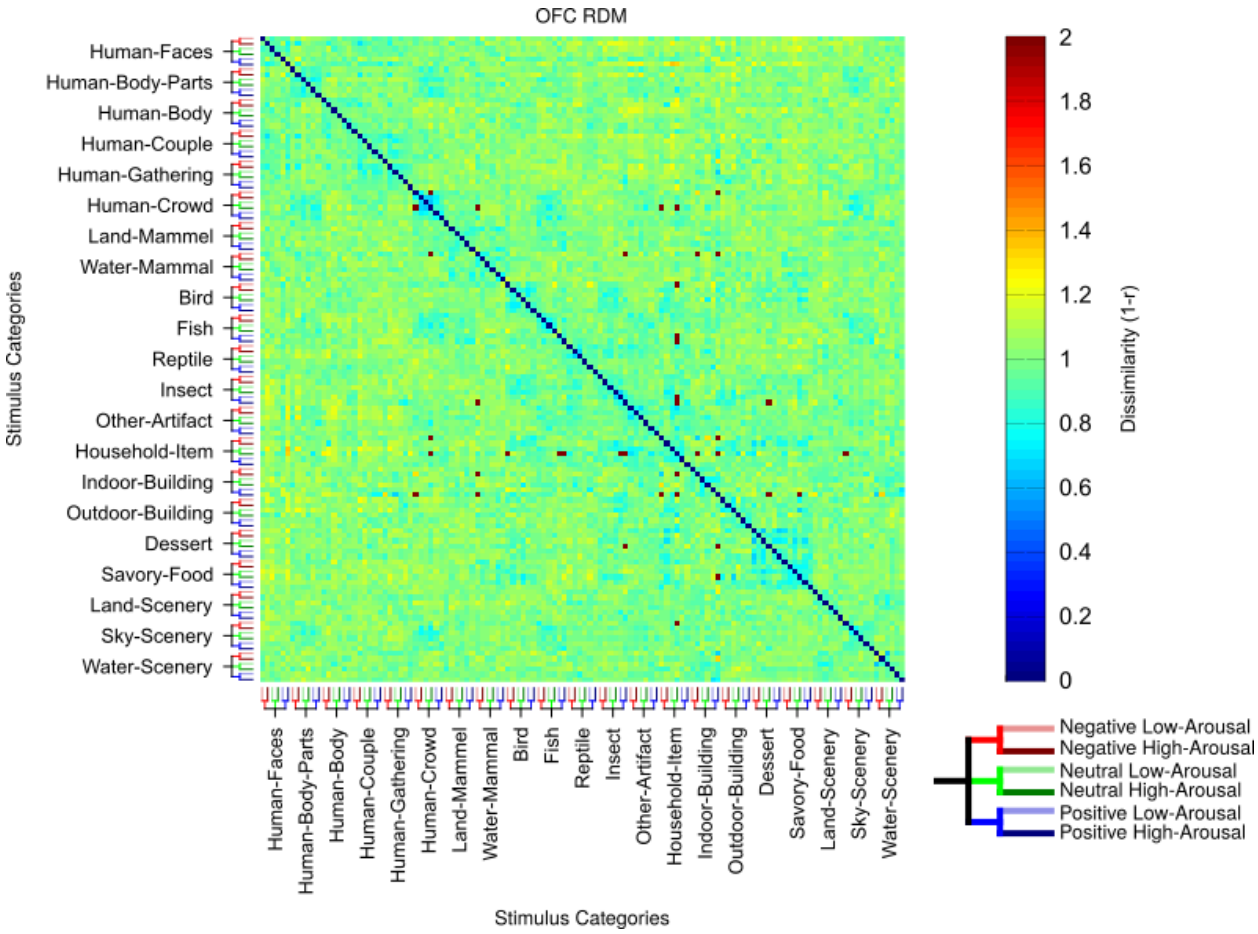


**Figure 3-16. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the OFC ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the orbitofrontal cortex (OFC) ROI (see Fig. 3-6 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. While clear similarity exists within similar semantic categories (e.g. all face stimuli are similar), there is not a high level of differentiation between valence by arousal categories.

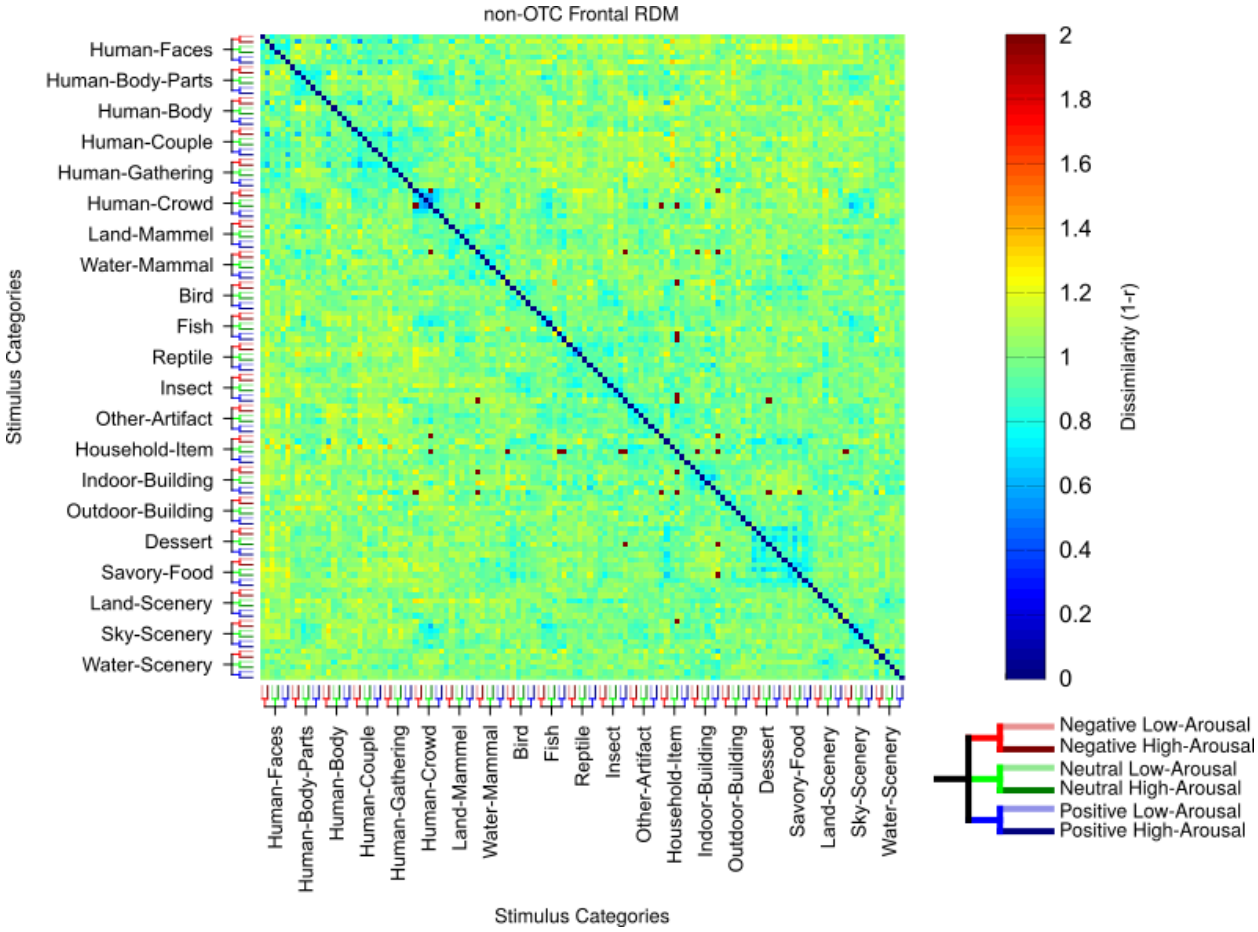


**Figure 3-17. Group-Averaged Brain RDMs showing Similarity of Semantic by Valence by Arousal Category Representation Within the non-OFC Frontal ROI.**

Matrix plotting group-averaged dissimilarity of stimulus representation across all voxels within the non-OFC Frontal ROI (see Fig. 3-7 for definition). Voxel stimulus representations were averaged across 126 semantic by valence by arousal conditions per subject, and all pairwise linear correlations of these stimulus representations for each condition were averaged across subjects. Each of the 21 semantic categories is labeled, while the six valence by arousal categories are represented by a colored tree, where red hues represent negative, green hues represent neutral and blue hues represent positive and saturation represents low vs. high arousal (as depicted in the figure key). Higher dissimilarity values are represented by hotter (yellow-red) colors and lower dissimilarity values are represented colder (blue-green) colors. Pixels colored with bright red were missing across all six subjects. A higher dissimilarity value is the result of lower Pearson's correlation between the multivoxel pattern of activity of two stimulus conditions, across the given region, and vice versa for lower dissimilarity values. While clear similarity exists within similar semantic categories (e.g. all face stimuli are similar), there is not a high level of differentiation between valence by arousal categories.

Theoretical models consisting of low-level features (Gabor wavelets), semantics, animacy, valence, arousal, and interactions between animacy, valence and arousal were specified for each of the 1440 stimuli used in model estimation of the VWM analysis in chapter 2, resulting in 16 theoretical models of interest (see RSA methods section for model details). RDMs were then calculated for each of these 16 models of interest (see RSA methods section for

distance metrics used for each model RDM). RSA was then conducted by correlating (Spearman rank correlation) each brain ROI RDM with each model RDM, resulting in an r (# of ROIs) by m (# of models) correlation matrix. Figure 3-18 shows a plot of this matrix for the full pairwise correlations of ROI brain RDMs and model RDMs. These results showed that the 2-level Animacy model (animate, inanimate) was significantly represented within EVC ($p<.05$), and within OTC ($ps<.001$) and VTC ($ps<.05$), both including and excluding EVC, but in neither of the frontal ROIs ($ps>.1$). The 21-feature semantic model showed significant representation within all ROIs tested ($ps<.05$), indicating that representations of semantic within frontal regions are more nuanced than the broad animate vs. inanimate gradient found within OTC (Sha et al., 2013; Grill-Spector & Weiner, 2014). Representation of arousal only, and the valence by arousal interaction, were significant only within OTC and non-EVC OTC ROIs ($ps<.05$), whereas valence only was not significantly represented within any of the ROIs. The interaction of animacy by valence, animacy by arousal, and animacy by valence by arousal were all significant within EVC, OTC including and excluding EVC and VTC including EVC ($ps<.05$). These findings are consistent with our findings reported in chapter that a combined semantic, valence, and arousal model (CSVA) out performed both a semantic only and an emotion model (valence by arousal) within OTC (Fig. 2-2 & figs. 2-S1 - 2-S3). Within non-EVC VTC, the animacy by arousal and animacy by valence by arousal interactions were also significant ($ps<.05$). Of the 3 animacy by emotion interactions, representation within frontal regions was only significant for animacy by arousal ($p<.05$), and only within the non-OFC Frontal ROI.

To further interrogate these animacy by emotion interactions, we tested for differential representations of emotion within either only animate or only inanimate stimuli. This also allowed for a comparison between results from this RSA with those from the PCA we reported in chapter 2. Breaking down the animate stimuli by either valence, arousal, or valence by arousal, we found significant representation of all three of these animate emotion models within OTC and VTC, both including and excluding EVC ($ps<.05$). These findings are inline with results from the PCA conducted in chapter 2, which found that the top 3 dimensions of representation within OTC contain information pertaining the the arousal and valence of animate stimuli. Furthermore, representation of arousal and valence by arousal for animate stimuli were significant in both frontal ROIs ($ps<.05$), but valence for animate stimuli was significant in neither. Turning to inanimate stimuli, representation of valence, arousal, and valence by arousal were also all significant for OTC & VTC including and excluding EVC ($ps<.05$). These findings are somewhat different from those from the PCA of voxel feature weights done in chapter 2. Both studies found evidence for representation of arousal of inanimate stimuli within OTC, but only this RSA found evidence for the valence of inanimate stimuli within OTC. Finally, OFC showed significant representation for valence and valence by arousal of inanimate stimuli ($ps<.05$), but not arousal of inanimate stimuli, and all 3 inanimate emotion models were significant within the non-OFC Frontal ROI ($ps<.05$).

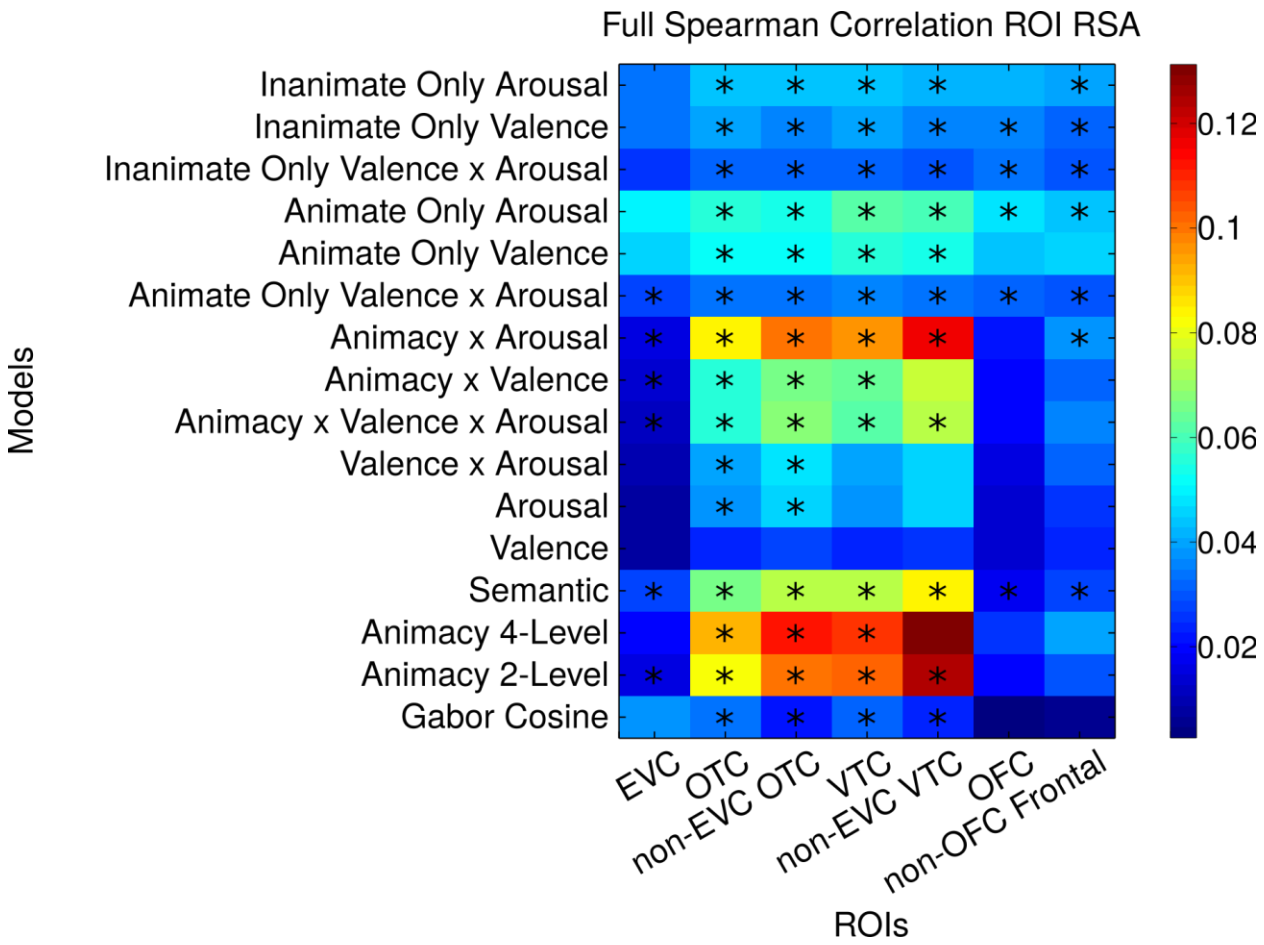


**Figure 3-18. Group Averaged Results from an RSA done within ROIs, across 16 models.** Correlation matrix showing group-averaged similarity, as quantified by full Spearman rank correlation, between brain RDMs created from seven ROIs, and 16 model RDMs (see methods for details of model definitions). Each row represents a single model, and each column a single ROI. Stars indicate those correlations that are significant from a one-tailed t-test of Fisher transformed correlations, across subjects ($p<.05$, Bonferroni corrected).

In addition to the full Spearman correlations just reported, partial Spearman correlations were also calculated between ROI brain RDMs and model RDMs. First, the Gabor wavelet RDM was partialed out of the RSA between all ROIs and all 15 remaining model RDMs (Fig. 3-19). The pattern of significance was exactly the same across all r (# of ROI) by (m-1) (# of models minus the Gabor model itself) representational similarity analyses, suggesting that these results were not influenced by low-level structural features. A second set of partial correlations were calculated which controlled for gabors, animacy, valence, and arousal in all RSAs (see RSA methods section for details). Crucially, when controlling for effects of animacy, valence and arousal, the pattern of significance across correlations was nearly identical here, bar two exceptions (Fig. 3-20). The arousal only model was no longer significant within OTC once gabors, animacy and valence were controlled for, and the valence by arousal model was no longer significant within non-EVC OTC once gabors and animacy were controlled for.

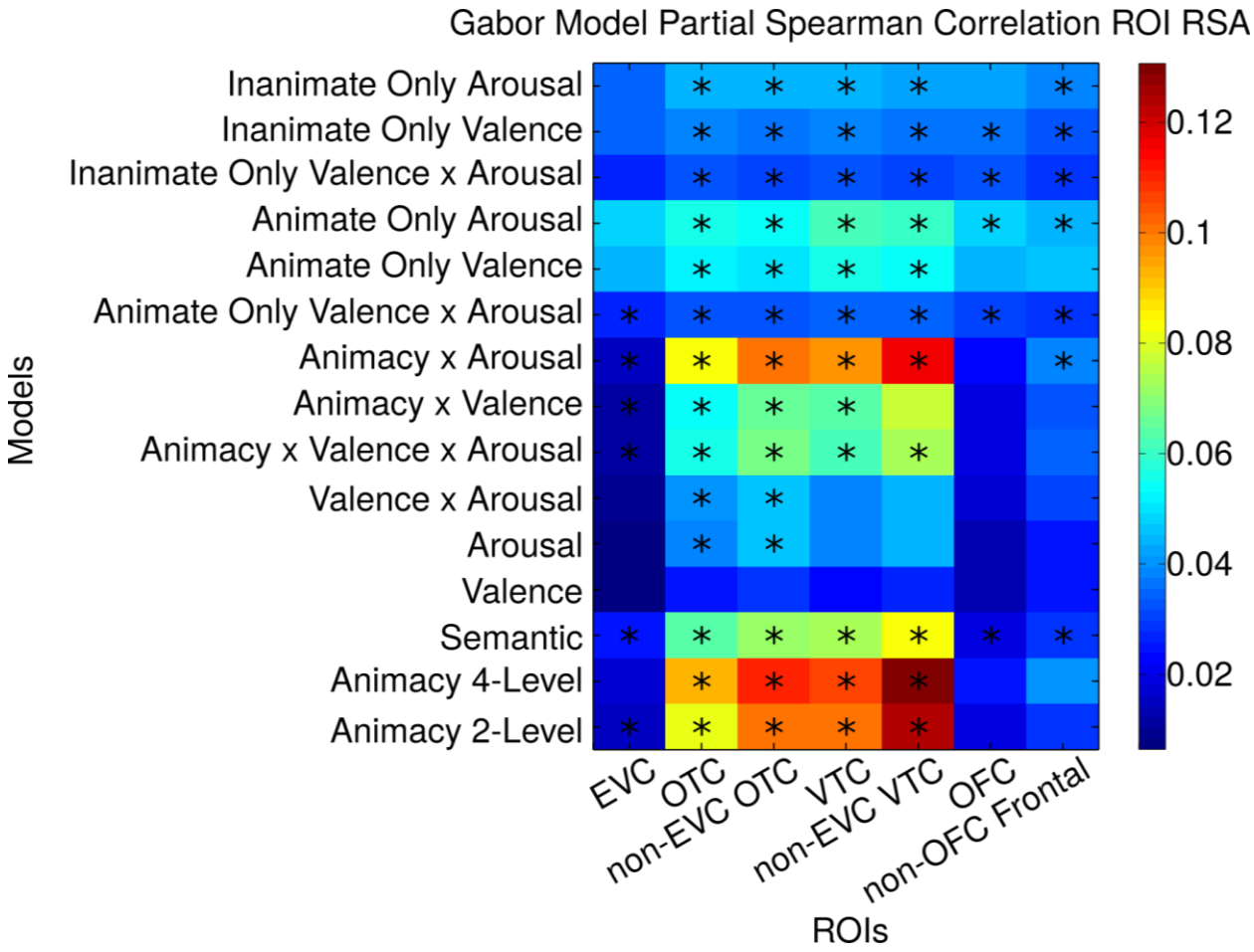


**Figure 3-19. Partial Correlations ROI RSA accounting for variance from Gabors.**
Correlation matrix showing the group-averaged similarity, as quantified by partial Spearman rank correlation, between brain RDMs created from seven ROIs, and 15 model RDMs (see methods for details of model definitions). For each RSA result shown, the Gabors Cosine model was partialed out of both brain ROI RDM and model RDM before taking the Spearman rank correlation. The Gabors Cosine model was not shown in this figure because partialing it out would leave no variance left to explain. Each row represents a single model, and each column a single ROI. Stars indicate those correlations that are significant from a one-tailed t-test of Fisher transformed correlations, across subjects ($p<.05$, Bonferroni corrected).

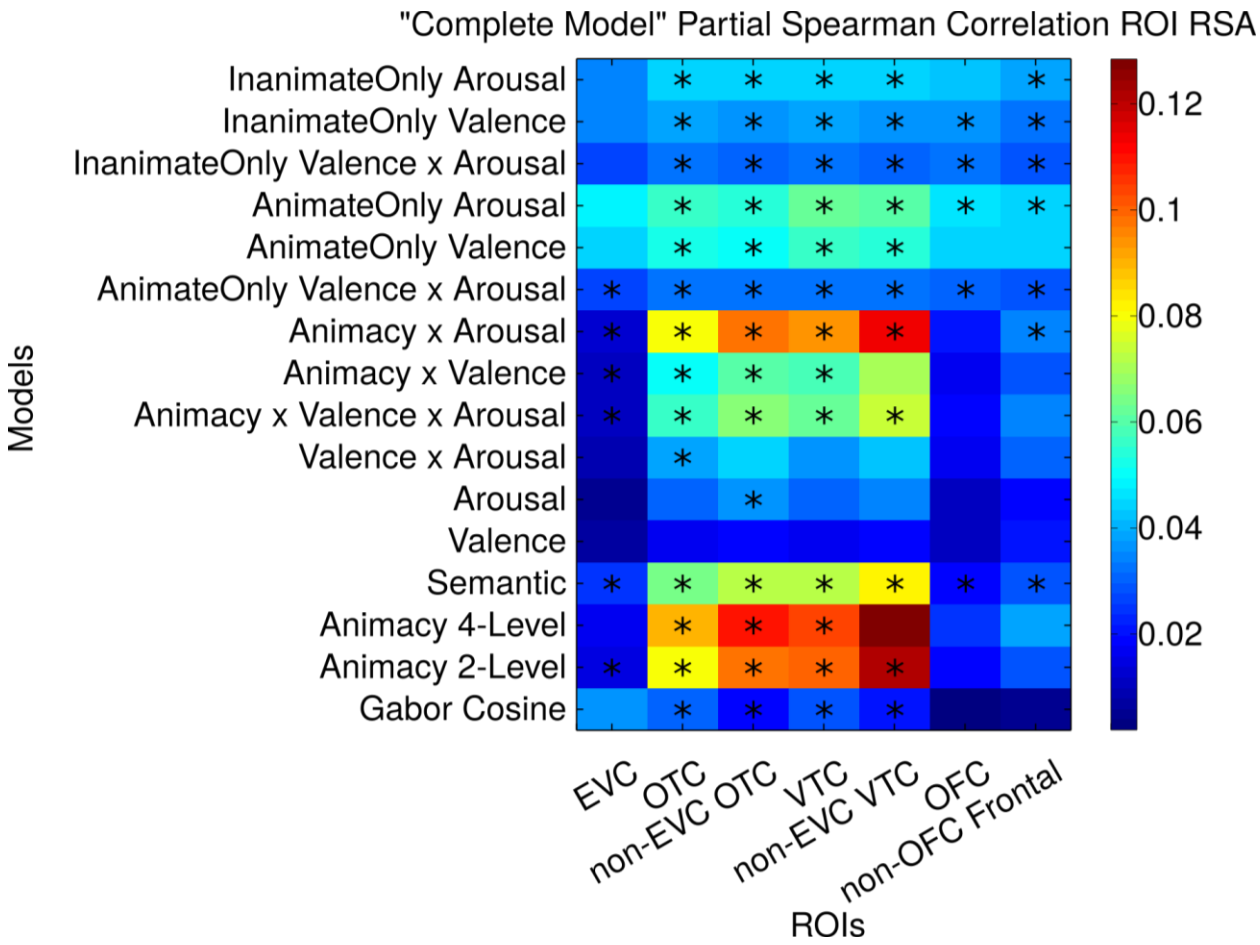


**Figure 3-20. Partial Correlations ROI RSA accounting for variance from Gabors, Animacy and Emotion.**

Correlation matrix showing the group-averaged similarity, as quantified by partial Spearman rank correlation, between brain RDMs created from seven ROIs, and 15 model RDMs (see methods for details of model definitions). For each RSA result shown (except the Gabors Cosine model itself), the Gabors Cosine model was partialed out of both brain ROI RDM and model RDM, the Animacy 2-Level model was partialed out of all models not containing Animacy, and the Valence and Arousal models were partialed out of all models not containing Valence or Arousal, respectively. Then the Spearman rank correlation was taken on the residuals of the brain ROI RDMs and the model RDMs. Each row represents a single model, and each column a single ROI. Stars indicate those correlations that are significant from a one-tailed t-test of Fisher transformed correlations, across subjects ($p<.05$, Bonferroni corrected).

In order to further investigate the spatial layout of these representations, searchlight analyses were conducted on the cortical surface. At each vertex on the cortical surface, the nearest 100 voxels were sampled, and a brain RDM constructed from those sampled voxel values. Partial Spearman correlations (controlling for Gabors, animacy, valence and arousal) between these brain RDMs and five selected model RDMs were then projected onto the cortical surface. In addition, partial Spearman correlations for the 2-Level Animacy model were done, controlling only for Gabors. Although a strict Bonferroni family-wise error (FWE) correction was conducted, much of cortex was found to be significant within all five models and across all

six subjects. In order to determine significance, most studies using searchlight RSA combine Fisher transformed single-subject searchlight correlations in a second level t-test analyses that is done within a normalized brain space. Since VWM is done on the single-subject level, we wanted to make the comparison between our VWM results and these RSA searchlight results in the native subject space. Thus, in order to calculate significance of RSA at a single subject level, p-values must be calculated on the correlation coefficients themselves. Due to the large number of stimuli in this study (n=1440), the RDM vectors correlated were very large (n=1,036,080) as they contained the pairwise similarity values between all 1440 stimuli. This meant that a very small correlation value resulted in a significant p-value, even after Bonferroni correction. We thus decided to threshold these correlation maps and only show values with a correlation coefficient (r-value) greater than 0.05, which is the smallest r-value that would be significant ($p<.05$, uncorrected) for a correlation between 1440 values, the number of independent stimuli (Figs. 3-21 - 3-26). While these maps display quantitatively derived values, they show the peaks of the similarity topography across the cortical surfaces for a somewhat arbitrary threshold. It could be argued that $p<.05$ is also an arbitrary threshold, however. In any case, we interpret these maps with this caveat in mind.

The results from the thresholded searchlight RSA (see previous paragraph) suggest that animacy is strongly represented within non-EVC OTC (panel A of Figs 3-21 - 3-26) across all 6/6 subjects. Additionally, controlling for valence and arousal by partialing them out did not qualitatively change the results for the Animacy model (panel B of Figs 3-21 - 3-26) for all 6/6 subjects. Valence for animate stimuli showed peaks within non-EVC OTC for all 6/6 subjects (panel C of Figs. 3-21 - 3-26), while valence for inanimate stimuli (panel D of Figs. 3-21 - 3-26) showed peaks only within 3/6 subjects (1 of those only showed 1 small cluster above threshold). Crucially, some of the same regions that showed peak representation to animacy also showed peak representation to valence for animate stimuli, which was mostly (although not exclusively) within FFA for five of the six subjects. These findings suggest that the significant representation of valence for inanimate stimuli within the non-EVC OTC ROI may have been driven primarily by a subset of the subjects. Peaks of representation for arousal of both animate (panel E of Figs. 3-21 - 3-26) and inanimate stimuli (panel F of Figs. 3-21 - 3-26) were present within non-EVC OTC for all six subjects. These findings are inline with findings from both the ROI RSA and the PCA analysis of OTC voxel features weights reported in chapter 2.

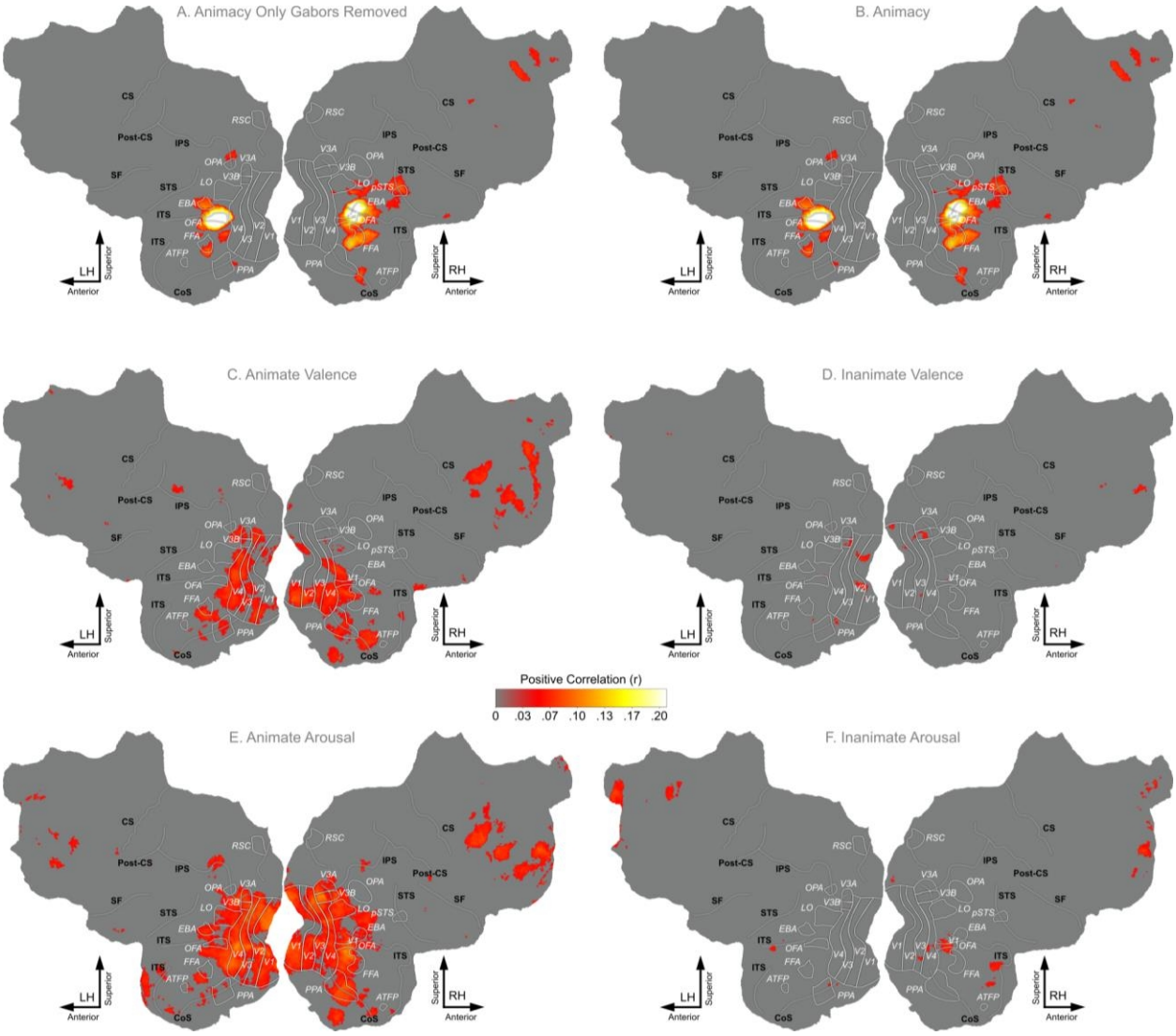


**Figure 3-21. Flatmaps showing Searchlight RSA Results for Subject 1.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

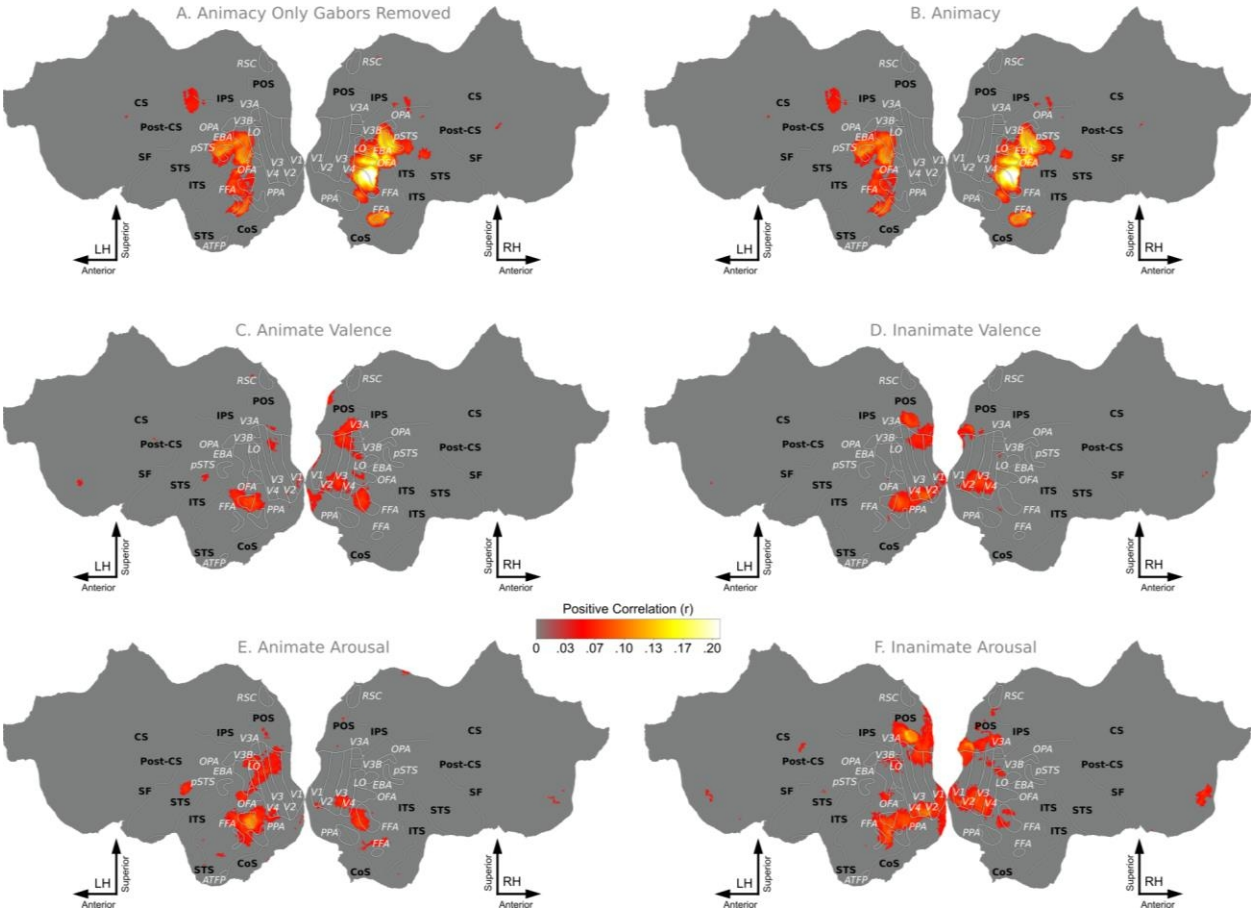


**Figure 3-22. Flatmaps showing Searchlight RSA Results for Subject 2.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

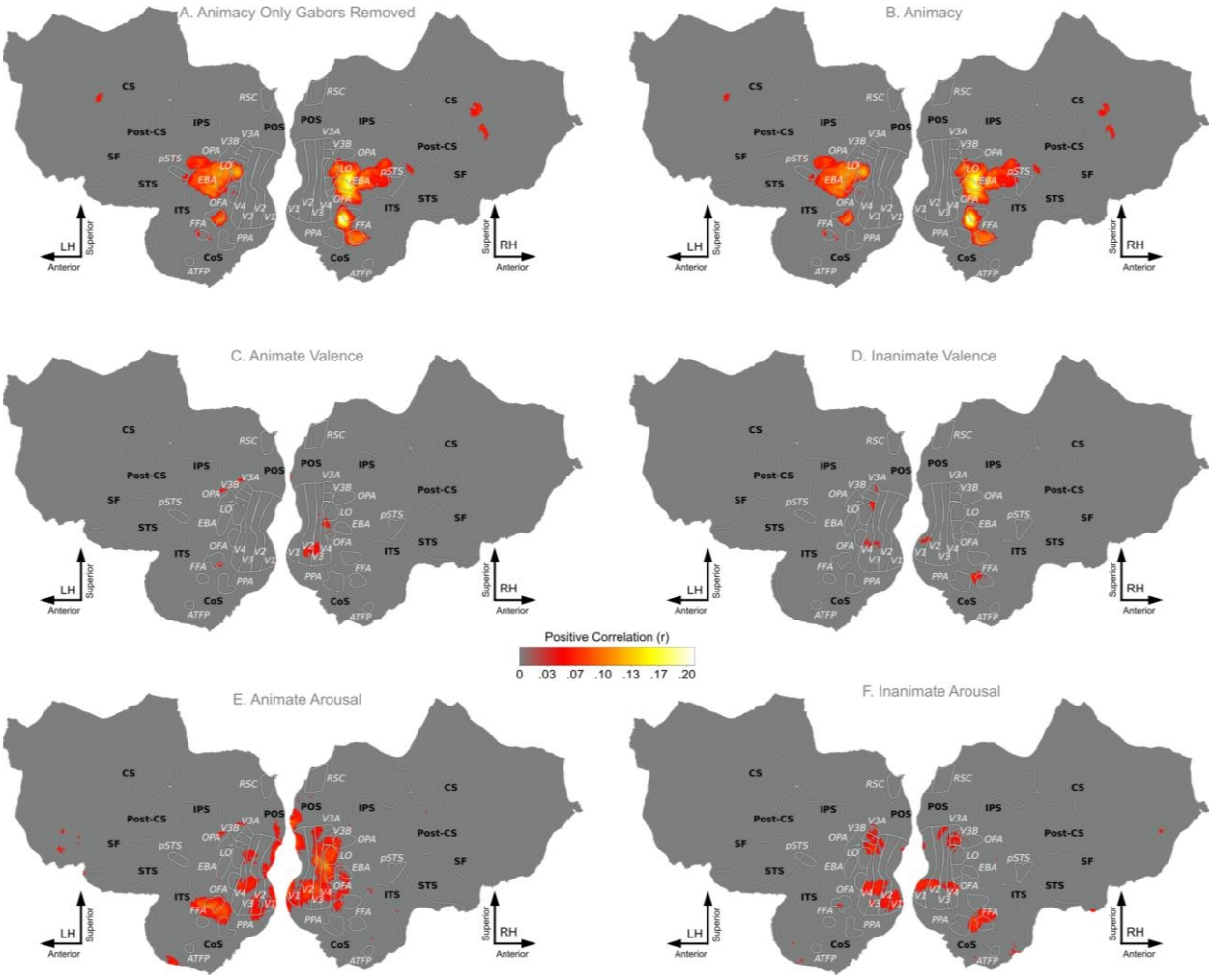


**Figure 3-23. Flatmaps showing Searchlight RSA Results for Subject 3.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

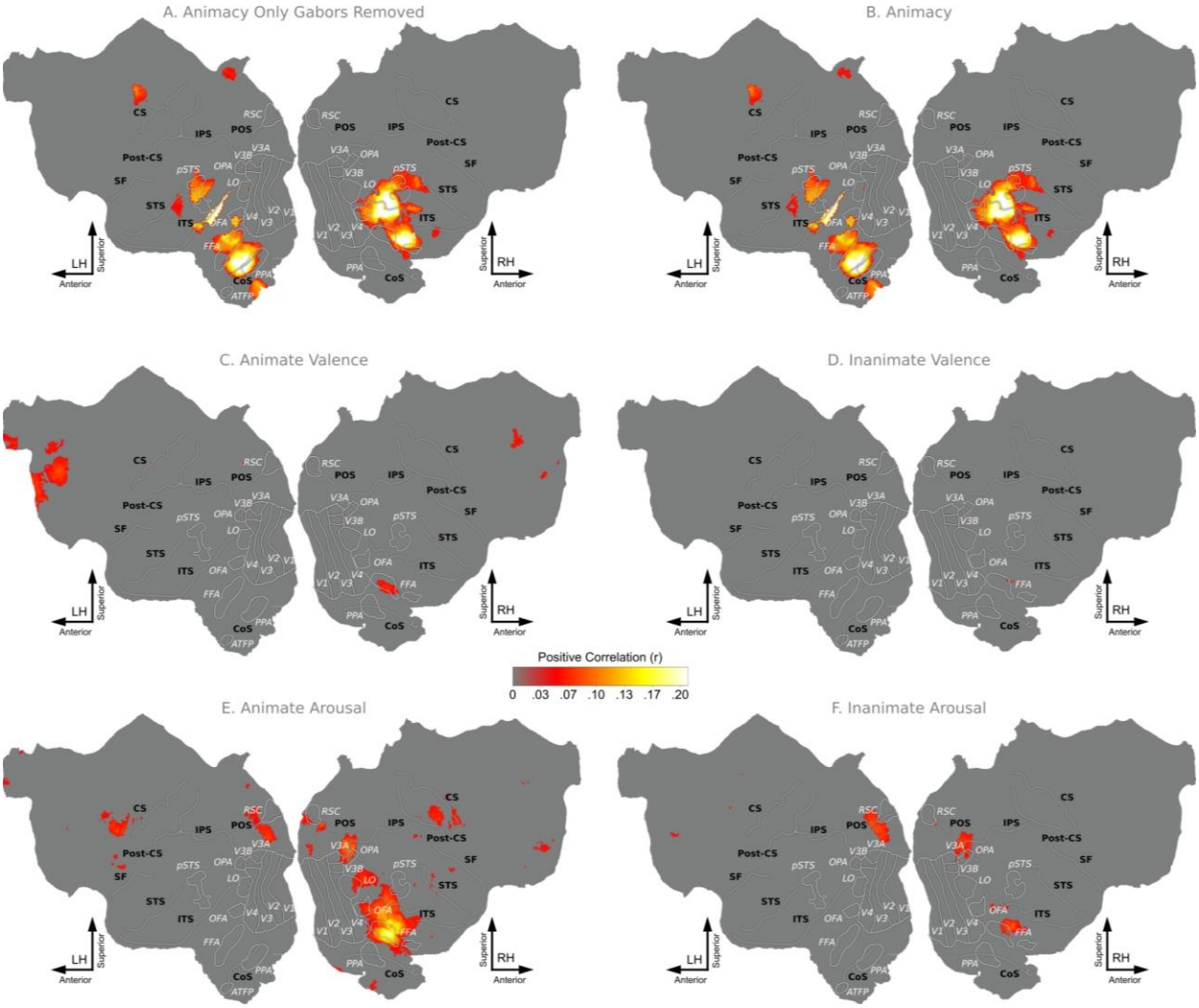


**Figure 3-24. Flatmaps showing Searchlight RSA Results for Subject 4.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

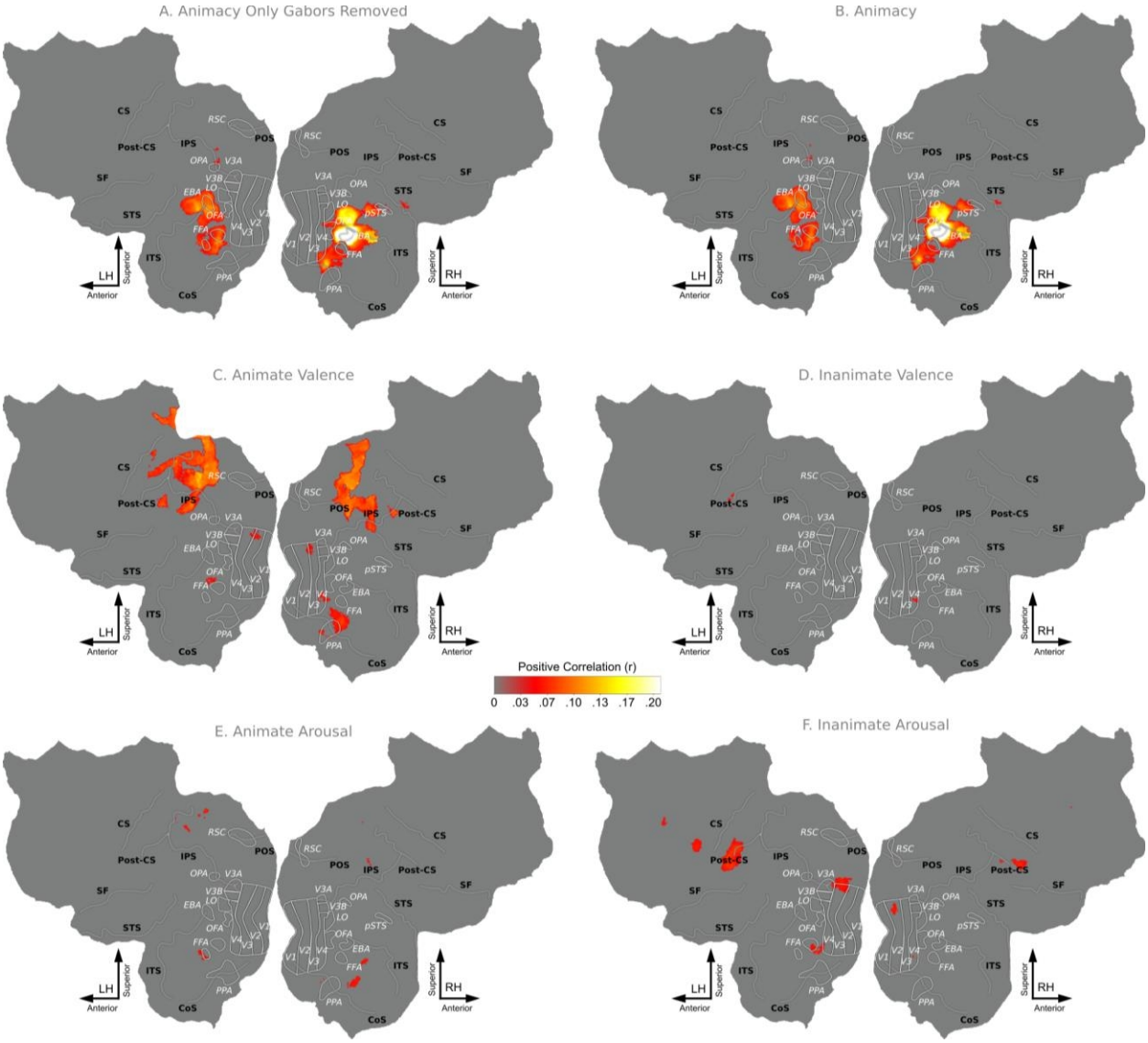


**Figure 3-25. Flatmaps showing Searchlight RSA Results for Subject 5.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

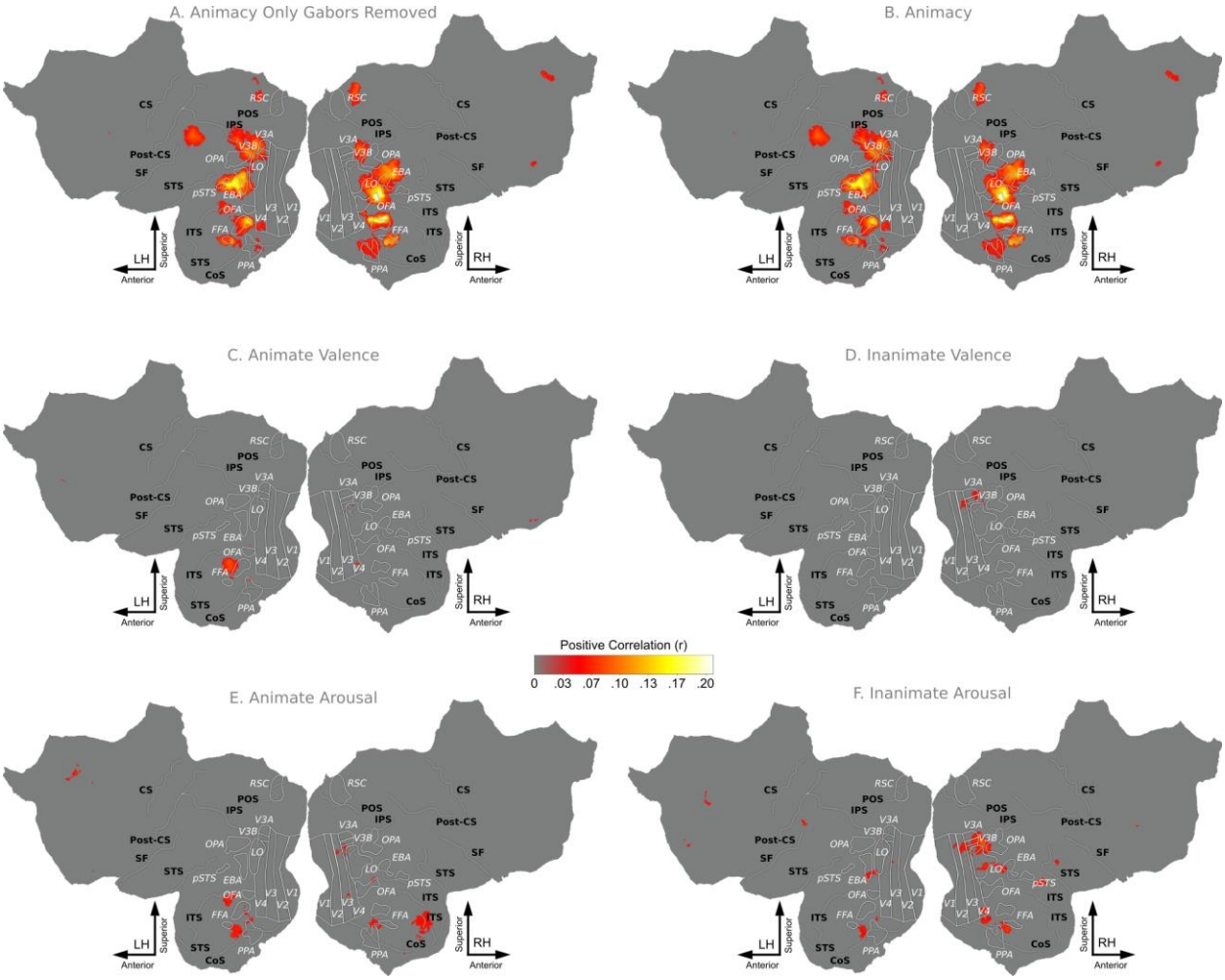


**Figure 3-26. Flatmaps showing Searchlight RSA Results for Subject 6.**

Flatmaps showing the results from doing searchlight RSAs on the cortical surface. Spearman correlation was used to determine similarity between brain and model RDMs, and significance of these correlations was determined by converting them into t-statistics and doing t-tests on the resulting values. Bonferroni correction at $p<.05$ was used for multiple comparisons correction. Due to the large number of stimuli, the resulting RDMs were large (>1,000,000 samples), and thus significance levels for the RSA was extremely low ($r>.0017$ for a $p<.05$, uncorrected). Thus, much of cortex was found to be "significant" and so voxels were further masked by including only those voxels where the Spearman correlation is greater than 0.05. This value (0.05) corresponds with the smallest correlation coefficient with a significant correlation ($p<.05$) 1440 elements, which is the number of independent stimuli used in the RSA.

# 4. Univariate SPM Analysis

## Introduction

The results reported in chapter 2 resulted from a voxel-wise modeling (VWM) analysis of BOLD data collected while subjects viewed a large number of naturalistic emotional images (n=1440 for model estimation, n=180 for model validation). VWM is a newer technique that offers several advantages over traditional univariate analysis (such as SPMs). Univariate analyses use a limited number of experimental conditions as regressors of interest, and calculate contrasts of those regressors to create statistical parametric maps (SPMs). The amount of data collected for univariate studies is often relatively small per subject, which limits the number of experimental conditions that can be tested. Additionally, inference is done using the same set of data that the model weights were estimated from, likely resulting in over-fitting. Finally, traditional univariate analyses using Gaussian kernels to spatially smooth the BOLD data which requires assumptions about the size of activation clusters expected. Spatially normalization it to a template brain is also used, which requires non-linear transformations and interpolation of the functional data. Both of these transformations lead to loss of information at the single voxel level. VWM overcomes these limitations of univariate analyses in three ways. First, by using regularized regression to estimate models, many more regressors (or features) can be fit to the same size data set as could be fit using standard ordinary least squares regression (OLS, used by SPM, FSL & AFNI neuroimaging packages). This allows for the construction of models that can more accurately describe the complex nature of representations within the brain. Second, model fit is quantified by predicting a held out set of validation, which greatly reduces the possibility of over-fitting to the data. Third, the BOLD data is not spatially smoothed nor spatially normalized to a template brain, all analyses are done within the subjects' native functional space, which prevents unnecessary loss of information. While all of these arguments can be made at a theoretical level, the question still remains: would the same results would have been found doing a traditional univariate analysis? Hence we conducted a univariate analysis using SPM in order to answer this question.

## Methods

### fMRI preprocessing

Data were preprocessed using Matlab version 8.0 (The Math- Works, Natick, MA) and SPM8 (Welcome Department of Imaging Neuroscience, London, UK). After conversion from DICOM to NIfTI format, diagnostics were run on the time series for each imaging run using the identical procedure defined in chapter 2 supplemental Method Details, in the "fMRI Preprocessing" section. Subsequent to this initial data-cleaning step, image realignment (correcting for head movement) was conducted, followed by slice time correction. No high-pass filter was used,

rather a 3rd order Savitzky-Golay filter with a window of 200 seconds was used to remove low-frequency scanner noise, as was done for the data used in the voxel-wise models fit in chapter 2. Finally, spatial smoothing was conducted using a Gaussian kernel with a full-width-to-half-maximum value of 5mm.

### fMRI Data Analysis

General linear modeling of the BOLD data was conducted using SPM 8 on data collected during 30 estimation runs (7.5 minutes each) of the experiment described in chapter 2. Across these thirty runs, a large set of naturalistic emotional images (n=1440) spanning 21 semantic categories and six valence (negative, neutral, positive) by arousal (high, low) conditions were presented for one second, with a three second inter-stimulus interval (ISI) (see supplemental "Method Details" for more details). The GLM was only fit to cortical voxels, as was done in the voxel-wise modeling from chapter 2. The per stimulus labels for animacy (animate/inanimate), valence (negative/neutral/positive) and arousal (high/low) described in chapter 2 were also used here. The valence and arousal ratings were defined by each participant separately in post-scan sessions. The animate/inanimate labels were defined by the mode of four independent raters. Onsets for each of the 12 animacy by valence by arousal conditions were modeled using delta functions and then convolved with the canonical hemodynamic response function (HRF), giving 12 regressors of interest. Six realignment (movement) nuisance regressors were included in order to reduce task-unrelated variance (noise). Per run constant bias terms were also included to account for between run differences in baseline BOLD signal. Voxel selection within SPM was done by providing an explicit mask which was the same voxel-selection mask as was used for the VWM analyses in chapter 2. Additionally, implicit masking was disabled by setting the threshold to negative infinity (default 80%), which tells SPM not to discard any voxels based on mean signal relative to the global mean signal. Following estimation of the GLM, contrasts of beta weights were created to isolate cortical regions responsive to various combinations of the animacy, valence and arousal labels. An F-contrast testing whether the full model (Animacy by Valence by Arousal) fit significantly in all brain voxels was calculated, and significant voxels ($p<.001$, uncorrected) were selected for use in the remaining contrasts of interest. For each contrast of interest, an F-statistic map was calculated, along with two t-statistic maps (one in each direction of the contrast, e.g. Animate minus Inanimate & Inanimate minus Animate for the Animacy contrast). Hypothesis testing was then done using these contrast maps, where significance of the F-statistic indicated that the contrast was significantly different from 0, and the t-contrasts indicating that the t-statistic was significantly greater than or less than 0. Correction for multiple comparisons was then applied across all selected cortical voxels (see above) using the false discovery rate (FDR), with a q-value of 0.05. These F & t statistic maps were then projected onto cortical flatmaps, and only voxels that survived the FDR correction were displayed. The following 10 contrasts were defined across all 12 regressors of interest, namely (Animate-Negative-Low, Animate-Negative-High, Animate-Neutral-Low, Animate-Neutral-High, Animate-Positive-Low, Animate-Positive-High, Inanimate-Negative-Low,

Inanimate-Negative-High, Inanimate-Neutral-Low, Inanimate-Neutral-High, Inanimate-Positive-Low, Inanimate-Positive-High).:

1. Animacy: (All Inanimate = -1, All Animate = +1)
2. Valence: (All Negative = -1, All Neutral = 0, All Positive = +1)
3. Arousal: (All Low-Arousal = -1, All High-Arousal = +1)
4. Valence by Arousal: (All High-Positive = +1, All High-Negative =-1, All Low-Positive = -1, All Low-Negative = +1, All Neutral=0)
5. Animacy by Valence (All Animate-Positive = +1, All Animate-Negative = -1, All Inanimate-Positive = -1, All Inanimate-Negative = +1, All Neutral=0)
6. Animacy by Arousal (All Animate-High-Arousal = +1, All Animate-Low-Arousal = -1, All Inanimate-High-Arousal = -1, All Inanimate-Low-Arousal = +1)
7. Animate Only Valence (All Animate-Negative = -1, All Animate-Neutral = 0, All Animate-Positive = +1, All Inanimate = 0)
8. Animate Only Arousal (All Inanimate-High-Arousal = +1, All Inanimate-Low-Arousal = -1, All Animate=0)
9. Inanimate Only Valence (All Inanimate-Negative = -1, All Inanimate-Neutral = 0, All Inanimate-Positive = +1, All Animate = 0)
10. Inanimate Only Arousal (All Inanimate-High-Arousal = +1, All Inanimate-Low-Arousal = -1, All Animate = 0)

## Results

SPM was used to fit an Animacy (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high, low) model, with 12 regressors of interest, to all 30 runs of estimation data collected for the experiment described in chapter 1. Animacy was used instead of all 21 semantic categories from the CSVA model given that regularization is not used in standard univariate SPM analyses and given that we sought to test whether the effects and interactions of animacy, valence and arousal observed with voxelwise modeling could be identified using a standard univariate SPM analysis. ). An F-contrast was calculated across all 12 model regressors of interest and those voxels with a significant fit ($p < .001$, uncorrected) were selected for further use. Across all six subjects F-tests revealed that the main effects of Animacy, Valence & Arousal (Figs. 3-27,3-30,3-33,3-36,3-39,3-41) were significant within at least some regions of OTC ($p<.05$, FDR corrected within voxels selected as described above). Main effects of Animacy showed significant activity across most of OTC for all six subjects. Activity corresponding to the main effect of Valence was significant across many regions of OTC for four of the six subjects, with more sparse activity in response to valence seen for two of the six subjects. Activity associated with the main effect of Arousal was significant throughout most of OTC for five out of six subjects, and within many, but not most, regions of OTC in the remaining subject. The interaction of Animacy by Valence (Figs. 3-27,3-30,3-33,3-36,3-39,3-41) was significant within some regions of OTC for three of the six subjects, with two additional subjects showing a single

small cluster. Additionally, the interaction of Animacy by Arousal (Figs. 3-27,3-30,3-33,3-36,3-39,3-41) showed significant activation within OTC for four out of six subjects.

In order to examine the direction of the main effects of Animacy, Valence & Arousal in OTC, uni-directional t-contrast maps were created. Breaking down the main effect of Valence, t-contrast maps are shown for both the positive greater than negative contrast and the negative greater than positive contrast (Figs. 3-28,3-31,3-34,3-37,3-40,3-42). Within OTC, five of the six subjects showed significant activation for the positive greater than negative contrast, although activation for one of those subjects was predominantly within EVC. Significant activity for the positive greater than negative contrast was found within bilateral FFA for four of the six subjects, and cortical regions neighboring RSA for five of the six subjects. In comparison, PC 3 scores from the PCA reported in chapter 2 were high within FFA across all six subjects. PC 3 was correlated with animate valence, and thus high scores on PC 3 indicate representation of positive greater than negative stimuli. Thus, FFA seems to more strongly represent positive stimuli than negative stimuli. Significant activity to valence within frontal regions was also observed for all six subjects. Across much of bilateral OFC, the positive greater than negative contrast was significant for all six subjects. PC 3 scores from chapter 2 were also high within left OFC across all six subjects, suggesting representation of positive greater than negative stimuli there. The negative greater than positive contrast was significant for a different subset of five of the six subjects. Regions in and near PPA and LO showed significant activity to the negative greater than positive contrast for the five subjects were it was significant anywhere within OTC. Additionally, PPA showed low PC 3 scores in the PCA reported in chapter 2, inline with these univariate findings. Regions of the non-OFC frontal ROI also showed significant activation in the negative greater than positive contrast across five of the six subjects. Thus, the main effects of Valence within OTC were not driven exclusively by either positive or negative stimuli, although patterns of activity within several specific subregions of OTC were consistent across subjects.

For the main effects of Arousal, in contrast (Figs. 3-28,3-31,3-34,3-37,3-40,3-42), the high greater than low arousal contrast was significant across much of OTC for all six subjects, while no subjects showed significant activity for the low greater than high arousal contrast in OTC. Additionally, the high greater than low arousal contrast was significant within similar non-OTC frontal regions for five of the six subjects. Interestingly, the low greater than high arousal contrast was significant across regions of parietal and frontal lobes for four of six subjects, and this activity was right-lateralized for three of those four subjects. This effect was not seen across the same four subjects when inspecting PC 1 scores from the PCA done in chapter 2. Effects of the animate greater than inanimate contrast were seen across wide swaths of OTC for all six subjects (Figs. 3-28,3-31,3-34,3-37,3-40,3-42), while only four of six subjects showed significant activity for the inanimate greater than animate contrast, with two of those four subjects showing only several small clusters of activity. The medial to lateral animacy gradient across VTC observed in previous findings (Grill-Spector & Weiner, 2014; Sha et al., 2013) was also observed in four of the six subjects using a univariate SPM analysis. The two subjects where

this gradient was not fully observed still showed significant activation to animate stimuli within fusiform cortex, but there was no significant activation to inanimate stimuli anywhere within VTC. In contrast, inspecting results of the PCA from chapter 1 reveals that PC 1 scores of all six subjects showed evidence of the animacy gradient across VTC. This difference between univariate SPM findings and VWM PCA findings may be explained by the broad range of inanimate stimuli modeled using the same regressors in the SPM analysis (the six inanimate valence by arousal regressors), whereas they were modeled using different, more specific features in the VWM model (54 inanimate semantic category valence by arousal features).

Next we looked at the interactions of Valence and Arousal with Animacy. The Animacy by Valence and Animacy by Arousal interactions were further broken down, and t-contrast maps were shown for Valence and Arousal as just described, but for animate and inanimate stimulus categories separately. The Animacy by Valence interaction was significant within any OTC regions for five of six subjects (subjects 1-4,6), and so we only showed these interaction maps for those five subjects. Within animate stimuli, four of five subjects showed significant activity (Figs. 3-21,3-24,3-27,3-30,3-35) to either the positive greater than negative contrast (two of five subjects) or the negative greater than positive contrast (three of five subjects). Regions between RSA and EVC, as well as inferior temporal sulcus (ITC) showed significant animate positive greater than negative contrast activity for both subjects where that contrast was significant anywhere within OTC. Activity within superior temporal sulcus (STS) was seen for the animate negative greater than positive contrast across all 3 subjects showing activity to that contrast anywhere within OTC. A different picture for inanimate stimuli emerged (Figs. 3-29,3-32,3-35,3-38,3-43), as only one of five subjects showed significant activity within OTC for the positive greater than negative contrast, and that activity was exclusively located within EVC. However, three of five subjects had significant activation to the negative greater than positive contrast for inanimate stimuli. Two of these three subjects showed similar activation patterns within PPA and OPA for the inanimate negative greater than positive contrast. Thus the interaction between Animacy and Valence seems to be driven by both negative and positive animate stimuli, but only negative inanimate stimuli. Turning to the Animacy by Arousal interaction contrast (Figs. 3-29,3-32,3-35,3-38,3-43), only four of six subjects (subjects 1,3,4,6) had any significant activity within OTC and so maps were shown for only those four subjects. All four of four subjects showed vast regions of OTC significant in the high greater than low arousal contrast for animate stimuli, while only one of four subjects were significant within OTC for the low greater than high contrast, and that subject only showed significant activity within EVC. For inanimate stimuli, three of four subjects showed significant activity within OTC to the high greater than low arousal contrast, and none of the subjects had significant low greater than high arousal contrast within OTC. It seems that high arousal stimuli evoke more activity across much of OTC than do low arousal stimuli, and that activity was seen across more subjects for animate than inanimate stimuli.

Given these results, can we conclude whether this univariate SPM analysis revealed the same conclusions as the VWM analysis reported in Chapter 2? This analysis found strong

evidence that in addition to animacy, valence and arousal are both represented within OTC, just as the VWM study did (see Fig.2- 2 and figs. 2-S1 - 2-S3). Particularly, regions of FFA, PPA, and RSC showed similar activation patterns both across subjects in the univariate analysis and between the SPM and VWM analyses. Additionally, similar activity to valence and arousal were found in both OFC and non-OFC frontal ROIs in both the SPM and VWM analyses. The results investigating evidence for an interaction between animacy and valence, and animacy and arousal, are mixed however. Using principal components analysis (PCA) on the feature weights of OTC voxels across all size subjects revealed that the top 3 dimensions of representation within OTC contain information about arousal for animate stimuli, and to a much lesser degree inanimate stimuli, and valence for animate stimuli. Similarity between these group PCs and the single subject PCs were highly significant ($ps<10^{-8}$), suggesting a shared representational space across all six subjects. Results from this univariate analysis showed some evidence for interactions between animacy and both valence and arousal, but only within subsets of the subjects. The Animacy by Valence contrast was significant across multiple OTC regions in three of six subjects, although activation was in different OTC regions for those three subjects. An additional 2 subjects showed a single small cluster for the animacy by valence contrast. Next, we investigated the differential effects of animacy on the interaction between animacy and valence. Four of the five subjects with any evidence of an OTC Animacy by Valence interaction showed increased activation for valence of animate stimuli, while three of five subjects showed increased activity for valence in the inanimate stimuli. This effect of valence was driven by positive and negative animate stimuli, while exclusively by negative inanimate stimuli. Thus, we see some evidence of representation of valence for animate (and less for inanimate) stimuli within OTC, although it is only within a subset of subjects, unlike the similar representational space across all six subjects in the PCA of chapter 2. In those subjects showing evidence of valence activity for animate stimuli there was similarity within several OTC regions such as RSA and STS. A similar story emerged for the interaction of animacy by arousal, as four of six subjects showed significant activation to that contrast. Looking at animate and inanimate stimuli separately for those four subjects revealed evidence that both animate (four of four subjects) and inanimate (three of four subjects) stimuli had increased activity for high relative to low arousal images, albeit with lower magnitude, across fewer voxels and in fewer subjects for the inanimate stimuli. This is in contrast to extremely similar PC 1 loadings across subjects in chapter 2. Again here, there is some evidence for representation of an interaction of animacy with arousal, but only in a subset of the subjects, and thus the picture is less clear than that presented by the PCA of CSVA model feature weights presented in chapter 2.

One important caveat to this analysis is that, although a traditional univariate SPM analysis was conducted, the experimental design was created with a VWM analysis in mind. This meant that a much larger number of stimuli were used, nearly 10 times comparable studies (see Chikazoe et al., 2014 which used 128 stimuli), that spanned numerous semantic and emotional categories. The additional power gained from such a design likely contributed to the fact that some evidence was found using a univariate analysis that were similar to those found using

VWM. In line with our approach, several recent papers (Grice et al., 2017; Normand, 2016; Smith & Little, 2018) have argued for the advantages of doing "small-N design" experiments for psychology and neuroscience studies, where a large number of samples are collected for a small number of subjects. Smith & Little (2018) argue that by focusing experimental power at the single subject level, "small-N design" experiments provide high-powered tests of the psychological or neuroscientific effects of interest in individuals without worrying about the uncontrolled effects of individual differences. By focusing experimental power at the group level using "large-N designs", individual differences are averaged over for the sake of population inference. What this exact population is hardly ever gets specified, and given the sampling biases known to occur in psychological studies (Henrich et al., 2010), can hardly be considered to be all humans. Thus, for phenomena where individual differences do exist, large group averages can actually decrease one's statistically power, as the models specified do not account for all known variables. For these reasons, and others, "small-N designs" offer advantages over "large-N designs", and the fact that this study used a small number of subjects and a large number of samples could account for the partial success of this univariate analysis in confirming some of the findings from Chapter 1.

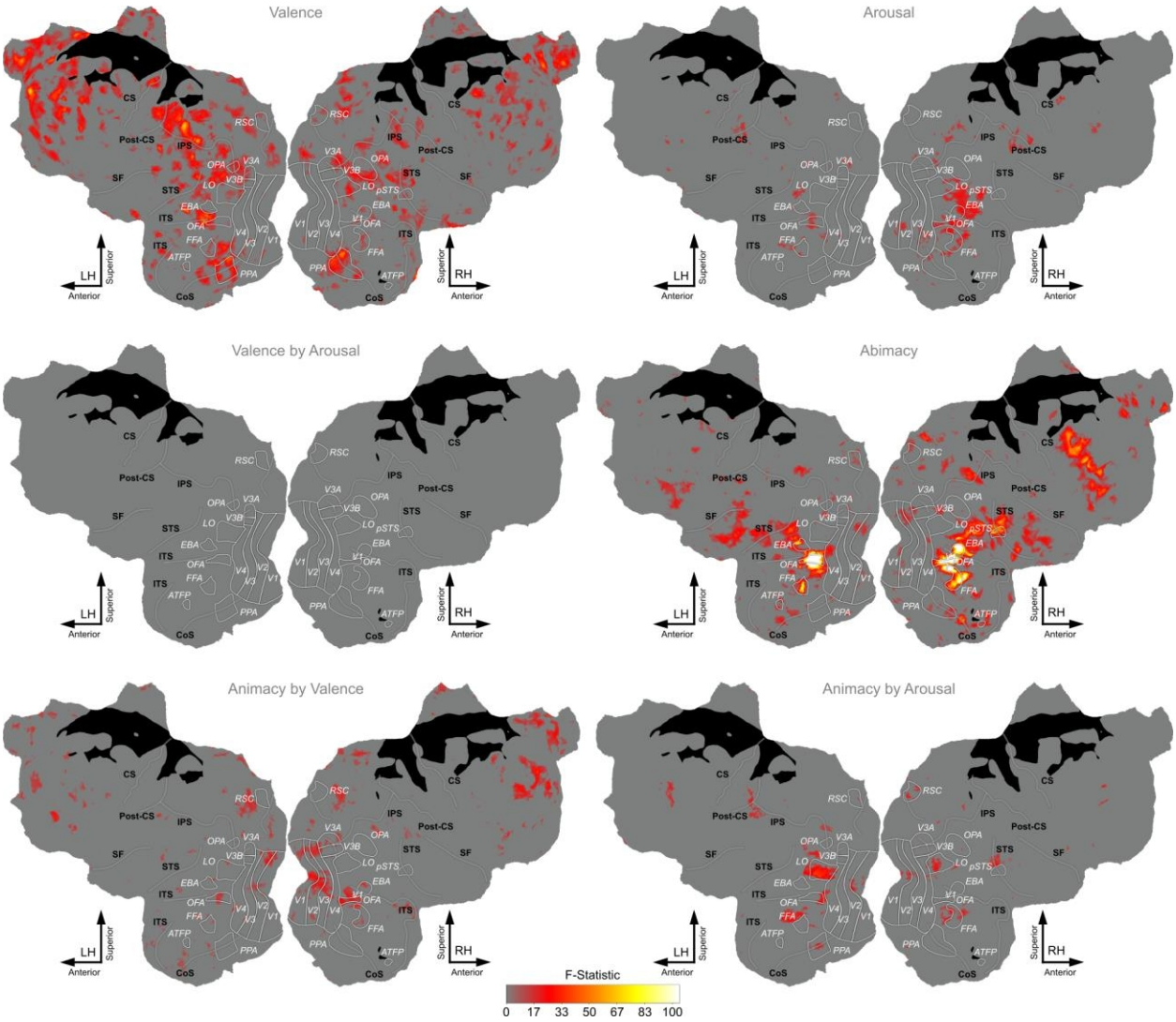


**Figure 3-27. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 1.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 1. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left), and Arousal (top-right) were significant within regions of OTC, but their interaction (middle-left) was not. The main effect of Animacy (middle-right) was significant across much of OTC, and both the interaction of Animacy by Valence (bottom-left) and Animacy by Arousal (bottom-right) showed some significant activation within OTC.

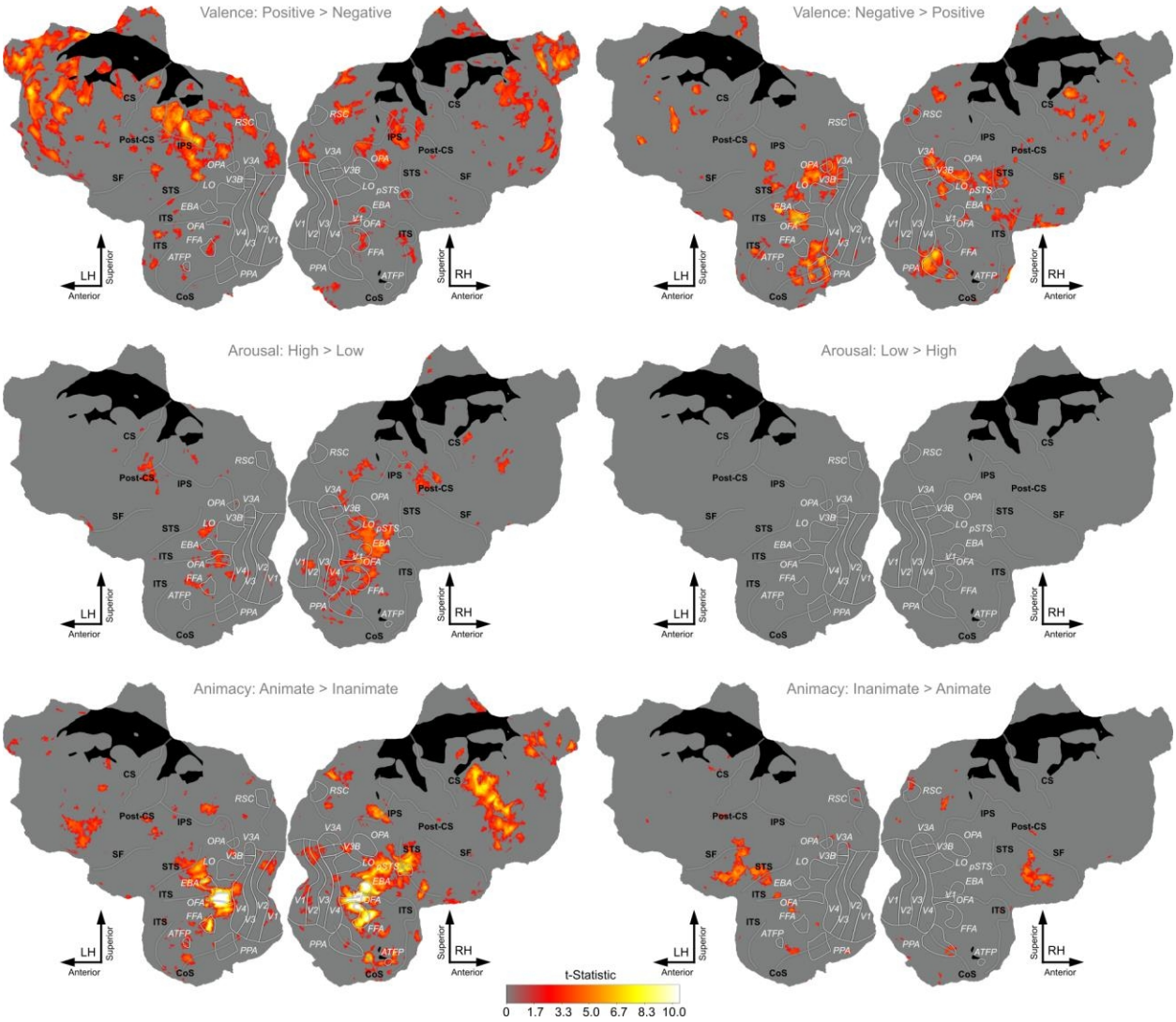


**Figure 3-28. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 1.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 1. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, both the positive greater than negative (top-left) and the negative greater than positive (top-right) contrast showed significance within many regions of OTC. Many regions of OTC showed significant activity to the high greater than low arousal contrast (middle-left), while no significant activity for the low greater than high arousal contrast was seen in OTC. The animate greater than inanimate contrast (bottom-left) showed significant activation across much of OTC, while the inanimate greater than animate contrast (bottom-right) showed limited significant activity within OTC. The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA.

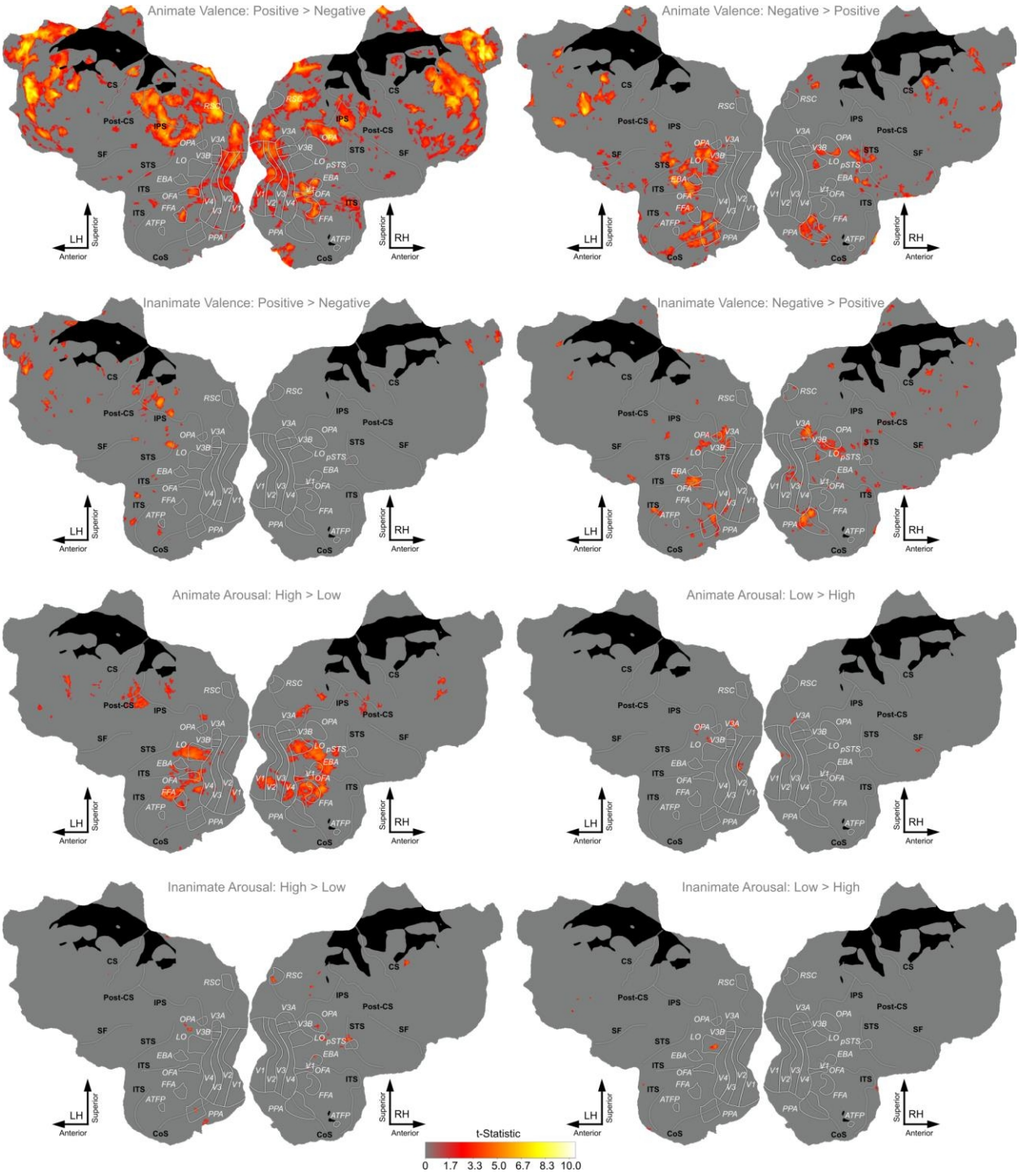


**Figure 3-29. t Contrast Maps breaking down the Interactions of Animacy by Valence & Animacy by Arousal for Subject 1.**

Flatmaps for subject 6 showing t-contrasts which breakdown the interactions of Animacy by Valence (rows 1 & 2) and Animacy by Arousal (rows 3 & 4) for animate and inanimate stimuli. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. Row 1 shows both t-contrasts for Valence of just animate stimuli. The positive greater than negative contrast (row 1-left) for animate stimuli was significant within much of OTC, as was the negative greater than positive contrast (row 1-right). For inanimate stimuli in row

2, the positive greater than negative contrast (row 2-left) reveals no significant activity in OTC, while the negative greater than positive contrast (row 2-right) showed significant activity across several regions within OTC. For the Animacy by Arousal interaction, row 3 shows contrast maps for animate stimuli. The high greater than low arousal contrast (row 3-left) for animate stimuli is significant across much of OTC, while the low greater than high arousal contrast (row 3-right) for animate stimuli is not significant within OTC. For inanimate stimuli, the high greater than low arousal contrast is not significant within OTC. The low greater than high arousal contrast is significant for one small cluster within OTC for inanimate stimuli.

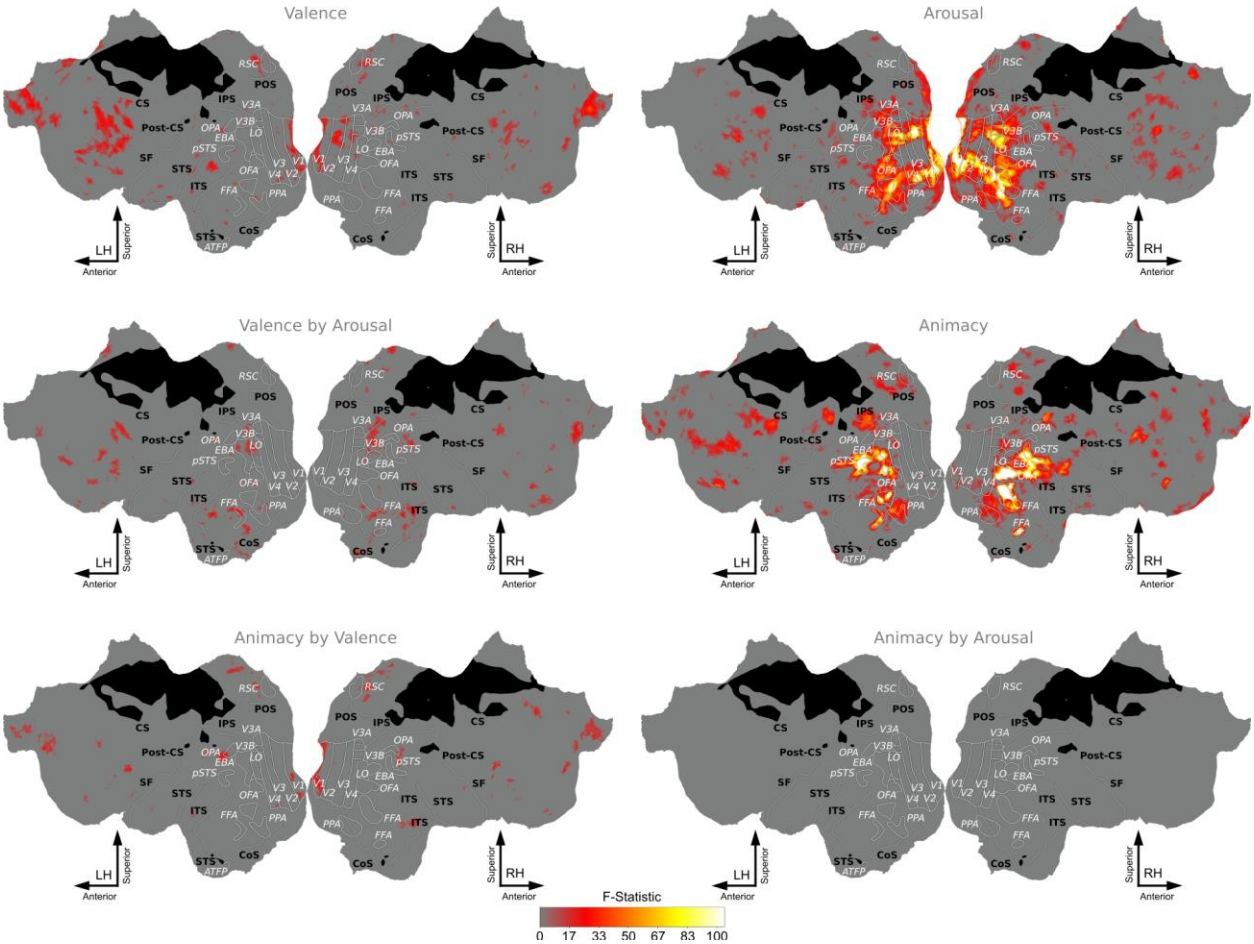


**Figure 3-30. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 2.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 2. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left) were significant within early-visual cortex (EVC), Arousal (top-right) was significant across much of OTC, and their interaction (middle-left) was significant within some of OTC. The main effect of Animacy (middle-right) was significant across much of OTC, the interaction of Animacy by Valence (bottom-left) was significant in some regions of V1, and Animacy by Arousal (bottom-right) showed no significant activation within OTC.

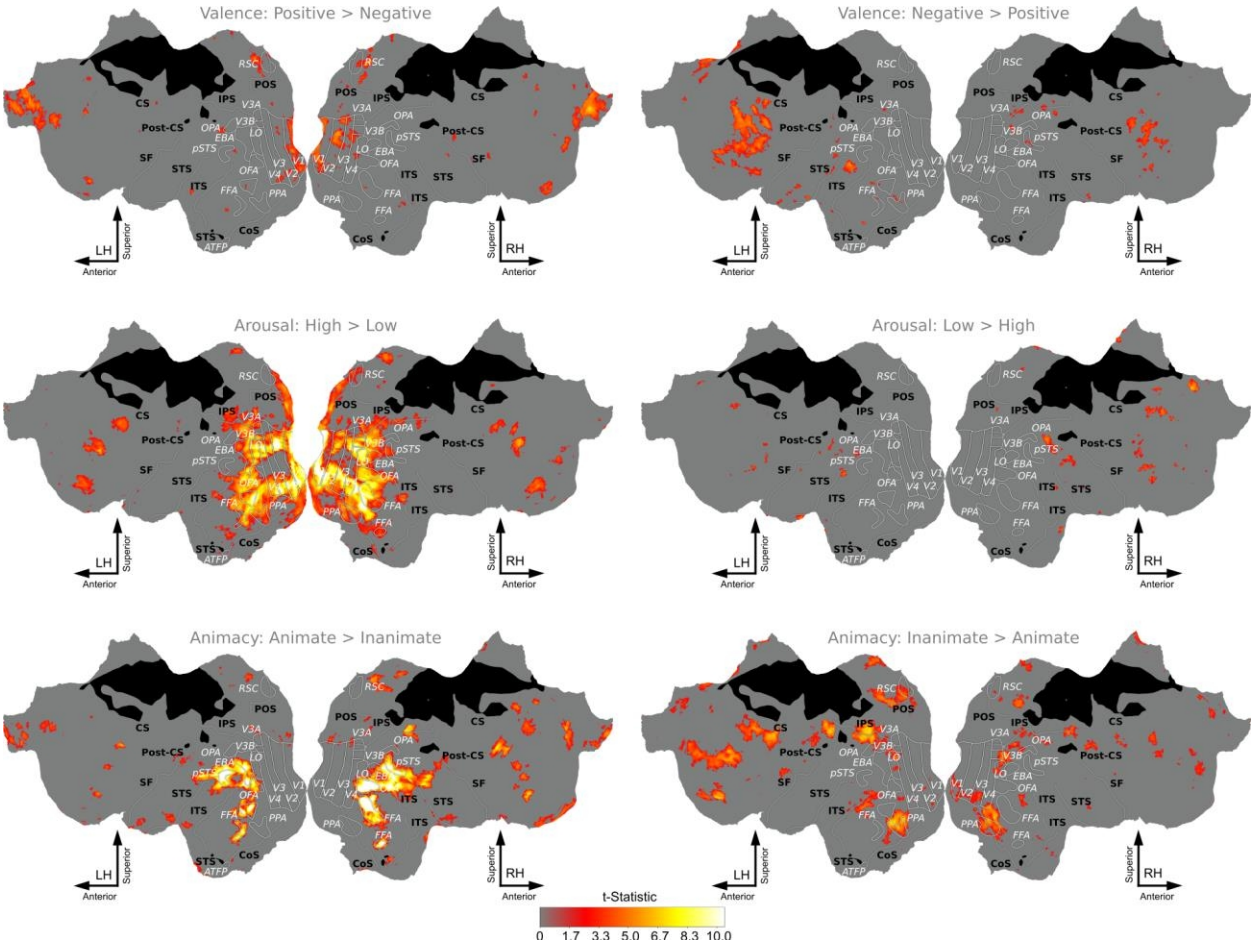


**Figure 3-31. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 2.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 2. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, the positive greater than negative (top-left) contrast showed significant activity within earl-visual cortex (EVC. The negative greater than positive (top-right) showed significance within limited regions of OTC. Many regions of OTC showed significant activity to high greater than low arousal (middle-left), while no significant activity for low greater than high arousal was seen in OTC (it was seen in some Parietal regions however). Both the animate greater than inanimate (bottom-left) and inanimate greater than animate (bottom-right) both showed significant activation across much of OTC. The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA, and the inanimate greater than animate contrast was significant in regions known to represent inanimate stimuli such as PPA, RSA and LO.

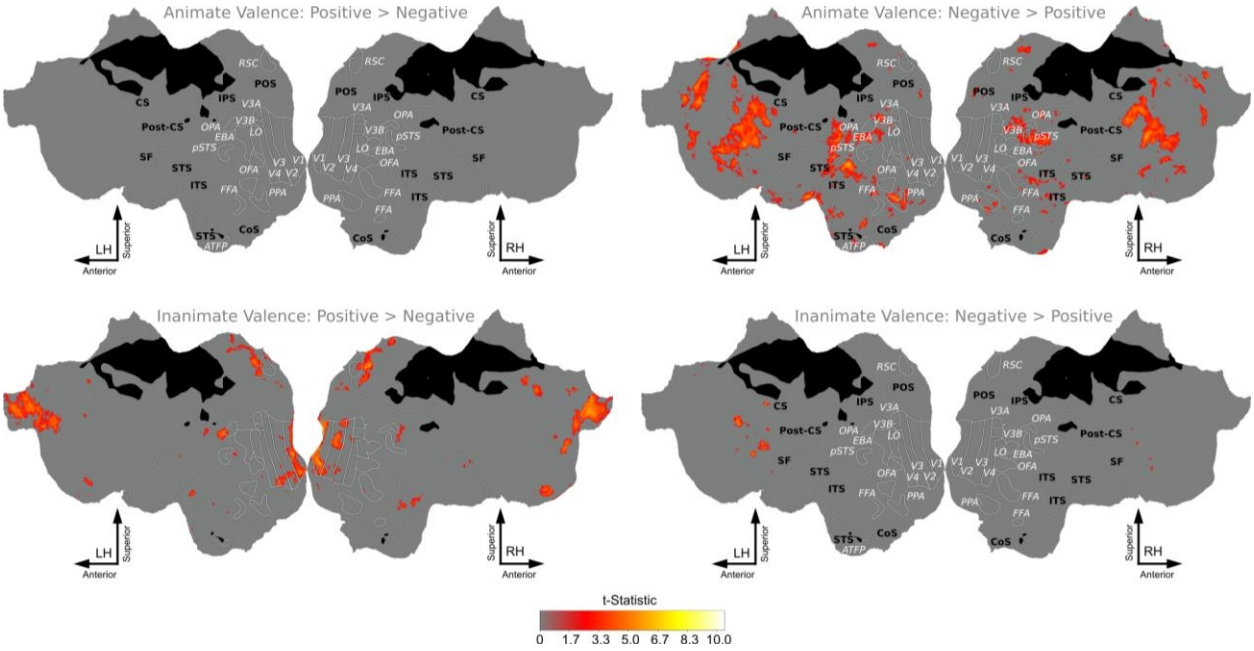


**Figure 3-32. t Contrast Maps breaking down the Interactions of Animacy by Valence for Subject 2.**

Flatmaps for subject 6 showing t-contrasts which breakdown the interactions of Animacy by Valence (rows 1 & 2) for animate and inanimate stimuli. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. Row 1 shows both t-contrasts for Valence of just animate stimuli. The positive greater than negative contrast (row 1-left) for animate stimuli shows no significant activity within OTC, while the negative greater than positive contrast (row 1-right) for animate stimuli shows some significant activation within OTC. For inanimate stimuli in row 2, neither the positive greater than negative contrast (row 2-left) nor the negative greater than positive contrast (row 2-right) are significant in OTC.

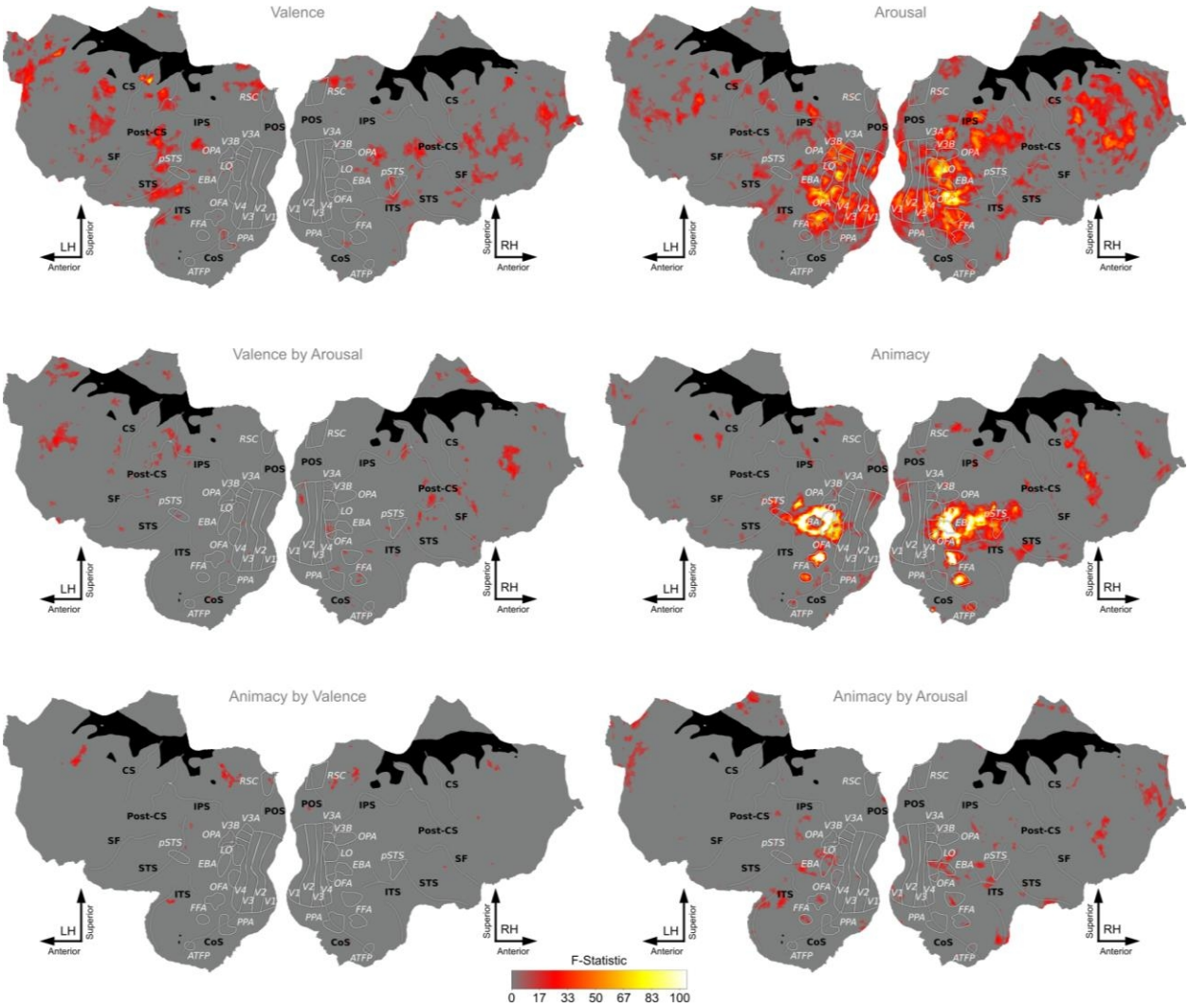


**Figure 3-33. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 3.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 3. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left) were significant within limited regions of OTC while Arousal (top-right) was significant across much of OTC, and their interaction (middle-left) was again significant within limited regions of OTC. The main effect of Animacy (middle-right) was significant across much of OTC, the interaction of Animacy by Valence (bottom-left) was not significant in any of OTC, and Animacy by Arousal (bottom-right) showed significant activation within some of OTC.

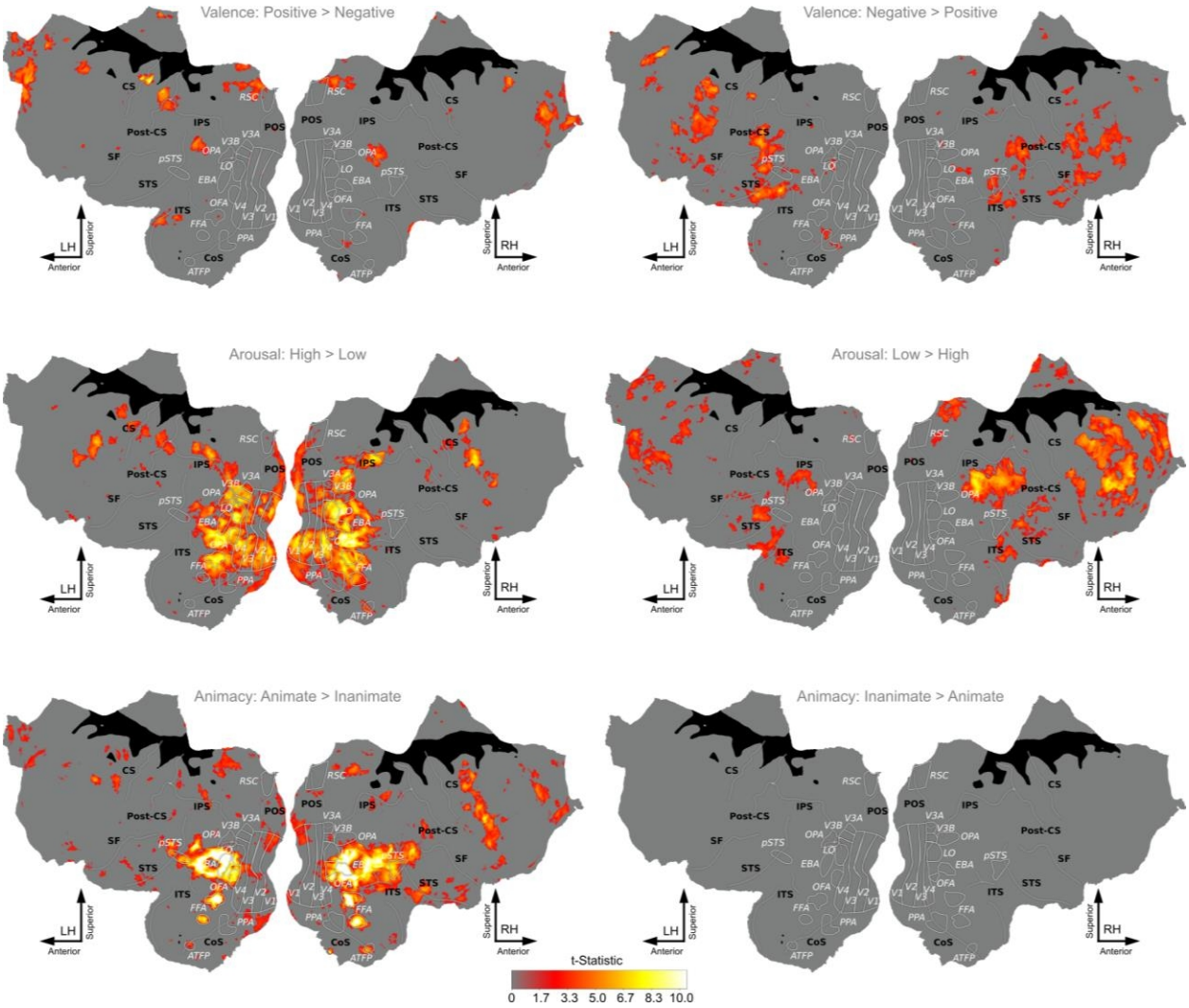


**Figure 3-34. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 3.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 3. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, both positive greater than negative (top-left) and negative greater than positive (top-right) were significant across limited regions of OTC. Many regions of OTC showed significant activity to high greater than low arousal (middle-left), while no significant activity for low greater than high arousal was seen in OTC (it was seen in Parietal regions however). Only the animate greater than inanimate (bottom-left) showed significant activation to the Animacy main effect, across much of OTC. The inanimate greater than animate (bottom-right) contrast showed no significant activity within OTC.The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA.

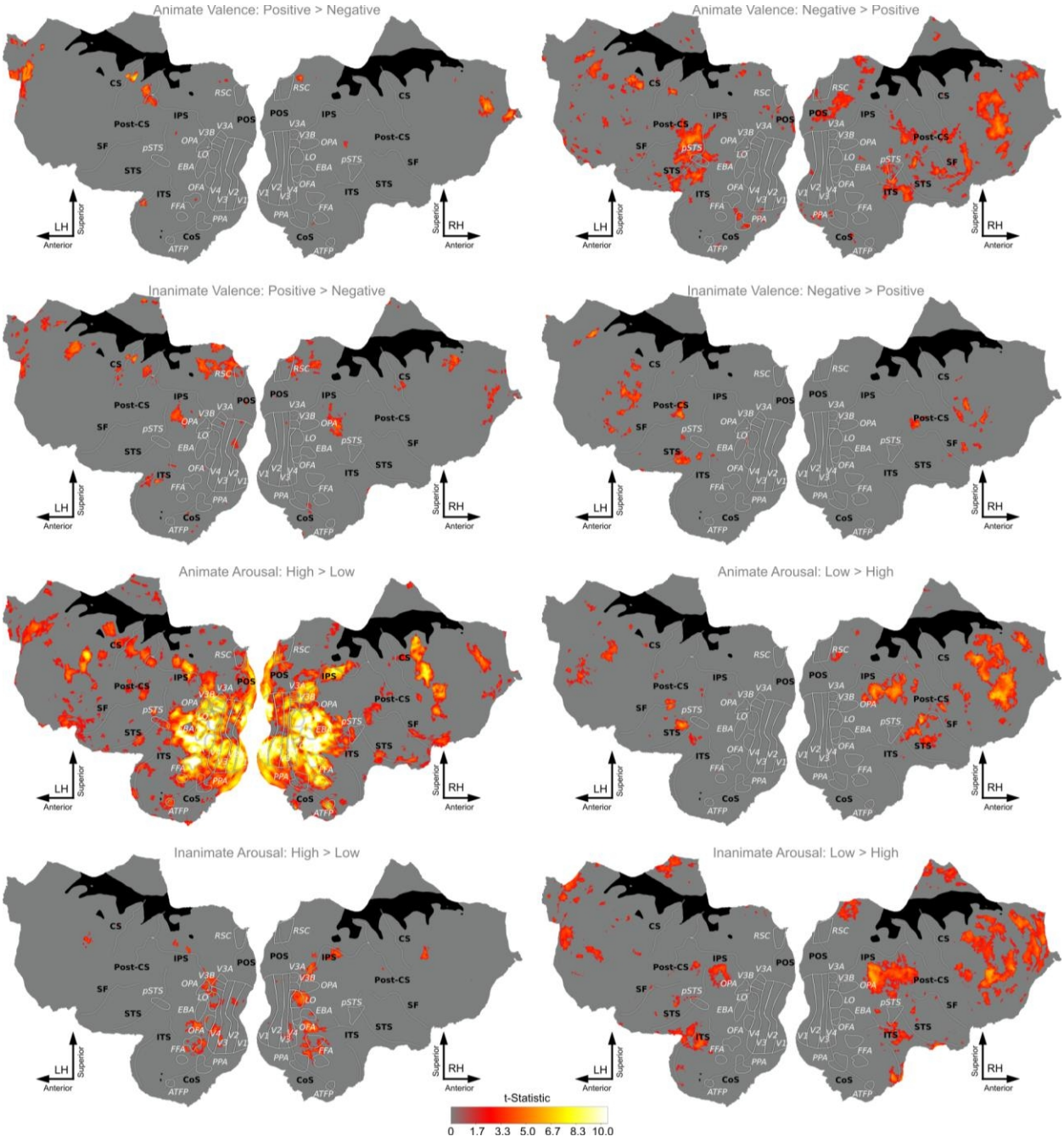


**Figure 3-35. t Contrast Maps Breaking Down the Interactions of Animacy by Valence, & Animacy by Arousal for Subject 3.**

Flatmaps for subject 6 showing t-contrasts which breakdown the interactions of Animacy by Valence (rows 1 & 2) and Animacy by Arousal (rows 3 & 4) for animate and inanimate stimuli. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. Row 1 shows both t-contrasts for Valence of just animate stimuli. The positive greater than negative contrast (row 1-left) for animate stimuli shows no significant activity within OTC, while the negative greater than positive contrast (row 1-right) for animate stimuli shows some significant activation within OTC. For inanimate stimuli in row 2, the positive greater than negative contrast (row 2-left) reveals some significant activity in early-visual cortex (EVC), while the negative greater than positive contrast (row 2-right) is not significant in OTC. For the Animacy by Arousal interaction, row 3 shows

contrast maps for animate stimuli. The high greater than low arousal contrast (row 3-left) for animate stimuli is significant across much of OTC, while the low greater than high arousal contrast (row 3-right) for animate stimuli is not significant within OTC (although it is significant through some of Parietal and Frontal lobes). For inanimate stimuli, the high greater than low arousal contrast is significant within OTC. The low greater than high arousal contrast is not significant within OTC for inanimate stimuli, however (although it is significant through some of Parietal and Frontal lobes).

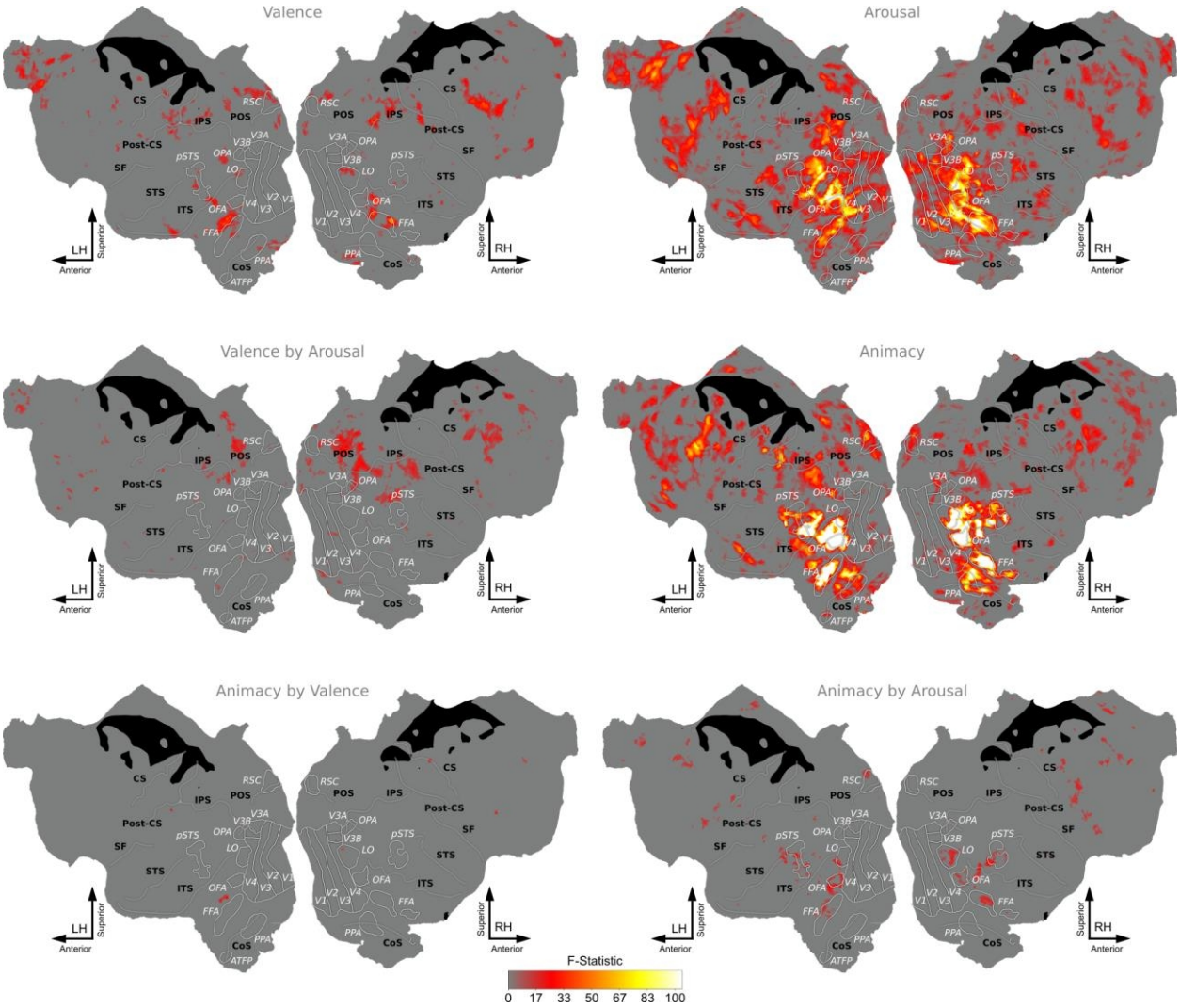


**Figure 3-36. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 4.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 4. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left) were significant within some regions of OTC, Arousal (top-right) was significant across much of OTC, and their interaction (middle-left) was no significant within OTC (although it was within Parietal regions). The main effect of Animacy (middle-right) was significant across much of OTC, the interaction of Animacy by Valence (bottom-left) showed only one small significant cluster within OTC, and Animacy by Arousal (bottom-right) was significant within some regions of OTC.

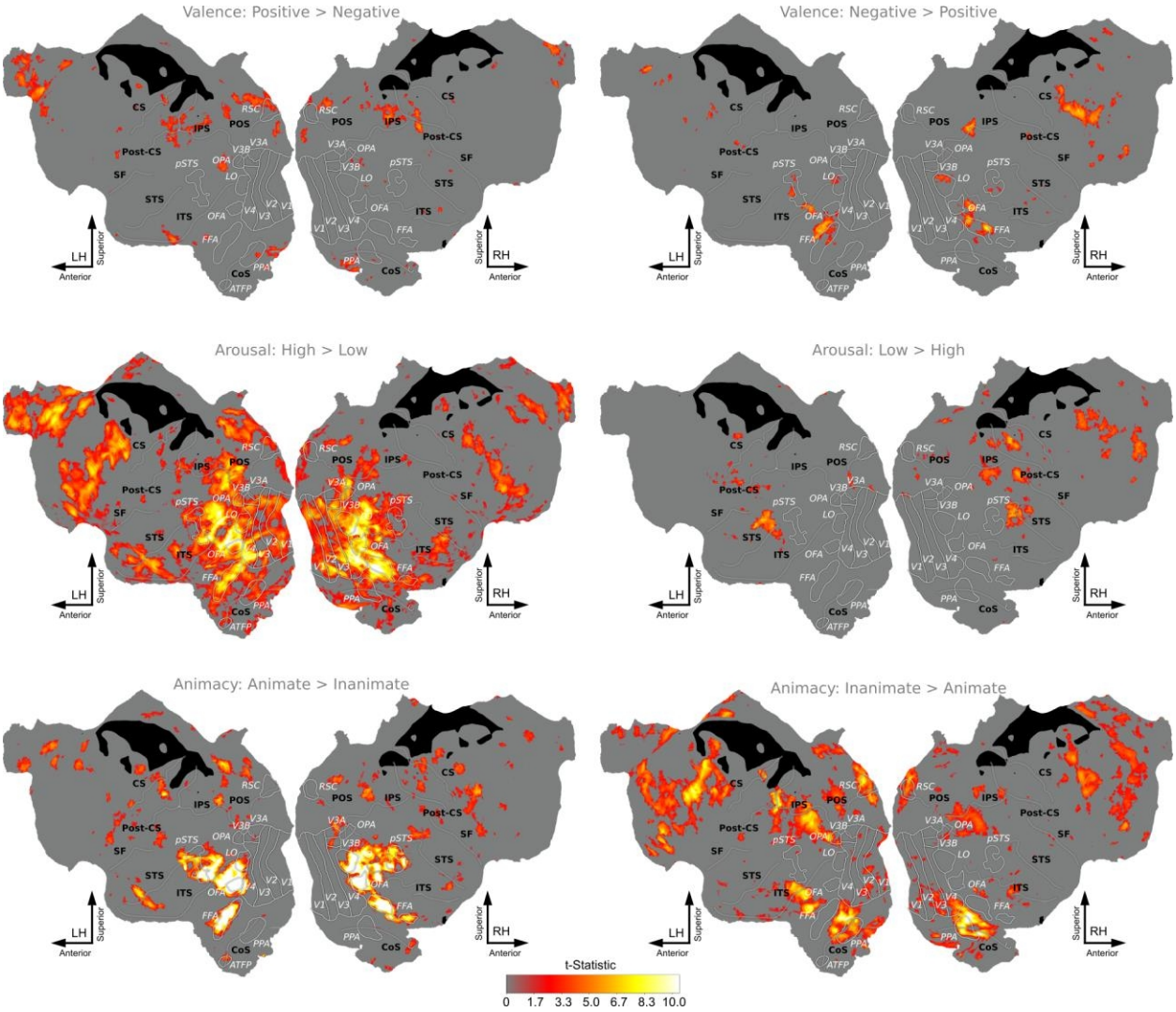


**Figure 3-37. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 4.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 6. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, positive greater than negative (top-left) was significant within limited regions of OTC, but negative greater than positive (top-right) was significant across many regions of OTC. Many regions of OTC showed significant activity to high greater than low arousal (middle-left), while no significant activity for low greater than high arousal was seen in OTC (it was seen in Parietal regions however). Both the animate greater than inanimate (bottom-left) and inanimate greater than animate (bottom-right) both showed significant activation across much of OTC. The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA, and the inanimate greater than animate contrast was significant in regions known to represent inanimate stimuli such as PPA, RSA and OPA.

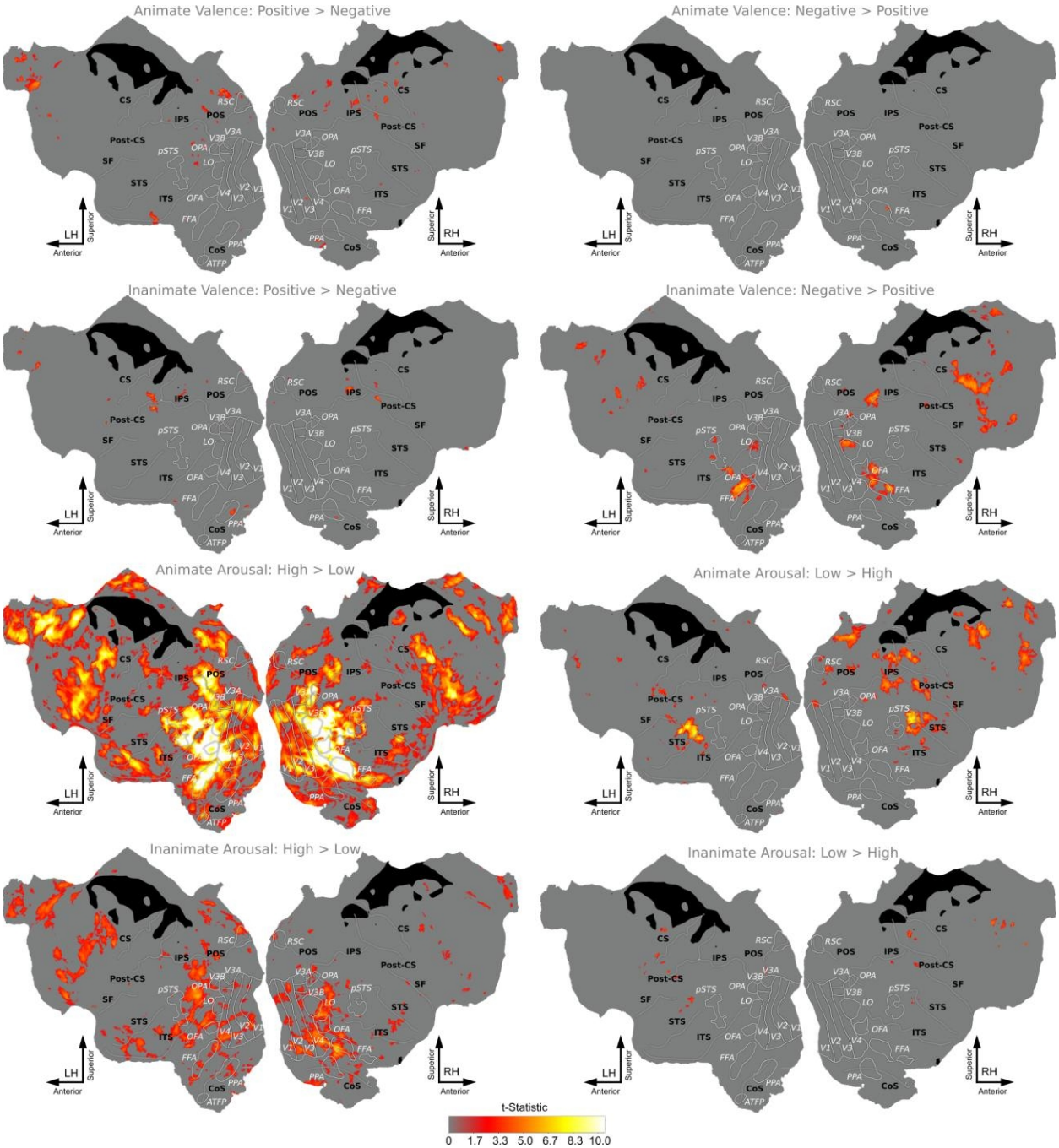


**Figure 3-38. t Contrast Maps Breaking Down the Interactions of Animacy by Valence & Animacy by Arousal for Subject 4.**

Flatmaps for subject 4 showing t-contrasts which breakdown the interactions of Animacy by Valence (rows 1 & 2) and Animacy by Arousal (rows 3 & 4) for animate and inanimate stimuli. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. Row 1 shows both t-contrasts for Valence of just animate stimuli. The positive greater than negative contrast (row 1-left) for animate stimuli shows significant activity within one small cluster in OTC, while the negative greater than positive contrast (row 1-right) for animate stimuli shows one small cluster of significant activation in OTC. For inanimate stimuli in row 2, the positive greater than negative contrast (row 2-left) reveals several small clusters of significant activity in OTC, while the negative

greater than positive contrast (row 2-right) is significant across OTC. For the Animacy by Arousal interaction, row 3 shows contrast maps for animate stimuli. The high greater than low arousal contrast (row 3-left) for animate stimuli is significant across much of OTC, while the low greater than high arousal contrast (row 3-right) for animate stimuli is not significant within OTC (although it is significant through some of Parietal and Frontal lobes). For inanimate stimuli, the high greater than low arousal contrast is significant within OTC, while the low greater than high arousal contrast is not significant anywhere for inanimate stimuli.

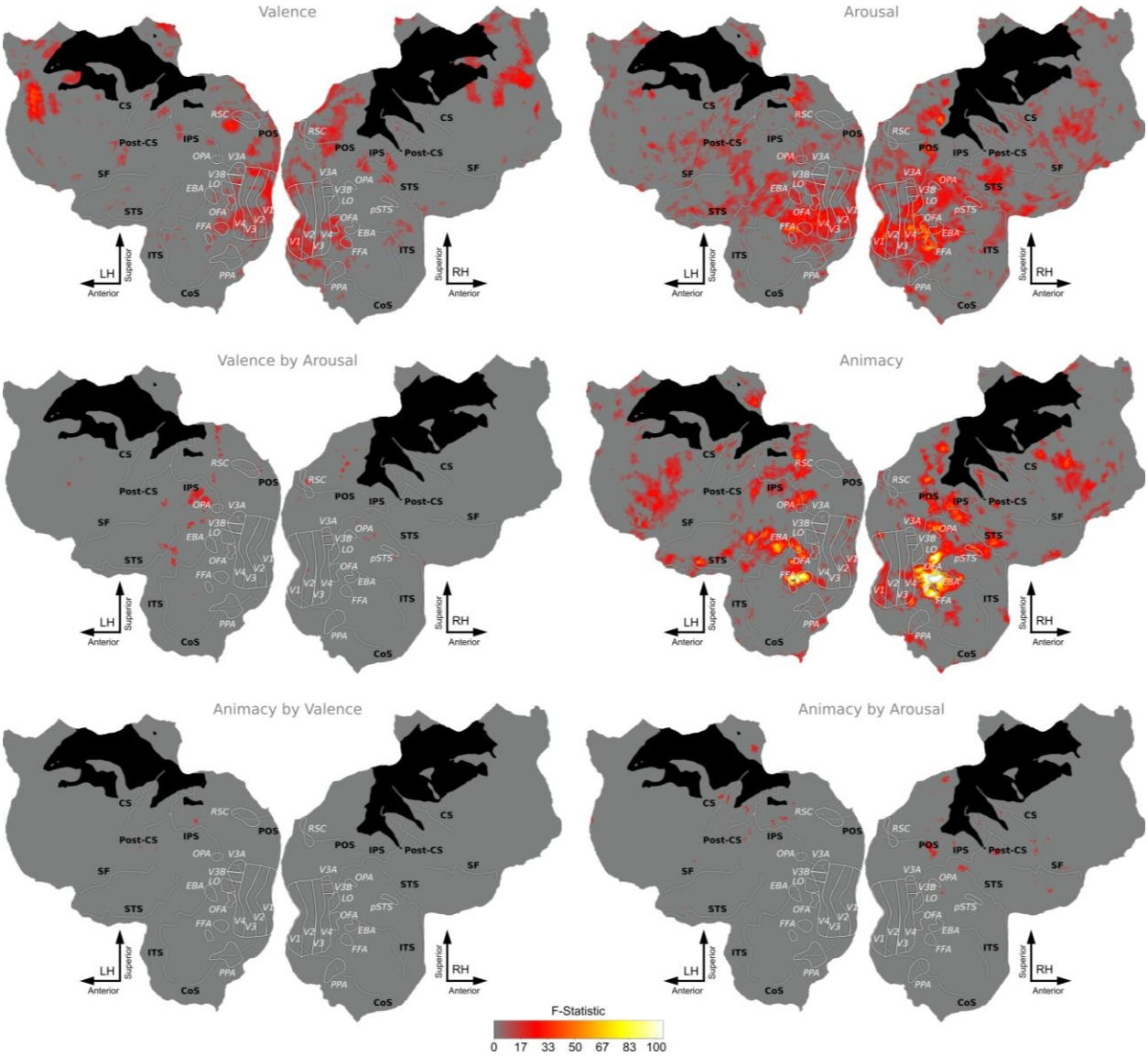


**Figure 3-39. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 5.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 5. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left) and Arousal (top-right) was significant across much of OTC, but their interaction (middle-left) showed no significant activity within OTC. The main effect of Animacy (middle-right) was significant across much of OTC, but the interactions of Animacy by Valence (bottom-left) and Animacy by Arousal (bottom-right) showed no significant activation within OTC.

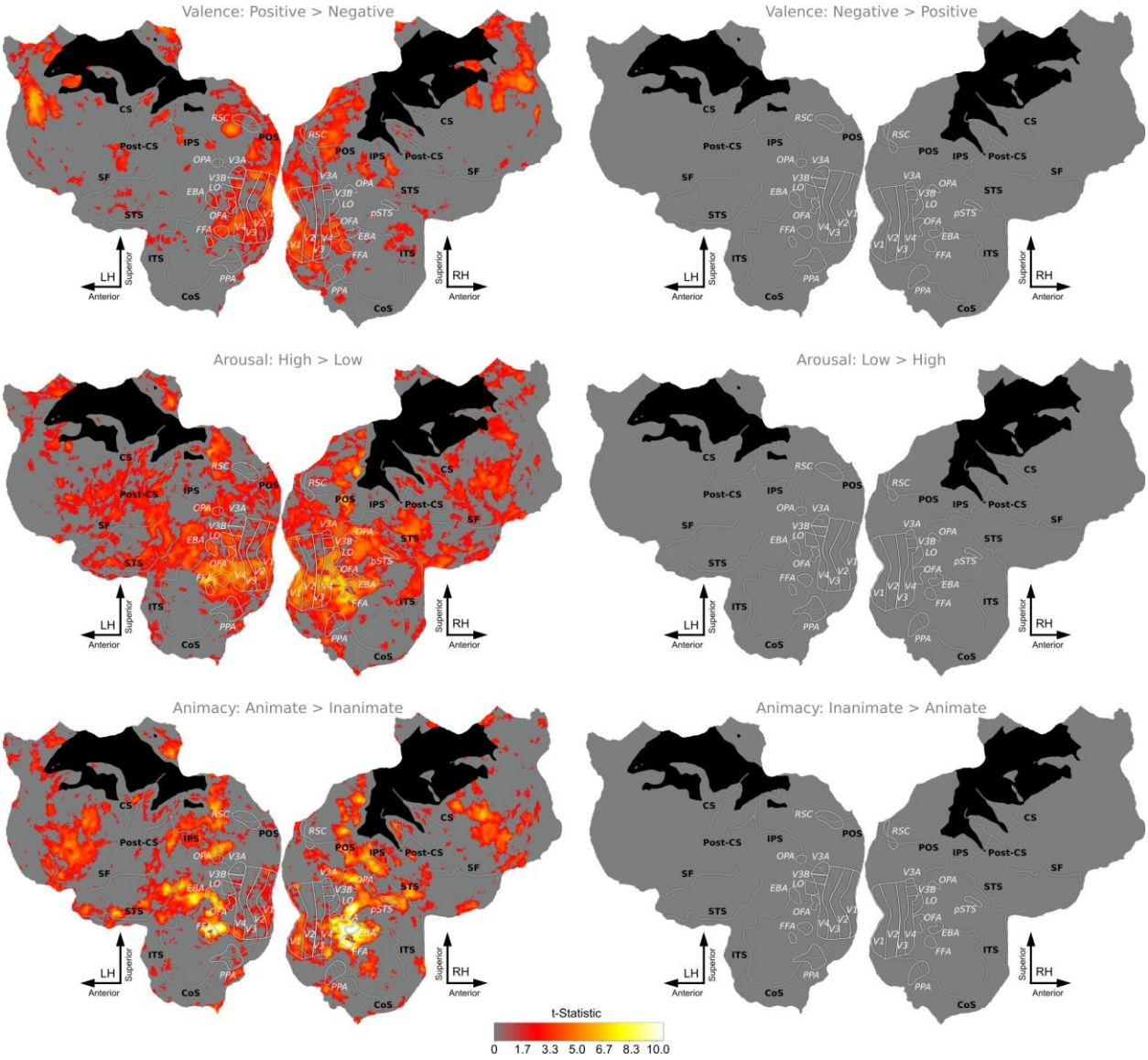


**Figure 3-40. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 5.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 5. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, positive greater than negative (top-left) was significant across much of OTC, but negative greater than positive (top-right) was not significant within OTC. Many regions of OTC showed significant activity to high greater than low arousal (middle-left), while no significant activity for low greater than high arousal was seen in OTC. Only the animate greater than inanimate (bottom-left) showed significant activation across much of OTC. The inanimate greater than animate (bottom-right) showed no significant activation. The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA.

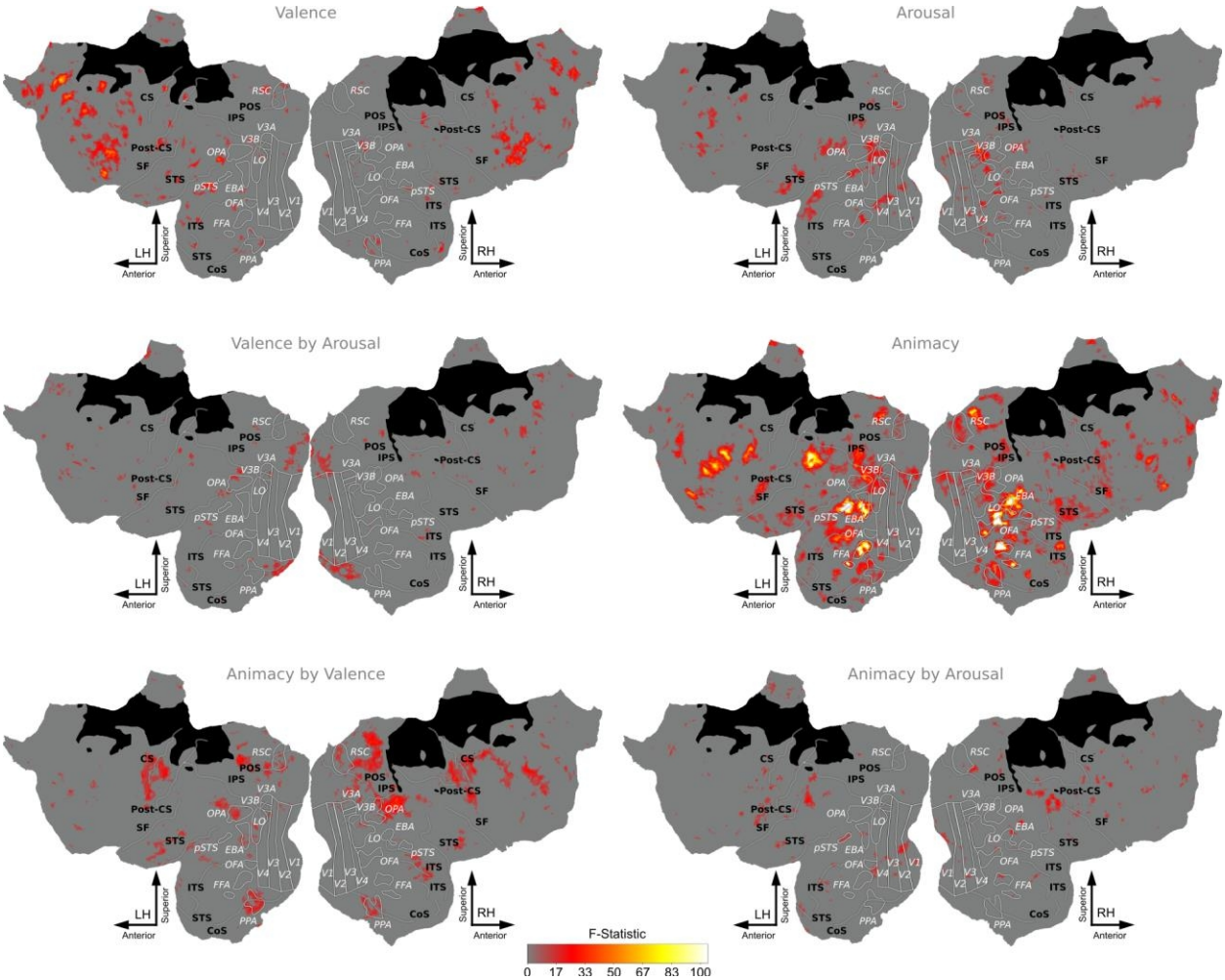


**Figure 3-41. Main Effects and Interactions between Animacy, Valence, & Arousal for Subject 6.**

Flatmaps showing results from 6 contrasts calculated from a univariate SPM analysis for subject 6. An Animate (animate, inanimate) by Valence (negative, neutral, positive) by Arousal (high,low) model consisting of 12 regressors was fit to the BOLD data. A voxel selection mask was created by calculating an F-contrast across all model regressors of interest and selecting those voxels with significant fit ($p < .001$, uncorrected). Secondary F-contrasts of interest were then calculated using this voxel-selection mask, and multiple comparison correction was done using FDR ($q=.05$). Main effects of Valence (top-left) and Arousal (top-right) were significant within some regions of OTC, as was their interaction (middle-left). The main effect of Animacy (middle-right) was significant across much of OTC, and the interaction of Animacy by Valence (bottom-left) and Animacy by Arousal (bottom-right) were significant across some regions within OTC.

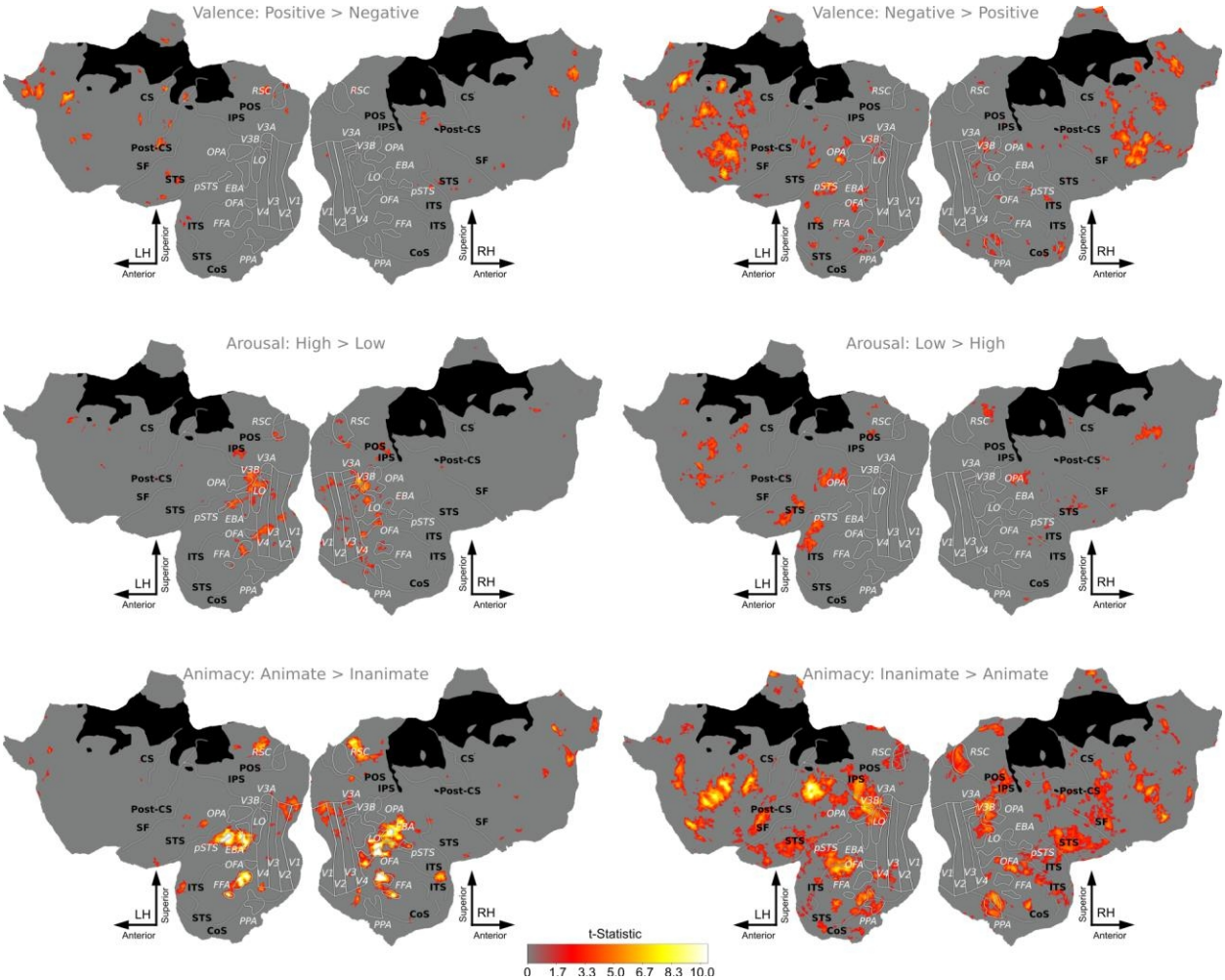


**Figure 3-42. t Contrast Maps for the Main Effects of Animacy, Valence, & Arousal for Subject 6.**

Flatmaps showing univariate SPM t-contrasts of the main effects of Valence (first row), Arousal (second row) & Animacy (third row) for subject 6. Each row shows both t-contrasts for the three main effects whose F-contrasts where shown in the previous figure. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. For the main effect of Valence, positive greater than negative (top-left) was not significant within OTC, but negative greater than positive (top-right) was significant across many regions of OTC. Many regions of OTC showed significant activity to high greater than low arousal (middle-left), while no significant activity for low greater than high arousal was seen in OTC (it was seen in Parietal regions however). Both the animate greater than inanimate (bottom-left) and inanimate greater than animate (bottom-right) both showed significant activation across much of OTC. The animate greater than inanimate contrast was significant in regions known to represent animate stimuli, such as FFA and EBA, and the inanimate greater than animate contrast was significant in regions known to represent inanimate stimuli such as PPA, RSA and LO.

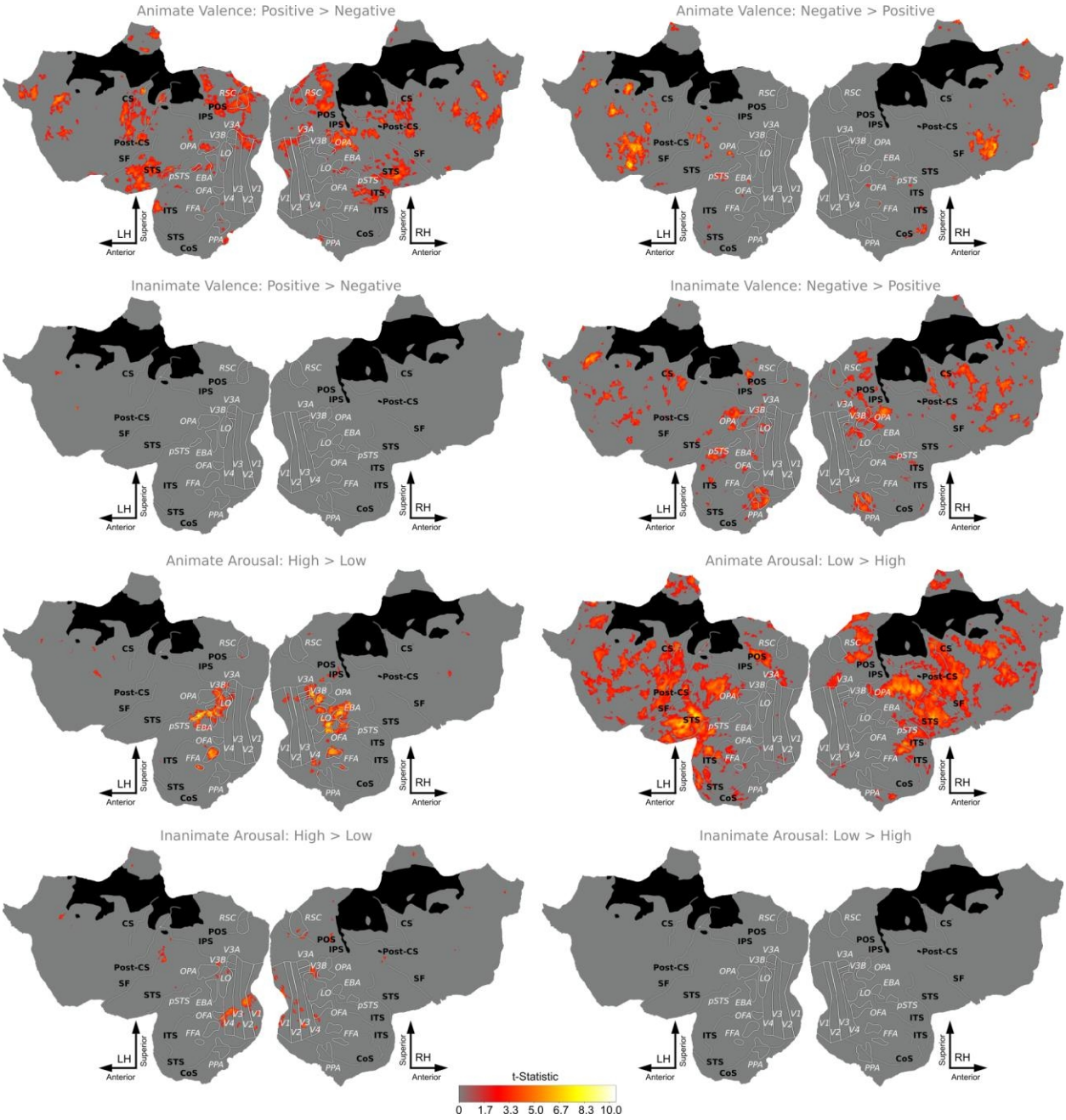


**Figure 3-43. t Contrast Maps Breaking Down the Interactions of Animacy by Valence & Animacy by Arousal for Subject 6.**

Flatmaps for subject 6 showing t-contrasts which breakdown the interactions of Animacy by Valence (rows 1 & 2) and Animacy by Arousal (rows 3 & 4) for animate and inanimate stimuli. Voxel selection and multiple comparisons correction was done in the same manner as was done for the F-contrasts. Row 1 shows both t-contrasts for Valence of just animate stimuli. The positive greater than negative contrast (row 1-left) for animate stimuli shows significant activity within some regions of OTC, while the negative greater than positive contrast (row 1-right) for animate stimuli shows several small clusters of significant activation in OTC. For inanimate stimuli in row 2, the positive greater than negative contrast (row 2-left) reveals no significant activity in OTC, while the negative greater than positive contrast (row 2-right) is

significant across OTC, especially within known inanimate-selective regions such as PPA. For the Animacy by Arousal interaction, row 3 shows contrast maps for animate stimuli. The high greater than low arousal contrast (row 3-left) for animate stimuli is significant across much of OTC, while the low greater than high arousal contrast (row 3-right) for animate stimuli is not significant within OTC (although it is significant through much of Parietal and Frontal lobes). For inanimate stimuli, the high greater than low arousal contrast is significant within early-visual cortex (EVC). The low greater than high arousal contrast is not significant anywhere for inanimate stimuli, however.

# 5. Variance Partitioning of CSVA Model Prediction Accuracy

## Introduction

Variance partitioning (also known as commonality analysis [Nimon & Reio, 2011]) allows for the allocation of explained variance in Y (e.g. BOLD data) across multiple feature spaces that have correlations between their respective features. Explained variance in Y is partitioned into unique variance terms for each feature space, and a shared variance term for all combinations of the features spaces under investigation (de Heer, et al., 2017; Nimon & Reio, 2011). In order to precisely calculate the proportions of variance explained by the combined semantic, valence, and arousal (CSVA) model reported in chapter 2, we here conducted a variance partitioning of the out-of-sample prediction accuracy scores of the CSVA model.

## Methods

To conduct variance partitioning, first an all possible subsets (APS) regression is conducted. In APS, a separate model is fit for every possible combination of feature sets under investigation. For example, in the case of two feature sets, A & B, the following three models would be fit:

1. $t = t_t \cdot t_t + t$, for feature space A
2. $t = t_t \cdot t_t + t$, for feature space B
3. $t = t_t \cdot t_t + t_t \cdot t_t + t$, for the union of feature spaces A or B, denoted AB

were $t$ is the BOLD data, $\beta_t$ , $\beta_t$ are the estimated beta-weights for feature spaces A & B respectively, and $t_t$ , $t_t$ are the design matrices for feature sets A & B respectively. The second step is to calculate the variance explained for all models fit. This can be done using either $t^2$ or $t^2$. The equations for both are given below:

1. $t^2 : t^2 = tttt \quad (\hat{t}, t)^2$
2. $t^2 : t^2 = \frac{\sum_t (t_t - \hat{t}_t)^2}{\sum_t (t_t - \underline{t})^2}$

were $\hat{t}$ is the predicted values of Y for a given model, and $\underline{t}$ is the mean of Y. The third step is to derive a formula that combines the variance explained from the models fit for each unique and combined variance term. In this example there are 2 unique variance terms (denoted $\sigma^2_{t(t)}$ and $\sigma^2_{t(t)}$), and one combined variance term (denoted $\sigma^2_{t(tt)}$), whose formulas are:

1. $\sigma^2_{t(t)} = t^2_{tt} - t^2_t$
2. $\sigma^2_{t(t)} = t^2_{tt} - t^2_t$
3. $\sigma^2_{t(tt)} = t^2_t + t^2_t - t^2_{tt}$

were $t^2_{\ t}$, $t^2_{\ t}$ and $t^2_{\ tt}$ are the amount of variance explained for models of feature spaces A, B and AB respectively, as defined above. Note that $t^{\ 2}$could be used instead of $t^{\ 2}$. See de Heer et al., 2017 for an explanation of variance partitioning using set theory and Nimon & Reio, 2011 for expansion of the unique and shared variance terms for 3 & 4 feature sets.

For this study, we are interested in determining the unique variance explained by the following feature spaces:

(a) $t$ =Semantic features (21 total features)
(b) $t$ =Valence by Arousal features (6 total features)
(c) $t$ =Semantic-Emotion Modifiers (18 total features)
(d) $t \cdot t$ =Semantic by Valence by Arousal features (126 total features)

To do so, 9 models were fitted to each voxel, where the ridge coefficient (lambda) was allowed to vary across voxels. 20 values of lambda were selected from a log-space ranging from 0-1000, and the best fitting lambda was determined using k-fold cross-validation with 50 partitions. This high number of possible lambdas and cross-validation partitions allow for an accurate estimation of the optimal lambda, and thus highest explained variance, per voxel. The following 9 models, which comprise an all possible subsets regression across the 4 feature spaces of interest, were fit to be used in the variance partitioning:

(a) $t$
(b) $t$
(c) $t$
(d) $t \cdot t$
(e) $t + t$
(f) $t + t$
(g) $t + t$
(h) $t + t + t$
(i) $t \cdot t + t$

Note that because we have an interaction model of indicator variables (Semanic by Valence by Arousal), there is no need to fit a full interaction model that would contain the Semantic terms, the Valence by Arousal terms, and the Semantic by Valence by Arousal terms. Both models explain the same amount of variance since the later's design matrix is rank deficient, and of the same rank as the former's.

Then, to calculate the unique and combined variance terms for the four feature sets of interest, the following formulas consisting of linear equations combining the above nine models were used. A system of nine binary equations (i.e. containing zeros and ones) with nine unknowns was defined. Each equation represented one variance partition, and each variable represented one of the nine all possible subsets (APS) regression models. Within each equation a one indicated that the current APS model contained at least one of the features sets in the current partition. To solve this system of equations nine linear equation solvers were used, one for each partition. Each solver was given the system of equations as the X variables. For the Y variables,

a vector of nine 0s was created containing a single 1 value in the location of the equation for the partition to solve. This resulted in nine sets of weights on the APS models (see Table 3-2) which, when multiplied by the $r^2$ values from each APS model regression gives the explained variance ($r^2$) for the given partition. See Table 3-3 below for the system of equations (the X values) given to the solver. Below the equations to calculate each partition from the nine APS model $r^2$ values are given. These equations contain the same information as Table 3-2, just in a more succinct form.:

1. $\sigma^2_{t\,(t)} = t^{\,2}_{(t+t+t)} - t^{\,2}_{(t+t)}$
2. $\sigma^2_{t\,(t)} = t^{\,2}_{(t+t+t)} - t^{\,2}_{(t+t)}$
3. $\sigma^2_{t\,(t)} = t^{\,2}_{(t\cdot t+t)} - t^{\,2}_{(t\cdot t)}$
4. $\sigma^2_{t\,(t\cdot t)} = t^{\,2}_{(t\cdot t+t)} - t^{\,2}_{(t+t+t)}$
5. $\sigma^2_{t\,(t+t)} = t^{\,2}_{(t+t)} + t^{\,2}_{(t+t)} - t^{\,2}_{(t+t+t)} - t^{\,2}_{(t)}$
6. $\sigma^2_{t\,(t+t)} = t^{\,2}_{(t+t)} + t^{\,2}_{(t+t)} - t^{\,2}_{(t+t+t)} - t^{\,2}_{(t)}$
7. $\sigma^2_{t\,(t+t)} = t^{\,2}_{(t+t)} + t^{\,2}_{(t+t)} - t^{\,2}_{(t+t+t)} - t^{\,2}_{(t)}$
8. $\sigma^2_{t\,(t+t+t)} = t^{\,2}_{(t+t+t)} + t^{\,2}_{(t)} + t^{\,2}_{(t)} + t^{\,2}_{(t)} - t^{\,2}_{(t+t)} - t^{\,2}_{(t+t)} - t^{\,2}_{(t+t)}$
9. $\sigma^2_{t\,(t\cdot t+t)} = t^{\,2}_{(t\cdot t+t)} + t^{\,2}_{(t\cdot t)} - t^{\,2}_{(t+t+t)} - t^{\,2}_{(t+t)}$

where the unique variance terms are denoted $\sigma^2_{t\,(ttt\quad)}$ for feature set XXX, and the combined variance terms are denoted $\sigma^2_{t\,(ttt\quad+ttt\quad)}$ for combined variance between feature sets XXX and YYY. Or, see Table 3-2 below for a depiction of these formulae using contrast weights.

Traditionally, the above technique for variance partitioning uses in-sample $t^{\,2}$ or $t^{\,2}$, meaning the explained variance is estimated on the same data that the models are fitted on (Nimon & Reio, 2011). Voxel-wise modeling (VWM) assesses model fit using out-of-sample correlation, or prediction accuracy ($t$). Conducting variance partitioning on out-of-sample data introduces additional technical challenges. When partitioning in-sample variance explained, the sum of all the unique and shared terms is guaranteed to equal the variance explained by the "full model" (the model consisting of the intersection of all feature spaces). However, when partitioning out-of-sample data it is possible that the sum of the unique and shared terms are greater than the explained variance of the full model, due to over-fitting in the estimation data and sampling error. To address this issue, de Heer et al. (2017) developed

> *"a post-hoc correction [that] was applied to the estimated variance explained by each model in each voxel. This correction moved the estimates to the nearest values that produced no nonsensical results. Mathematically, this involved estimating a bias term for the variance explained by each model in each voxel. We began by assuming that the estimated variance explained by some model ($R^2$), $t\hat{}$, is a biased estimate of the true variance explained, $t$ : $t\hat{} = t + t_t$ ." This correction yields one bias term ($t_t$) for each model fit. "Furthermore, because we know that the size of each variance partition must be at least to zero, the set theory equations that give the size of each partition can*

*be used to define seven inequality constraints on the bias terms. Assuming that we want to find the smallest set of bias parameters (in an L2 sense) that produce no nonsensical results, this allowed us to setup a constrained function minimization problem"*:

$ttt \quad \{|t||^2\}ttttttt \qquad tt \quad h_t(t) \geq 0\, ttt \quad t$
$= 1, \ldots t, thttt \quad h_t \; ttt \quad tht \; tttttttttt \qquad tttttttttt \qquad ttt \quad t \; tht \; ttttt \qquad tt \quad ttt$

This correction was applied separately to the estimated variance explained by all models, for each voxel.

**Table 3-2. Contrast Weights Used to Calculate the Nine Variance Partitions of the CSVA model.**

| | $t^2_{(t)}$ | $t^2_{(t)}$ | $t^2_{(t)}$ | $t^2_{(t \cdot t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t+t)}$ | $t^2_{(t \cdot t+t)}$ |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma^2_{t(t)}$ | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 1 | 0 |
| $\sigma^2_{t(t)}$ | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 1 | 0 |
| $\sigma^2_{t(t)}$ | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 1 |
| $\sigma^2_{t(t \cdot t)}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 1 |
| $\sigma^2_{t(t+t)}$ | 0 | 0 | -1 | 0 | 0 | 1 | 1 | -1 | 0 |
| $\sigma^2_{t(t+t)}$ | 0 | -1 | 0 | 0 | 1 | 0 | 1 | -1 | 0 |
| $\sigma^2_{t(t+t)}$ | -1 | 0 | 0 | 0 | 1 | 1 | 0 | -1 | 0 |
| $\sigma^2_{t(t+t+t)}$ | 1 | 1 | 1 | 0 | -1 | -1 | -1 | 1 | 0 |
| $\sigma^2_{t(t \cdot t+t)}$ | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 1 | -1 |

**Table 3-3. System of Equations Given to Linear Solver to Find the Contrast Weights.**

| | $t^2_{(t)}$ | $t^2_{(t)}$ | $t^2_{(t)}$ | $t^2_{(t \cdot t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t)}$ | $t^2_{(t+t+t)}$ | $t^2_{(t \cdot t+t)}$ |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma^2_{t(t)}$ | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| $\sigma^2_{t(t)}$ | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| $\sigma^2_{t(t)}$ | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| $\sigma^2_{t(t \cdot t)}$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| $\sigma^2_{t(t+t)}$ | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\sigma^2_{t(t+t)}$ | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\sigma^2_{t(t+t)}$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\sigma^2_{t(t+t+t)}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\sigma^2_{t(t \cdot t+t)}$ | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |

## Results

Variance partitioning was done to determine the unique and combined contributions to the prediction accuracy of all possible subsets of features spaces of the CSVA model, namely: Unique Semantic, Unique Emotion, Unique Modifiers, Semantic * Emotion, Semantic + Emotion, Semantic + Modifiers, Emotion + Modifiers, Semantic + Emotion + Modifiers, & Semantic * Emotion + Modifiers. Interrogation of these partitions (and combinations thereof) was done to determine the amount of variance in the CSVA model that can be attributed to just emotion, just semantics and various combinations and interactions of semantics and emotion. Figures 3-44 - 3-49 show flatmaps of the following 5 partitions (or combinations), with each figure showing all 5 partitions (and the full CSVA prediction accuracy) for one of the six subjects. Panel letters from the figure correspond with letters below:

A. Unique Emotion

B. CSVA - Semantic: Consists of the sum of the following 5 partitions - Unique Emotion, Modifiers Only, Semantic * Emotion, Emotion + Modifiers, & Semantic * Emotion + Modifiers
C. Unique Semantic
D. CSVA - Emotion: Consists of the sum of the following 5 partitions - Unique Semantic, Unique Modifiers, Semantic * Emotion, Semantic + Modifiers, & Semantic * Emotion + Modifiers
E. Full CSVA Prediction accuracy (for reference)
F. CSVA - Semantic - Emotion: Consists of the sum of the following 3 partitions - Unique Modifiers, Semantic * Emotion, & Semantic * Emotion + Modifiers

The features comprising the Modifier feature space (such as mutilated animals, rotten food, and erotic humans) contain combined semantic and emotional information, and so partitions including the Modifiers feature space will be considered to contain combined semantic and emotion information. Panel A of Figures 3-44 - 3-49 shows that variance explained uniquely by the Emotion model can be seen within OTC across all six subjects, however to a lesser degree, and across fewer voxels, than that of the full CSVA model prediction accuracy (panel E of Figs. 3-44 - 3-49). This unique partition of the Emotion model does not include shared variance with the Semantic or Modifier feature spaces, nor the interaction feature space of Semantics by Emotion. To explore the extent to which variance is explained considering emotion along with it's shared variance and interactions with other feature spaces, we constructed the flatmaps contained within Panel B of Figures 3-44 - 3-49. These flatmaps show all variance explained except from those partitions including the Semantic feature space alone. On these maps, across all six subjects nearly all the OTC voxels that were significantly predicted by the CSVA model are significantly explained, indicating that emotion, and interactions between emotion and semantics, explains much of the variance explained by the CSVA model in OTC. Extensive previous neuroscientific studies have found evidence for the representation of object category within OTC (e.g Epstein & Kanwisher,1998). As these studies would suggest, much of OTC is uniquely predicted by the Semantic model. Perhaps surprising however, is the amount of unique variance explained within OTC voxels by the Semantic model compared with the full CSVA model prediction accuracy. The unique Semantic model explained variance is markedly less in many voxels across all 6 subjects. Removing variance explained by all 4 partitions that include the Emotion Only feature space, shown in Panel D of Figures 3-44 - 3-49, does not reduce prediction accuracy to the same extent as removing variance explained by Semantic only partitions (panel B). This might be expected given the large role semantic information plays in representations of visual stimuli within OTC. Comparing the unique variance explained by Semantics (panel C) with variance explained by Semantics and it's interactions with emotion (panel D) however suggest that the interactions of semantics with emotion explain a large portion of the variance explained by the full CSVA model. Panel E contains flatmaps where variance explained is isolated to only those partitions containing combinations and interactions of

semantic and emotion information. Within these flatmaps in 5 of the 6 subjects, large portions of the voxels significantly explained by the CSVA model still explain significant variance, indicating that in addition to unique variance explained by just the Semantic and Emotion feature spaces, the combination of semantic and emotion information are represented across much of OTC.

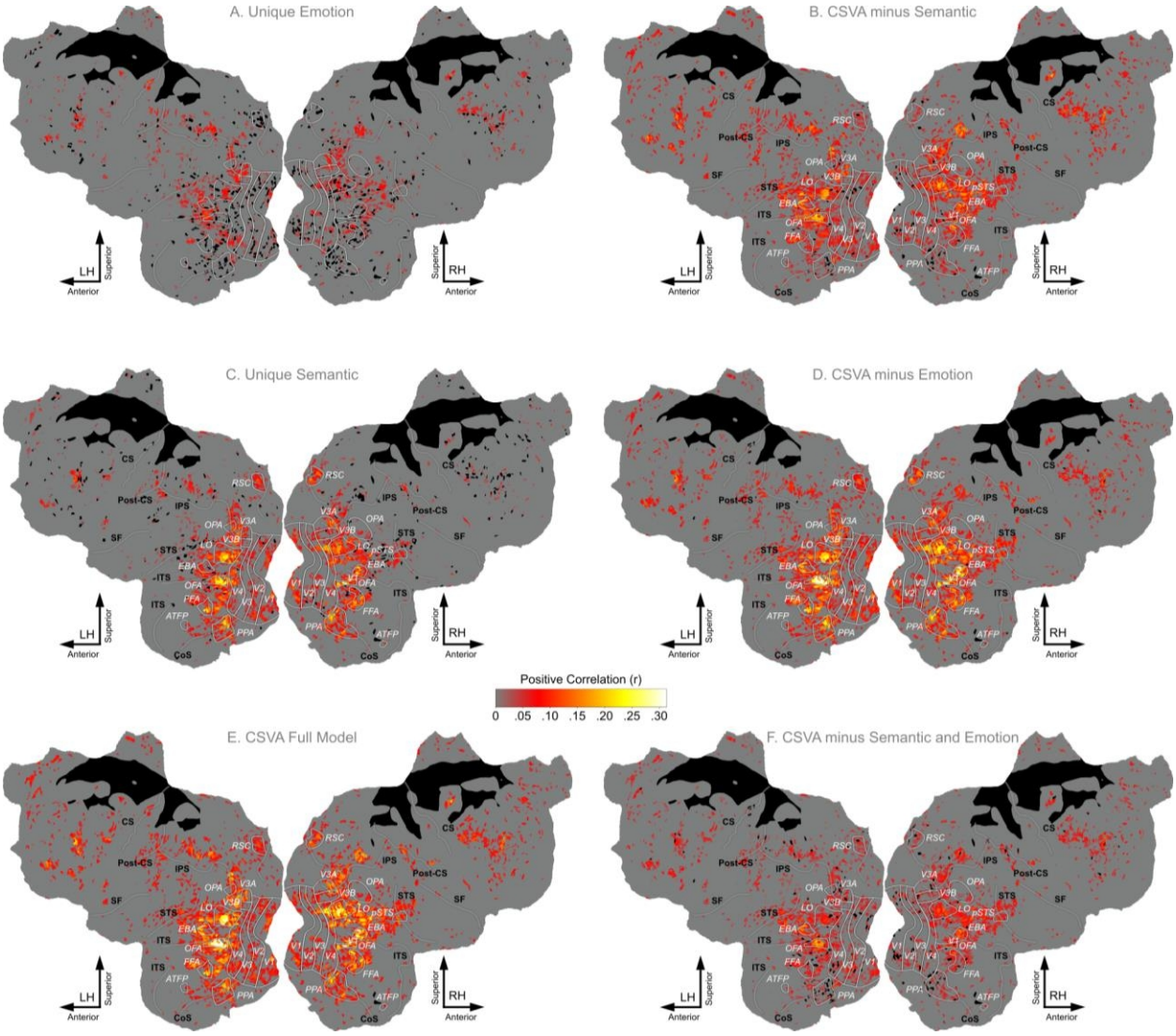


**Figure 3-44: Variance Partitioning of the CSVA model prediction accuracy for Subject 1.**
Variance partitioning was done on the prediction accuracy of the CSVA model, resulting in 9 partitions (see methods for details). These flatmaps of subject 1 show (A,B,C,D,F) the prediction accuracy obtained on 5 different combinations of partitions and (E) the prediction accuracy of the full CSVA model itself. (A) Unique contributions to the explained variance by the Emotion model. Some regions of OTC show significant variance explained by just the emotion model, although to a lesser extent, and across fewer voxels, than the full CSVA model. (B) Variance explained by emotion alone plus combinations of emotion with semantics and modifiers are shown. Specifically, all variance explained by the CSVA model except those partitions that included the semantic model (Semantic Only, Semantic + Emotion, Semantic + Modifiers, & Semantic + Emotion + Modifiers). Nearly all OTC regions that were significantly predicted by the CSVA model are explained by this combination of partitions, indicating that the emotion plus the interaction between emotion and semantics explains much of the variance explained by the CSVA model in OTC. (C) Unique contributions to the explained variance by the Semantic model. As previous work has suggested, much of OTC is well predicted by this model, although not to the same extent as the CSVA model itself. (D) Variance explained by semantic alone plus combinations of semantic with emotion and modifiers are shown. Specifically, all variance explained by the CSVA model except those partitions that included the emotion model (Emotion Only, Semantic + Emotion, Emotion + Modifiers, & Semantic + Emotion + Modifiers). (E) Prediction accuracy of the full CSVA model, shown for reference. (F) Variance explained by the CSVA model removing all

partitions including the Semantic or Emotion model alone. Specifically, this shows the sum of the 3 interaction partitions (Modifiers Only, Semantic x Emotion, Semantic x Emotion x Modifiers). Large portions of the voxels significantly explained by the CSVA model still explain significant variance when only looking only at the interactions between semantics and emotion, suggesting the combination of semantics and emotion are represented across much of OTC.

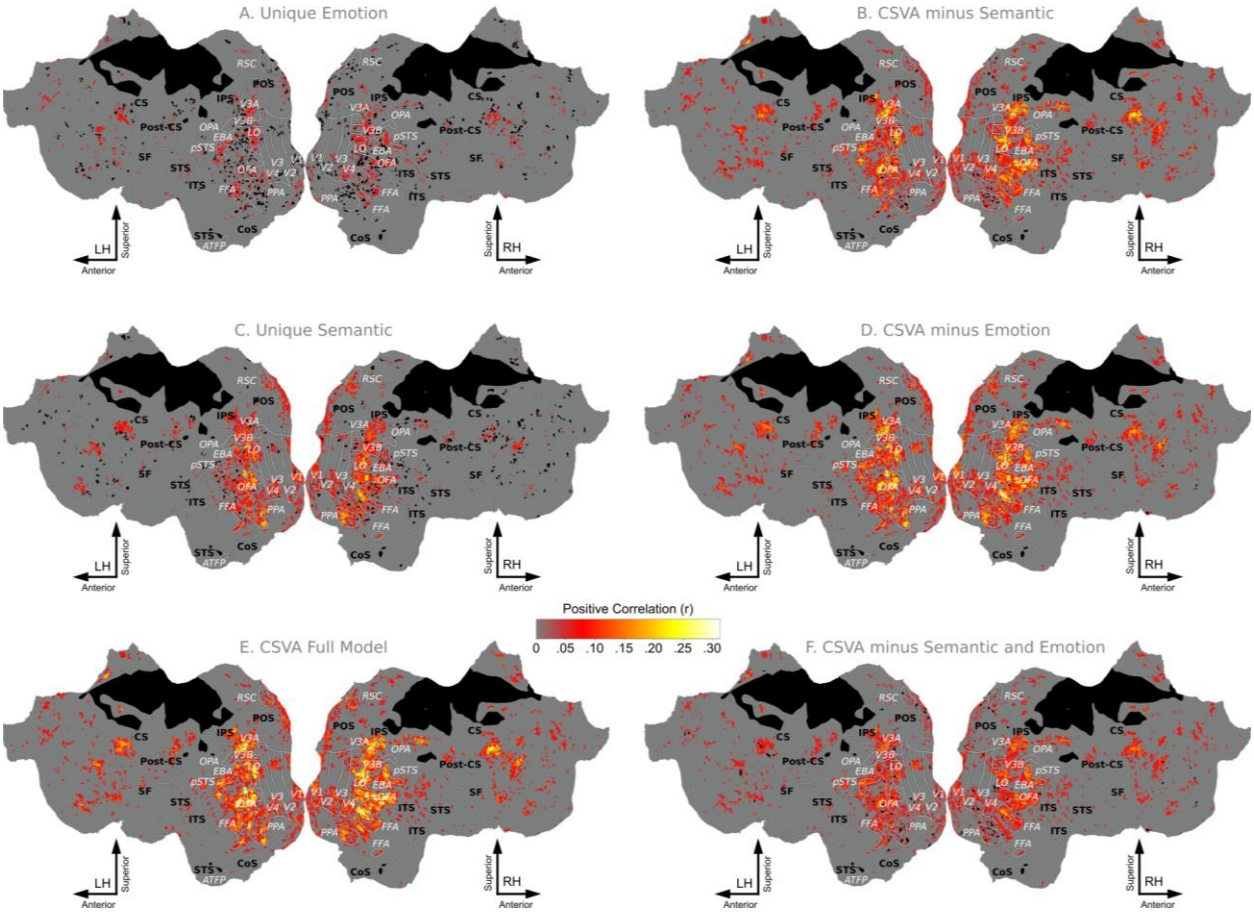


**Figure 3-45: Variance Partitioning of the CSVA model prediction accuracy for Subject 2.**
Flatmaps for Subject 2 are the same as those described for subject 1 in Figure 3-44. (A) Scattered regions within OTC show variance explained uniquely by Emotion. (B) Nearly all regions of OTC whose voxels are significantly explained by the CSVA model are also significantly explained when removing contributions from just Semantics, indicating emotion and it's interactions with Semantics are represented within OTC. (C) Although many of the same voxels show significant variance explained uniquely by Semantics, the amount of variance within these voxels is much less than the full CSVA model itself. (D) Removing all partitions including Emotion, but not interactions with emotion, slightly reduces the amount of variance explained in much of OTC. (E) Prediction accuracy of the full CSVA model, for comparison. (F). Only the interactions between Semantics, Emotion and the Modifiers still explain significant variance throughout much of OTC explained by the full CSVA model, indicating the interaction of semantics and emotion are crucial to representations within OTC.

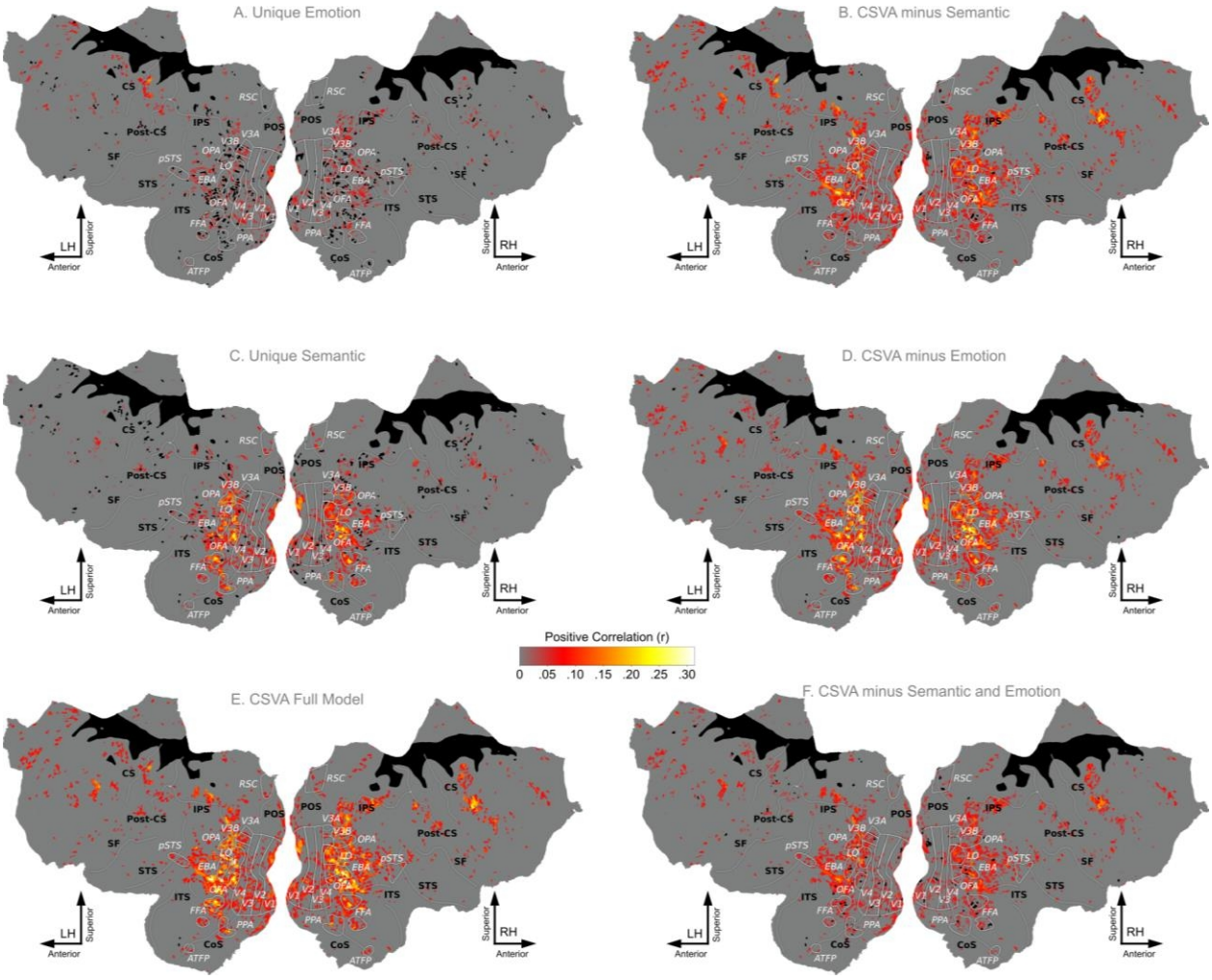


**Figure 3-46: Variance Partitioning of the CSVA model prediction accuracy for Subject 3.**
Flatmaps for Subject 3 are the same as those described for subject 1 in Figure 3-44. (A) Scattered regions within OTC show variance explained uniquely by Emotion. (B) Nearly all regions of OTC whose voxels are significantly explained by the CSVA model are also significantly explained when removing contributions from just Semantics, indicating emotion and it's interactions with Semantics are represented within OTC. (C) Although many of the same voxels show significant variance explained uniquely by Semantics, the amount of variance within these voxels is much less than the full CSVA model itself. (D) Removing all partitions including Emotion, but not interactions with emotion, slightly reduces the amount of variance explained in much of OTC. (E) Prediction accuracy of the full CSVA model, for comparison. (F). Only the interactions between Semantics, Emotion and the Modifiers still explain significant variance throughout much of OTC explained by the full CSVA model, indicating the interaction of semantics and emotion are crucial to representations within OTC.

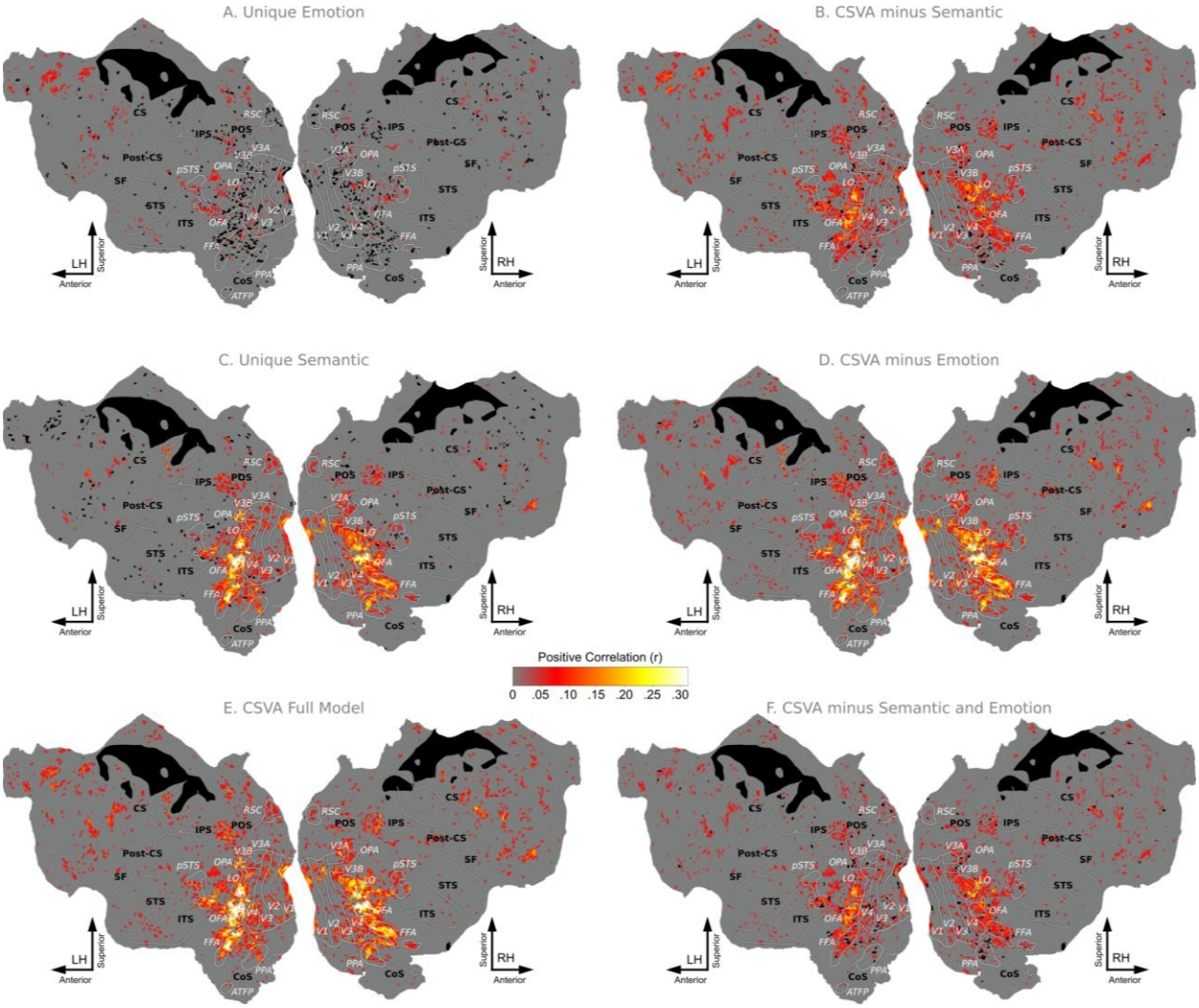


**Figure 3-47: Variance Partitioning of the CSVA model prediction accuracy for Subject 4.**
Flatmaps for Subject 4 are the same as those described for subject 1 in Figure 3-44. (A) Scattered regions within OTC show variance explained uniquely by Emotion. (B) Nearly all regions of OTC whose voxels are significantly explained by the CSVA model are also significantly explained when removing contributions from just Semantics, indicating emotion and it's interactions with Semantics are represented within OTC. (C) Although many of the same voxels show significant variance explained uniquely by Semantics, the amount of variance within these voxels is much less than the full CSVA model itself. (D) Removing all partitions including Emotion, but not interactions with emotion, slightly reduces the amount of variance explained in much of OTC. (E) Prediction accuracy of the full CSVA model, for comparison. (F). Only the interactions between Semantics, Emotion and the Modifiers still explain significant variance throughout much of OTC explained by the full CSVA model, indicating the interaction of semantics and emotion are crucial to representations within OTC.

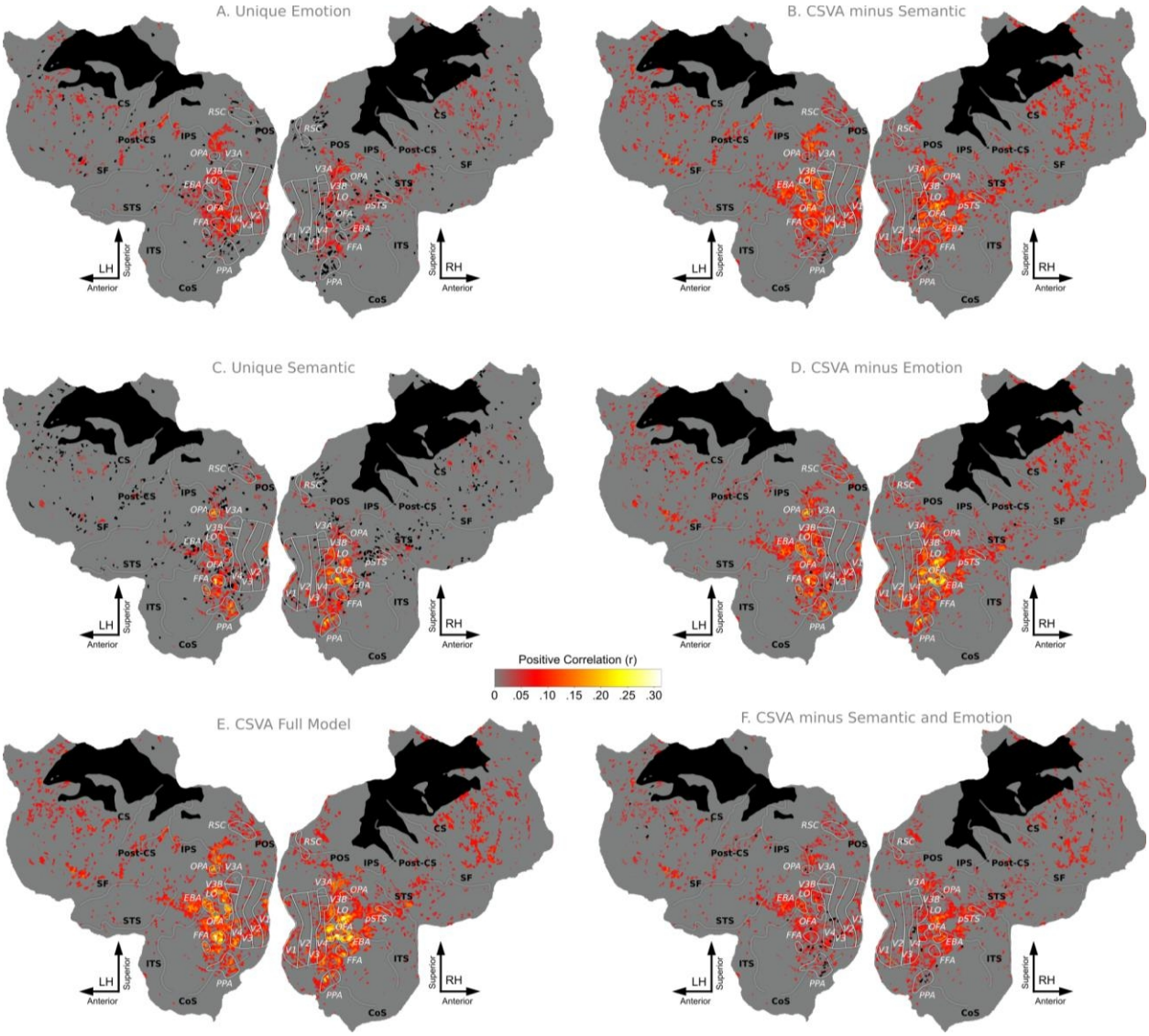


**Figure 3-48: Variance Partitioning of the CSVA model prediction accuracy for Subject 5.**
Flatmaps for Subject 5 are the same as those described for subject 1 in Figure 3-44. (A) Many regions within OTC show variance explained uniquely by Emotion, with more voxels in the left hemisphere explaining significant variance. (B) Nearly all regions of OTC whose voxels are significantly explained by the CSVA model are also significantly explained when removing contributions from just Semantics, indicating emotion and it's interactions with Semantics are represented within OTC. (C) Although many of the same voxels show significant variance explained uniquely by Semantics, the amount of variance within these voxels is much less than the full CSVA model itself. (D) Removing all partitions including Emotion, but not interactions with emotion, slightly reduces the amount of variance explained in much of OTC. (E) Prediction accuracy of the full CSVA model, for comparison. (F). Only the interactions between Semantics, Emotion and the Modifiers still explain significant variance throughout much of OTC explained by the full CSVA model, indicating the interaction of semantics and emotion are crucial to representations within OTC.

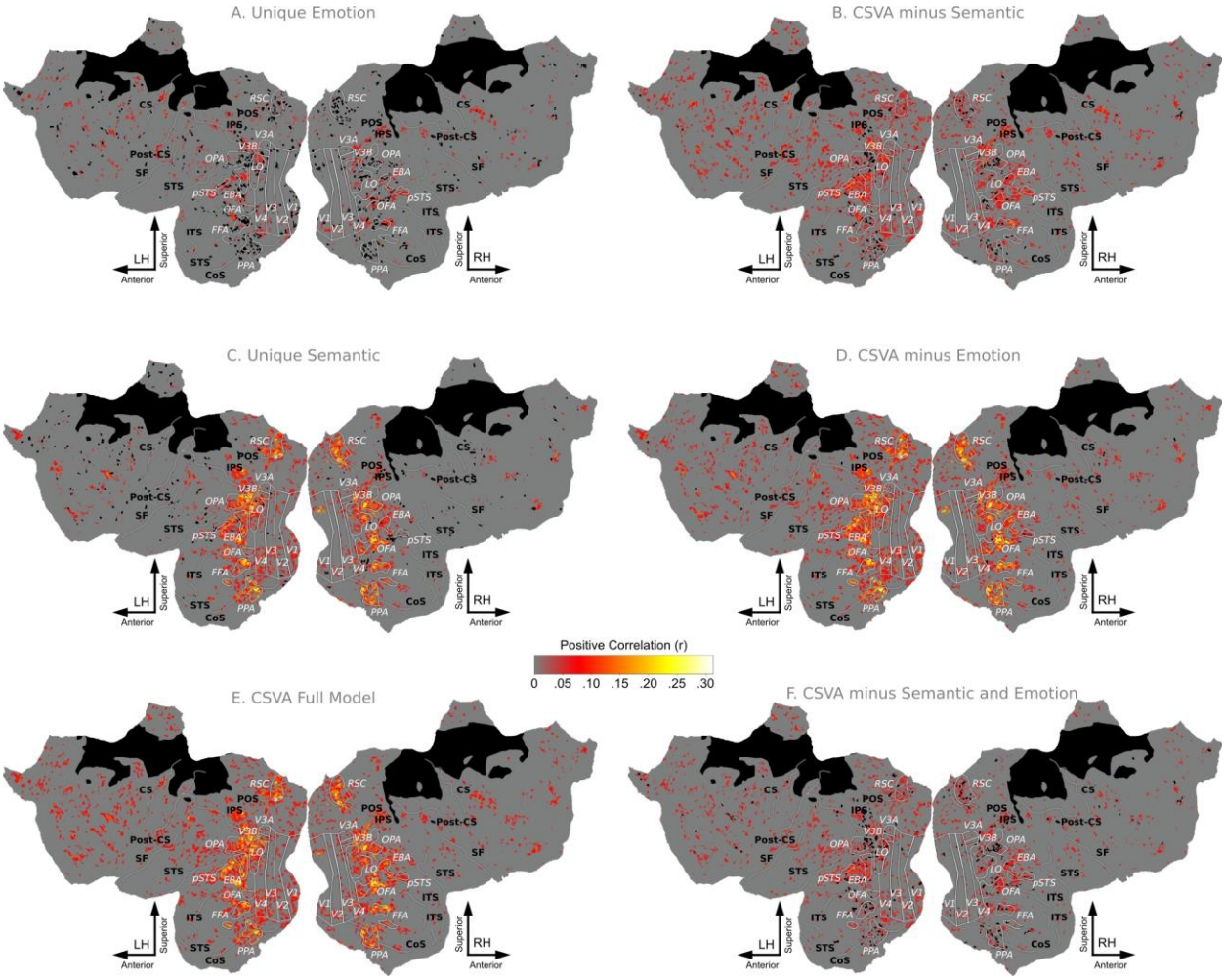


**Figure 3-49: Variance Partitioning of the CSVA model prediction accuracy for Subject 6.**
Flatmaps for Subject 6 are the same as those described for subject 1 in Figure 3-44. (A) Scattered regions within OTC show variance explained uniquely by Emotion. (B) Nearly all regions of OTC whose voxels are significantly explained by the CSVA model are also significantly explained when removing contributions from just Semantics, indicating emotion and it's interactions with Semantics are represented within OTC. (C) Although many of the same voxels show significant variance explained uniquely by Semantics, the amount of variance within these voxels is much less than the full CSVA model itself. (D) Removing all partitions including Emotion, but not interactions with emotion, slightly reduces the amount of variance explained in much of OTC. (E) Prediction accuracy of the full CSVA model, for comparison. (F). Only the interactions between Semantics, Emotion and the Modifiers still explain significant variance throughout some of OTC explained by the full CSVA model.

# 6. Principal Components Analysis (PCA) on CSVA Model Voxel Feature Weights and Behavioral Response Analysis within Occipital-Temporal Cortex (OTC) Subregions

## Introduction

Principal components analysis (PCA) was done on the CSVA model feature weights of all voxels within OTC where the CSVA model both significantly predicted the BOLD signal, and predicted more variance than the Semantic Only model, as reported in chapter 2. This included voxels from Early Visual Cortex (EVC), defined as retinotopic regions V1, V2, V3, & V4, as well as voxels from anterior occipital and ventral temporal cortex, defined as all OTC voxels not in EVC (referred to as non-EVC OTC from here on). Numerous studies have shown that regions in EVC are well modeled using Gabor wavelet models (Kay et al., 2008), while regions within non-EVC OTC are well modeled using a Semantic model. Thus the contributions to the overall spatial and representational structure of the OTC PCA from these two regions is likely to be very different than that of the two regions considered separately. In order to disentangle the contributions of these two regions, two additional PCAs were conducted on voxel feature weights from EVC and non-EVC OTC separately.

## Methods

### PCA

This analysis was done on the 4 ROIs listed above (EVC, non-OTC EVC, OFC, & non-OFC Frontal) exactly as described in two supplemental methods sub-sections of chapter 2 within the "fMRI Data Modeling" section of the METHOD DETAILS. They were titled:

1. "PCA of CSVA model feature weights"
2. "Interpreting the top 3 group-level PCs"

Statistical tests were done as described in the two subsections of the QUANTIFICATION AND STATISTICAL ANALYSIS section titled:

1. "Principal component significance testing"
2. "Relating PCs to hypothetical dimensions"

### Behavioral Analysis

This analysis was done on the 4 ROIs listed above (EVC, non-OTC EVC, OFC, & non-OFC Frontal) exactly as described in a supplemental methods subsection of chapter 2 within the "fMRI Data Modeling" section of the METHOD DETAILS. It was titled:

1. "Prediction of behavioral responses"

Statistical tests were done as described in the sub-section of the QUANTIFICATION AND STATISTICAL ANALYSIS section titled:

1. “Predicting behavioral responses from OTC tuning”

## Results

### PCA

Scree plots displaying the amount of variance explained by the top 10 group EVC CSVA PCs are shown in panel A of Figure 3-50. The top 3 group EVC PCs explain significantly more variance within EVC voxel feature weights than can be explained by the top 3 PCs from a PCA done on the stimulus features alone. Thus subsequent analysis of EVC PCs will be limited to the top 3 group PCs. Correlations between the top 3 group and single-subject EVC CSVA PCs are shown in panel B of Figure 3-50. While all correlations are significant (ps<.05), similarity between the group and single-subject PC loadings across subjects is lower, and less consistent, than that of the OTC PCs, suggesting that the representational space captured by the CSVA model within EVC varies across subjects. PC loadings from the top 3 group PCs of the EVC and OTC CSVA PCA were correlated to determine similarity of the representational space between the two regions, shown in Panel C of Figure 3-50. As was expected, similarity between the top 3 EVC and OTC CSVA PCs was low (PC1 r=.41, PC2 r=.49, PC3 r=.48), and confusability between the top 2 PCs across the two models was high (PC1 r=.70, PC2 r=.56). Flatmaps showing the scores of of the top 3 EVC CSVA PCs within EVC voxels are shown in Figure 3-51. These maps reveal a moderate consistency across subjects in spatial structure as defined by the top 3 EVC CSVA PCs.

Within non-EVC OTC, the top 3 group PCs explain significantly more variance than stimulus feature PCs, as shown in panel D of Figure 3-50. Additionally, the top PC explains approximately 33% of the variance in non-EVC OTC feature weights, compared with approximately 10% of EVC feature weight variance explained by the top EVC PC. Panel E shows that similarity between the top 3 group and single subject non-EVC OTC CSVA PCs is high, and significantly correlated (ps<.01) across all 3 PCs and 6 subjects. Furthermore, as seen in panel F of Figure 3-50, the correlations between the top 3 non-EVC OTC and full OTC group CSVA PCs are nearly perfect(rs>.99), and uncorrelated with the remaining other top 3 PCs (rs<.05). Thus, as might be expected, the spatial structure of the top 3 group CSVA PCs within non-EVC OTC (Fig. 3-52) and all of OTC is extremely similar, as can be seen in the flatmaps of PC scores in Figure 2-4 and Figure 2-S5. Taken together, these results indicate that the representational space and spatial structure recovered by the PCA on OTC voxels done in chapter 2 was largely driven by covariance across non-EVC OTC voxel feature weights, and not EVC voxel feature weights.

## Behavioral Response Analysis

Appropriately responding to external stimuli containing emotional content is crucial to survival. Thus, an important question is whether the stimulus information represented within OTC could be used to select an appropriate behavioral response to that stimulus. To address this question, OTC tuning to emotional stimuli, quantified via PC loadings from PCA on CSVA feature weights within OTC, was used to predict appropriate behavioral responses to the stimuli (reported in chapter 2). A separate set of subjects selected the behavioral responses appropriate to the content of each image (see chapter 2 methods for details), and a linear regression was used to predict out-of-sample explained variance by the scores of each image on a varying number of OTC CSVA PCs (1 to 21). To determine the separate contributions that EVC and non-EVC OTC make to the predictions of appropriate behavioral responses we repeated the same analysis as was done on OTC PCs in chapter 2 on both these regions.

Tuning to the CSVA model within EVC (as quantified by EVC CSVA PCs) predicted behavioral responses better than PCs from the stimulus features themselves for only the top 8 PCs (Fig. 3-53 panel A), compared with all 21 OTC PCs considered in chapter 2 (Fig. 2-6, panel B). In addition to the CSVA model, PCs of models with Semantic Only, Emotion Only, and Gabor wavelet features were also used to predict out-of-sample variance in behavioral responses. As can be seen in panel B of Figure 3-53, the Gabor model PCs did a poor job of predicting behavioral responses, which suggests that although EVC is well predicted by a Gabor wavelets model, that tuning to low-level image structural features is insufficient to guide behavior. Additionally, both the Semantic Only and Emotion Only models outperform the Gabor model in predicting behavior, and the Emotion Only model performs particularly well at low levels of dimensionality, while the semantic model is not significantly different from the CSVA model at any level of dimensionality. Unlike in OTC as a whole, tuning in EVC to the Emotion Only model at its maximal prediction of behavior (at n=6 PCs) is not significantly less than that achieved by the CSVA model using an equivalent number of components.

In contrast to tuning in EVC, tuning in non-EVC OTC explained nearly identical amounts of variance in behavioral responses compared with tuning in OTC as a whole. All 21 non-EVC OTC CSVA PCs considered explained significantly more variance than stimulus feature PCs and Gabor model PCs, as did OTC CSVA PCs. The Emotion Only and Semantic Only model PCs also perform well at low dimensionality, but are then significantly outperformed by non-EVC OTC CSAV PCs at their respective maximal prediction of behavior (n=6 and n=21 PCs). The extreme similarity (rs>.99) between the top 3 CSVA model PCs from the non-EVC OTC PCA and the entire OTC PCA was here shown to extend across higher dimensionality when considering the PCs ability to predict appropriate behavioral responses to emotional stimuli. These findings suggest that, as with the PCA, the contributions of EVC representations did little to alter the explained variance in behavioral responses within OTC, and when excluding EVC the results are nearly identical between non-EVC OTC and the entire OTC.

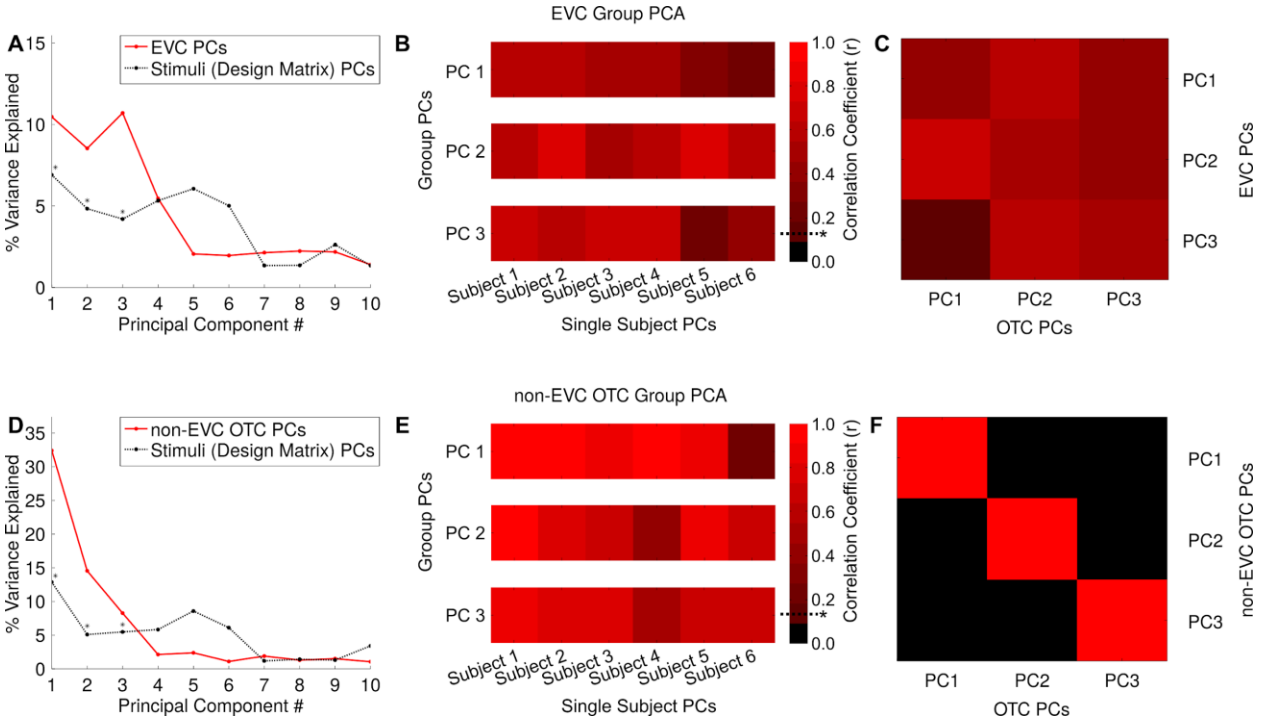


**Figure 3-50. Results of PCA on voxel feature weights within Early Visual Cortex (EVC) and non-EVC OTC.**

(A-C) A group-level principal components analysis (PCA) was conducted on the CSVA model feature weights across all Early Visual Cortex (EVC) voxels where model fit was significant and better than that of the Semantic Only model. (A) The scree plot shows the amount of variance explained by each of the top ten PCs (in red). PCs from a PCA analysis conducted on stimulus features (using the combined design matrix from all 6 subjects) are shown in black. Asterisks indicate group PCs that explain significantly more variance than the stimulus PCs (jackknife test, $p < 10^{-8}$ ), see chapter 2 Methods for details. (B) Correlations between the top 3 group PCs within EVC, and the top 3 PCs from single subject PCAs within EVC. The significant correlation coefficients ($r > .14$, indicated by the dotted line and asterisk on the colorbar) between each of the group PCs and the corresponding PCs from the single subject PCA for each subject ($ps < 10^{-8}$) indicate a shared representational structure across subjects. (C) Correlations between the loadings of the top 3 PCs from PCAs done on EVC voxels and all OTC voxels. Lack of similarity between the PCs suggest the representational space of EVC differs drastically from that of the OTC as a whole, even though the PCs on all OTC voxels contain those within EVC. (D-F) A second group-level principal components analysis (PCA) was conducted on the CSVA model feature weights across all OTC voxels not contained within EVC, where model fit was significant and better than that of the Semantic Only model. (D) Scree plots showing variance explained by the top 10 PCs for the non-EVC OTC PCs. The top 3 PCs explain significantly more variance than do the stimulus feature PCs. Of note is the increased amount of variance the non-EVC OTC PCs explain relative to the EVC PCs. (E) Similarity between the group and single subject non-EVC OTC PCs is significant, and higher than that seen in the EVC PCs. (F) Correlations between the non-EVC OTC, and all OTC PCs is extremely high for the top 3 PCs, suggesting that removing the EVC voxels from the PCA does not meaningfully alter the results.

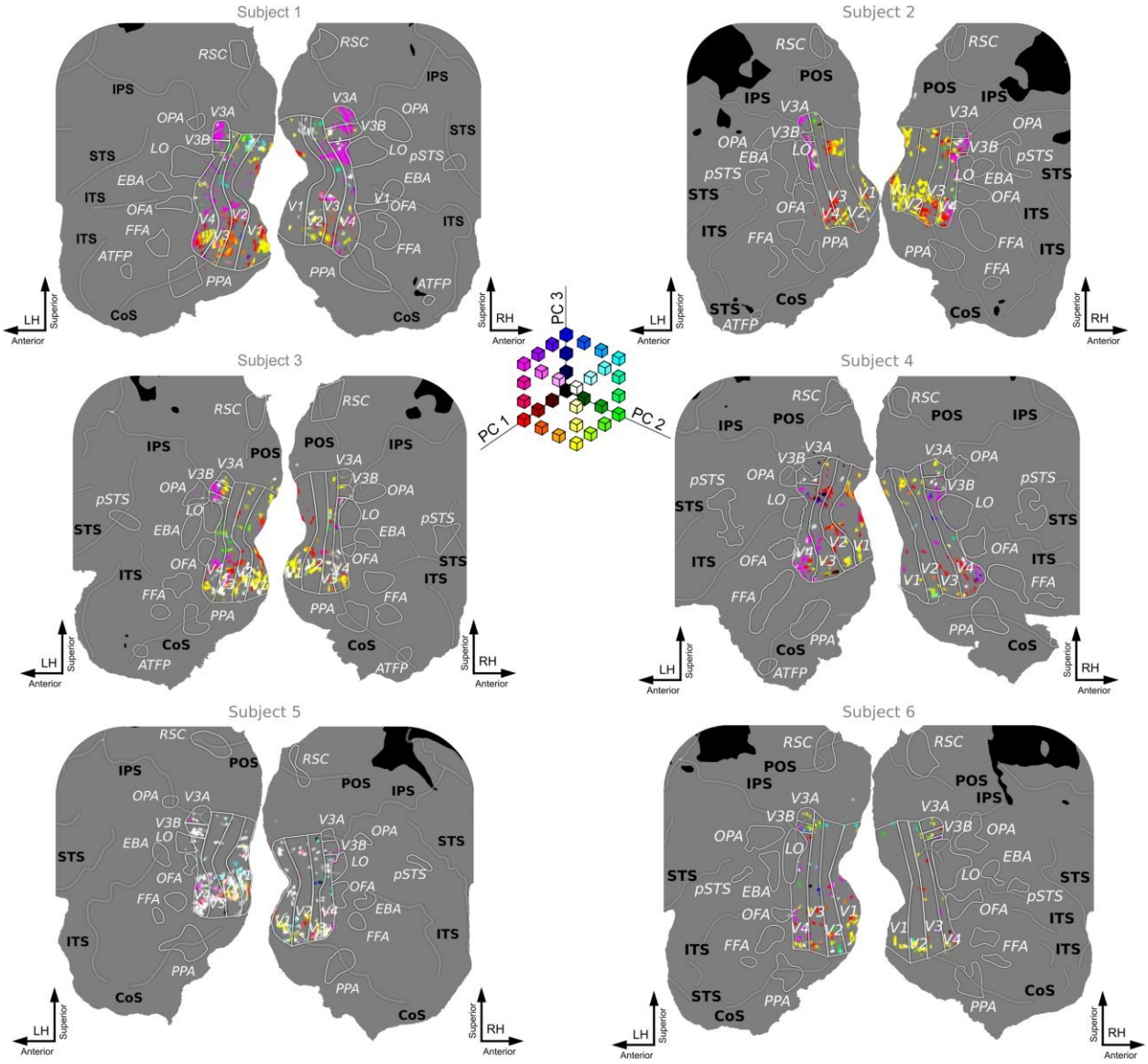


**Figure 3-51. Top 3 Group PC scores within Early Visual Cortex (EVC).**

A principal components analysis (PCA) was conducted on CSVA model feature weights for all EVC voxels where CSVA model fit was significant and superior to that of the Semantic Only model. PC scores were calculated as the product of CSVA feature weights for a given voxel by feature loadings for each PC. Here, a RGB color space is used to map PC scores onto cortex (red= scores on PC1, green = scores on PC2, blue = scores on PC3). PC scores are thresholded at 6 standard deviations above and below 0 with values beyond the threshold given the maximal (or minimal) color channel value. Moderately consistent spatial structure of voxel-wise tuning to the top three group PCs is observed across subjects. Note: Areas where MRI data was not acquired are shown in black. Both voxels where the CSVA model did not fit significantly and those where the CSVA model fit significantly but did not outperform the semantic only model were excluded from the PCA (these voxels are shown in grey).

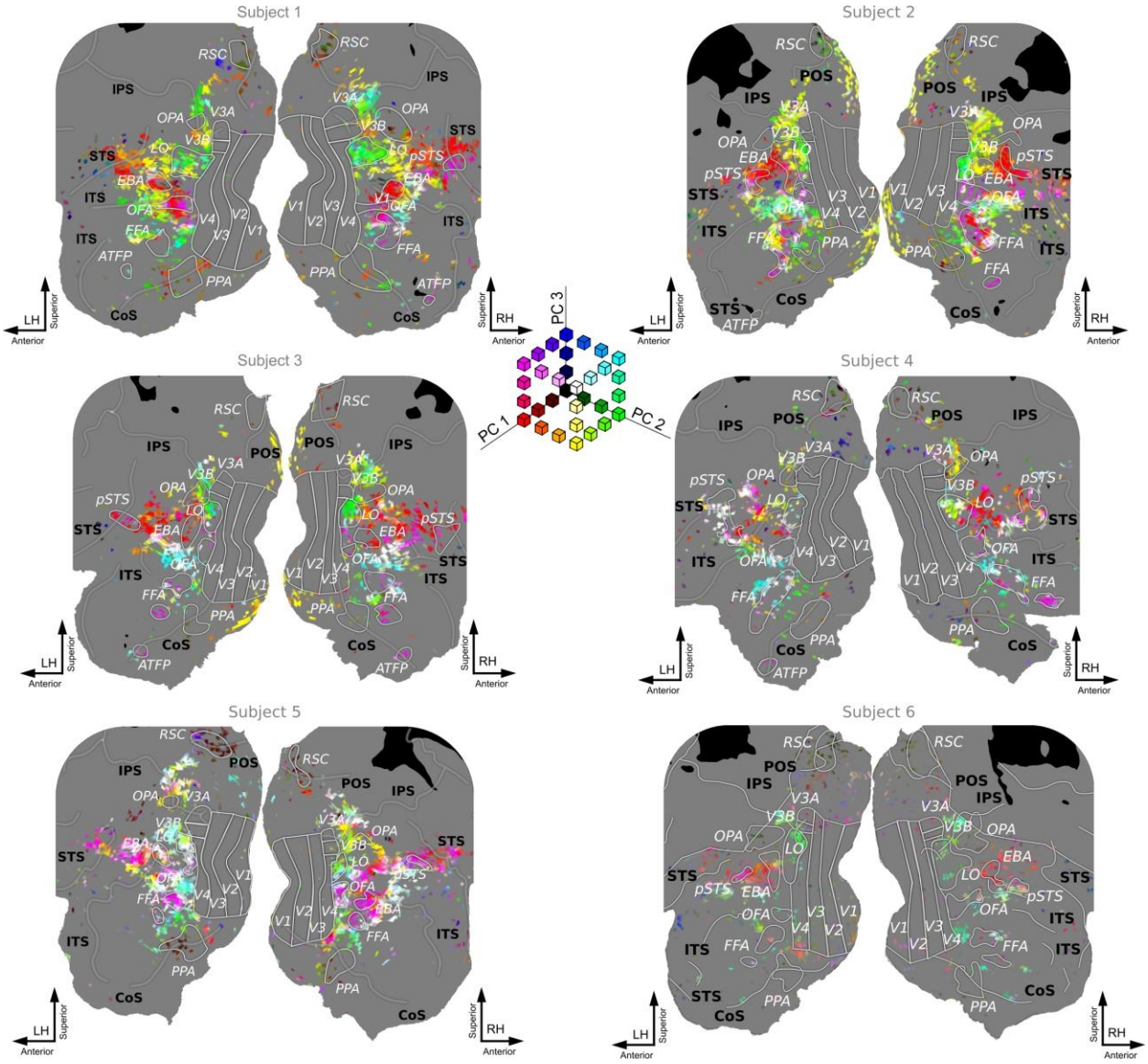


**Figure 3-52. Top 3 Group PC scores within non-EVC OTC.**

Here, we map PC scores onto cortex using the top three PCs from PCA on CSVA model weights across all OTC voxels not contained within EVC, where model fit was significant and better than that of the Semantic Only model. As illustrated in Fig. 3-50, when voxel selection is restricted in this manner, the top three PCs are highly correlated with those from the OTC analysis reported in chapter 2(rs>.99). As in Fig. 3-51, a RGB color space is used to map PC scores onto cortex (red = scores on PC1, green = scores on PC2, blue = scores on PC3). PC scores are thresholded at 6 standard deviations above and below 0.

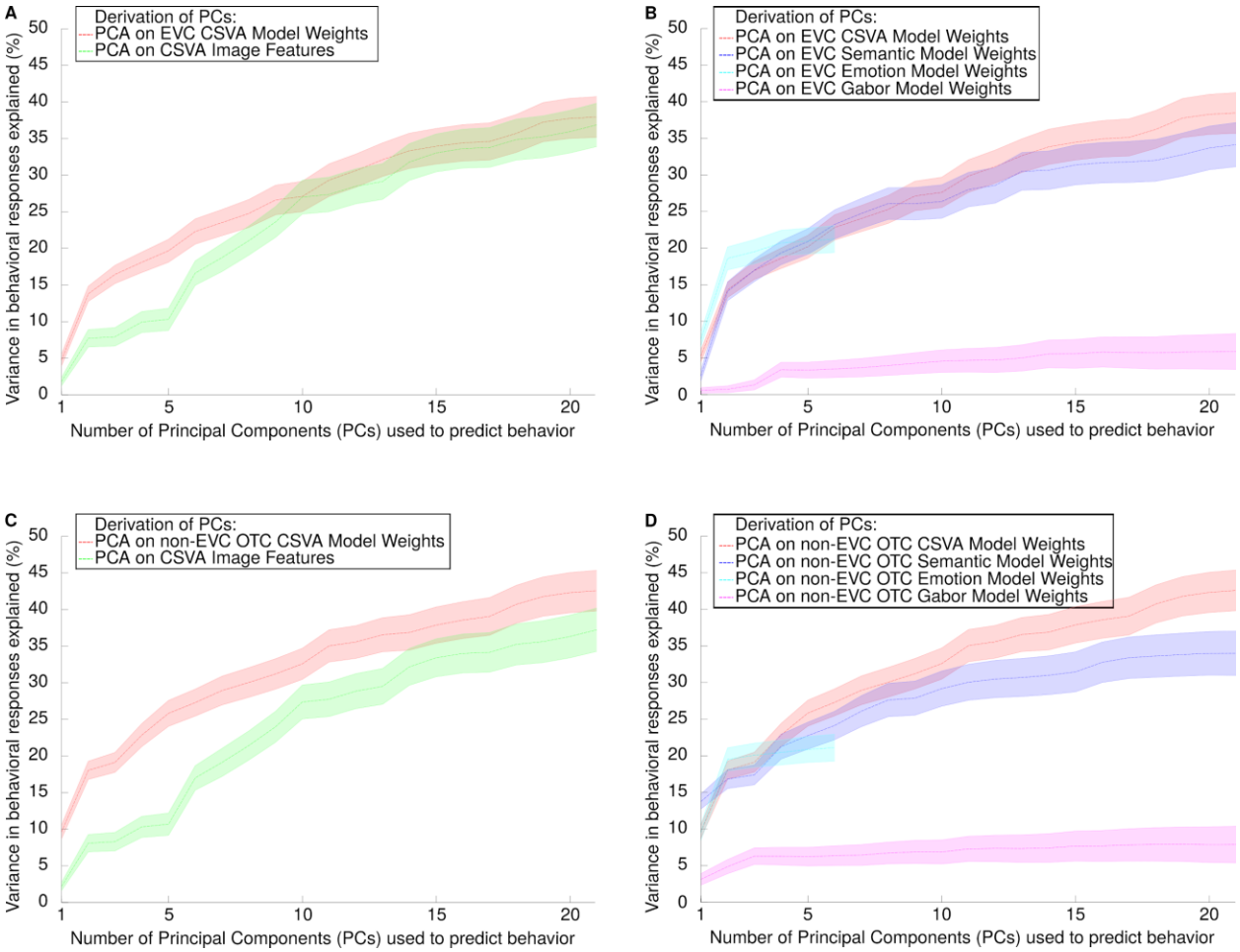


**Figure 3-53. Explained variance in behavioral responses by PCs within OTC.**

Subjects recruited through Amazon's Mechanical Turk platform were shown the emotional natural images used in this study and asked to pick one or more behavioral responses appropriate to the content of each image. (A) We examined the extent to which EVC tuning to emotional natural images, as captured by CSVA group-level PC loadings (red line), predicted behavioral responses selected, across images. The percentage of out-of-sample variance in behavioral responses explained (y axis) is plotted against the number of PCs included as predictors in an ordinary least squares regression analysis. The shaded area around the solid line represents the 95% confidence interval. We also calculated the variance in behavioral responses explained using PCs derived directly from PCA on CSVA image features, across images (green line). EVC tuning to CSVA features predicted behavioral responses significantly better than components from PCA conducted directly on the features themselves for only the first 8 PCs. (B). Here, we plot the variance in behavioral responses explained by PCs obtained from PCA on EVC feature weights for the CSVA model (red) versus (i) a low-level structural (Gabor) model (purple); (ii) the Semantic Only model (dark blue) and (iii) the Emotion Only model (light blue). Given the smaller feature space of the Emotion Only model, the maximum number of PCs that can be extracted for this model is 6. The poor performance of Gabor model PCs in predicting behavioral responses suggests that although EVC is well predicted by a Gabor wavelets model, tuning there to low-level image structural features is insufficient to guide behavior. Both the Semantic Only and Emotion Only models outperform the Gabor model in predicting behavior. The Emotion Only model performs particularly well at low levels of dimensionality, and it's maximal prediction of behavior (at n=6 PCs) is not significantly less than that achieved by the CSVA model using an equivalent number of components. The Semantic Only model never predicts significantly more variance than the CSVA model, and does significantly

worse than the CSVA model at its maximal prediction of behavior (n=21 PCs). (C-D) We likewise examined the extent to which tuning within non-EVC regions of OTC could predict variance in behavioral responses. (C) Variance explained by PCA on CSVA model features (shown in red) was significantly higher than that explained by PCA on stimulus features (shown in green) across all levels of dimensionality considered. This is consistent with OTC, and particularly non-EVC regions of OTC, showing selective representation of image semantic and emotional features pertinent to behavior. (D) Tuning in OTC as captured by the Gabor model was similarly poor to that captured by tuning in EVC, and significantly less than that captured by the CSVA model in all 21 PCs considered. The Semantic Only and Emotion Only models both performed particularly well at low levels of dimensionality, but both models explained significantly less variance than the CSVA models at their maximal prediction of behavior (at n=21 and n=6 PCs respectively). This finding is nearly identical to the that found when PCA was done on the entirety of OTC, as shown in chapter 2.

# 7. Principal Components Analysis (PCA) on CSVA Model Voxel Feature Weights and Behavioral Response Analysis within Frontal Cortex Subregions

## Introduction

Many regions of the prefrontal cortex are known to represent both semantics and emotion (Ongur & Price, 2000). Many regions of OTC also have direct connections with prefrontal regions (Rempel-Clower & Barbas, 2000). Thus, we were interested to know whether CSVA model tuning within prefrontal cortex could better predict appropriate behavioral responses to emotional stimuli than could OTC tuning, suggesting that prefrontal cortex provides additional processing of semantic and emotional information pertinent to appropriate behavioral response. Orbitofrontal Cortex (OFC) in particular is known to represent affective properties of visual stimuli (Shenhav, Barret, & Bar, 2013; Kawasaki, et al., 2001). We thus conducted two additional PCAs on voxel feature weights within OFC significantly predicted by the CSVA model, and all other frontal regions excluding OFC and motor cortex (known as non-OFC Frontal Cortex from here on, see ROI Definitions section above for details).

## Methods

### PCA

This analysis was done on the 4 ROIs listed above (EVC, non-OTC EVC, OFC, & non-OFC Frontal) exactly as described in two supplemental methods sub-sections of chapter 2 within the "fMRI Data Modeling" section of the METHOD DETAILS. They were titled:

1. "PCA of CSVA model feature weights"
2. "Interpreting the top 3 group-level PCs"

Statistical tests were done as described in the two subsections of the QUANTIFICATION AND STATISTICAL ANALYSIS section titled:

1. "Principal component significance testing"
2. "Relating PCs to hypothetical dimensions"

### Behavioral Analysis

This analysis was done on the 4 ROIs listed above (EVC, non-OTC EVC, OFC, & non-OFC Frontal) exactly as described in a supplemental methods subsection of chapter 2 within the "fMRI Data Modeling" section of the METHOD DETAILS. It was titled:

1. "Prediction of behavioral responses"

Statistical tests were done as described in the sub-section of the QUANTIFICATION AND STATISTICAL ANALYSIS section titled:

1. "Predicting behavioral responses from OTC tuning"

## Results

Results from group PCA on both OFC and non-OFC Frontal Cortex show very similar results. In both analyses, only the top group PC explains significantly more variance than the stimulus feature PCs (ps<.05) within their respective sets of voxel feature weights (panels A&D of Fig. 3-54). While this is true, if stimulus feature PCs were re-ordered by a decreasing amount of variance they explain, as are the OFC and non-OFC Frontal CSVA PCs, then the top 3 group PCs from both ROIs would explain more variance than the stimulus feature PCs. For this reason, and for comparison purposes with the OTC PCs, we will look at the top 3 PCs from OFC and non-OFC Frontal Cortex. The amount of variance explained by the top PC differs in the two regions however, with the top OFC PC explaining approximately 17% of the variance in OFC voxel feature weights, and the top non-OFC Frontal PC explaining approximately 23% of the variance in non-OFC Frontal voxel feature weights. Correlations between the top 2 group and single subject CSVA PCs from both regions show similarity (rs>.61), and are significantly correlated (ps<.05), although less so than OTC CSVA PCs (as seen in panels B & E of Figure 3-54. The correlations between the third group and single-subject CSVA PCs vary dramatically from subject to subject. Investigating the similarity in PC loadings from these two frontal regions with those from entire OTC (panels C & F of Figure 3-54) shows that although the top 2 PCs from both OFC (PC1 r=.69, PC2 r=.67) and non-OFC Frontal Cortex (PC1 r=.74, PC2 r=.72) are significantly similar to the top 2 PCs from OTC (ps < .05), PCs 1 & 2 are also highly confusable in their similarity with OTC PCs 1 & 2 (.52 <= rs <= .56). PC 3 from both frontal regions is moderately similar to all 3 PCs from OTC. Projecting voxel scores on the top 3 PCs from both OFC and non-OFC Frontal Cortex onto flatmaps (Figs. 3-55 & 3-56) revealed that the spatial structure of CSVA model representations within these frontal regions does not show large degrees of similarity across subjects. Within OFC this may be due to the relatively few number of voxels significantly predicted by the CSVA model itself.

Subsequent to doing PCA, the same analysis of appropriate behavioral responses done on OTC PCs was conducted on the PCs from both OFC and non-OFC Frontal Cortex (Fig. 3-57). OFC and non-OFC Frontal tuning to CSVA features predicted behavioral responses significantly better than components from PCA conducted directly on the features themselves for all dimensions considered, as was the case for OTC tuning (panels A&C, Fig 3-57). Crucially, the amount of variance explained in behavioral responses by either OFC or non-OFC Frontal tuning across all 21 PCs considered was not significantly greater than that explained by OTC tuning ps>.1). Additionally, both the Semantic Only and Emotion Only models outperform the Gabor model in predicting behavior in both frontal regions (panels B&D Fig. 3-57), and while the Emotion Only model performs better than the CSVA model at low dimensionality, both the Emotion Only and Semantic Only Models predict significantly less variance (ps <.05) at their maximal levels of prediction (at n=6 and n=21 PCs respectively). Thus, results from these PCAs indicate that while the CSVA model does significantly predict BOLD activity within frontal regions, there is not a high degree of shared representational space within these regions as seen

in the low correlations between group and single subject PCs beyond PC 2, this shared representational space explains a relatively small amount of variance as only one PC predicts significantly more variance than PCs of the features themselves in both OFC and non-OFC Frotnal ROIs, and the spatial structure of this representational space across these frontal regions lacks similarity, unlike within OTC (specifically non-EVC OTC). Furthermore, the amount of variance that OFC and non-OFC Frontal Cortex CSVA tuning can explain in appropriate behavioral response to emotional stimuli is no greater than that explained by OTC CSVA tuning. This suggests that the processing of the combination of semantic and emotion information done by OTC is just as informative in the selection of appropriate behavioral responses to situations with emotional content as is that of frontal regions.

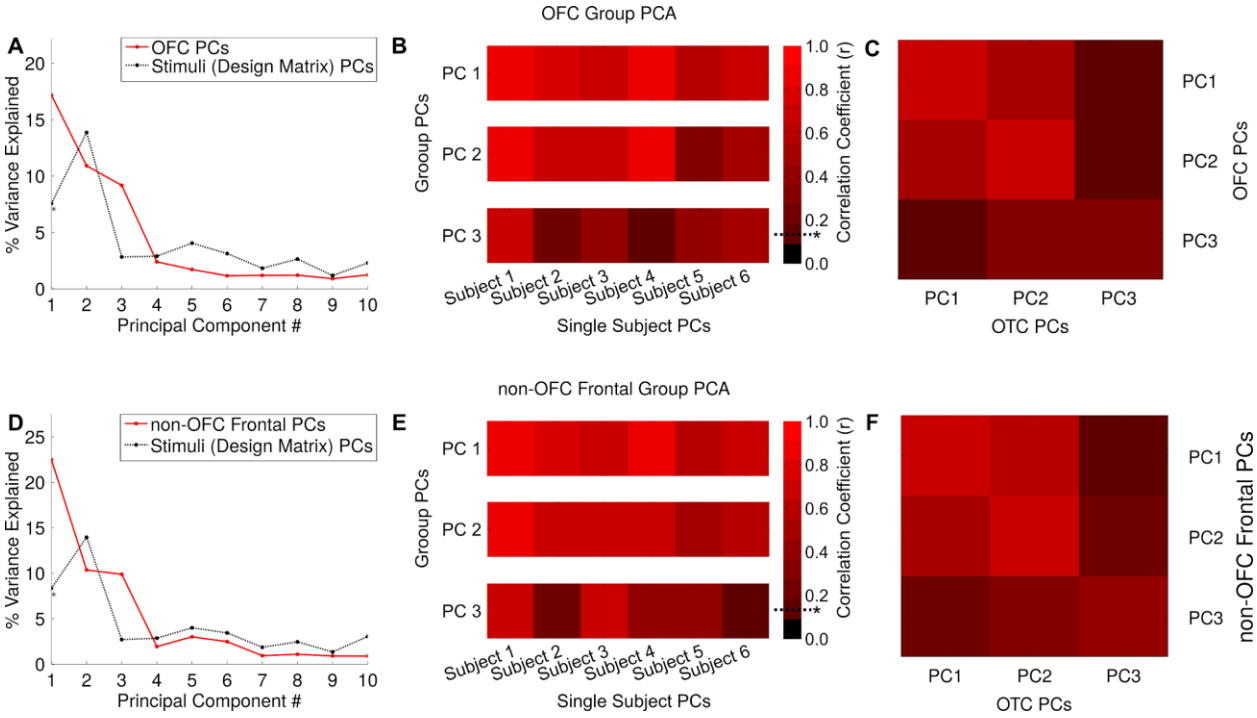


**Figure 3-54. PCA on voxels within Frontal Cortex: Variance explained and similarity of PCs.**

(A-C) A group-level principal components analysis (PCA) was conducted on the CSVA model feature weights across all Orbitofrontal Cortex (OFC) voxels where model fit was significant (all panels here use the same format as described in Fig. 3-50). (A) Only the top OFC PC explains significantly more variance than PCA conducted on the stimulus features themselves. The 2nd stimulus PC explains more variance in the OFC voxels' feature weights than the 1st stimulus PC, and as such re-ordering the top 3 stimulus PCs would result in the top 3 OFC PCs explaining more variance than the stimulus PCs. Thus the top 3 PCs will be shown in subsequent analysis for comparison purposes. (B) Correlations between the top 2 OFC group and single subject PC loadings are significant, however several correlations between the 3rd group and single subject PC fall below significance. (C) Correlations between the loadings of the top 2 PCs are high for PCA done within OFC and OTC, however there is a relatively high level of confusability between the two PCs, as the correlations between PCs 1 & 2 from both PCAs are also high. Lack of similarity between the third PCs indicates they represent different aspects of semantic and emotional information. (D-F) A second group-level principal components analysis (PCA) was conducted on the CSVA model feature weights across all Frontal voxels not contained within OFC, where model fit was significant. (D) The pattern of variance explained by the top 3 PCs of this model is similar to that of the OFC PCA, with the exception that the top non-OFC Frontal PC explains approximately 5% more variance than the top OFC PC. (E) As with the OFC PCs, the top 2 group non-OFC Frontal PCs are similar to the single subject PCs, but not the 3rd. (F) Similarity between the top 2 non-OFC Frontal PCs was also high with OTC PCs, as was the confusability between the 1st and 2nd PCs. The 3rd PCs from non-OFC Frontal and OFC were also not similar to each other, nor to the top 2 PCs.

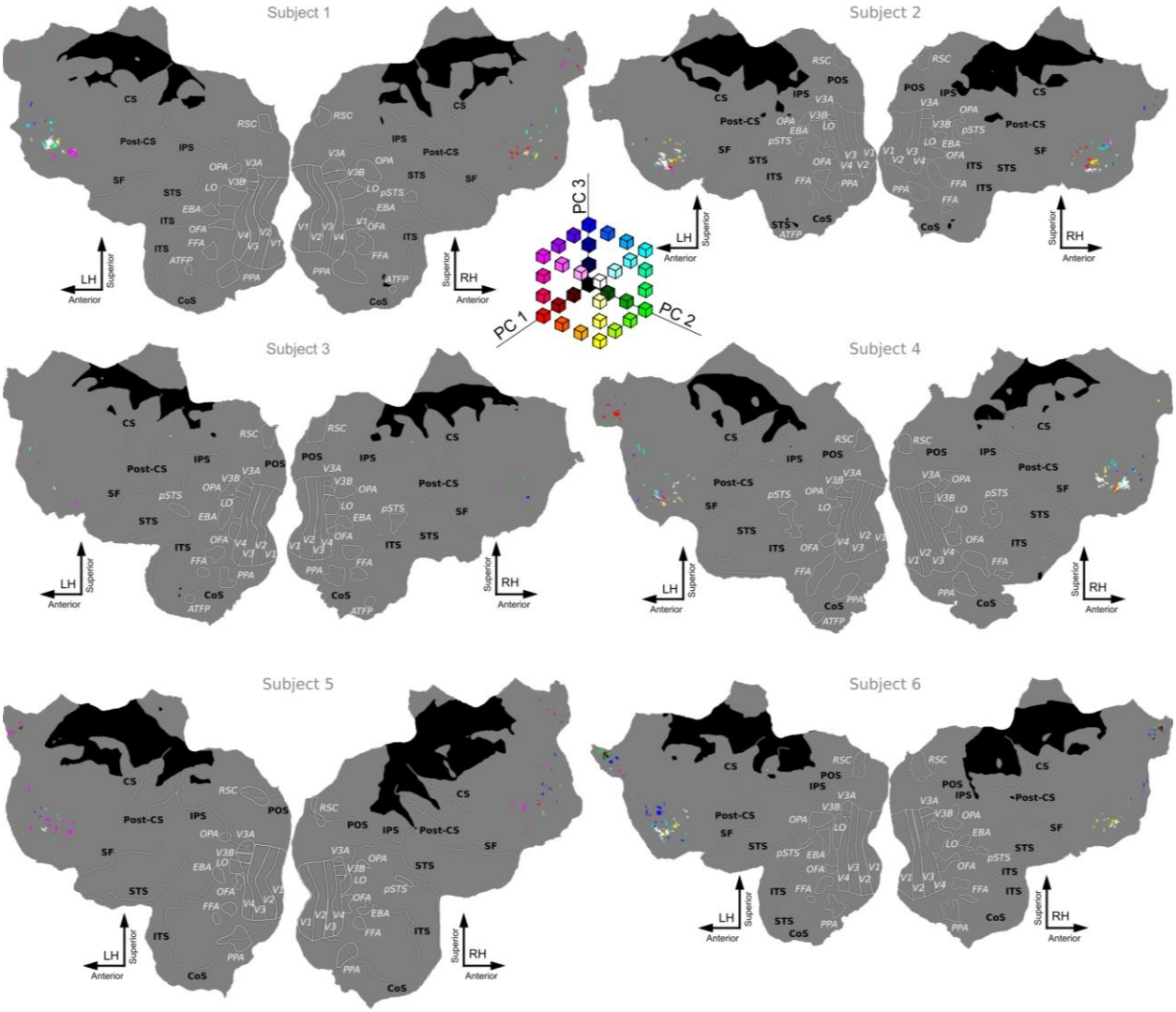


**Figure 3-55. Top 3 Group PC scores within Orbitofrontal Cortex (OFC).**

PC scores from the top three PCs from PCA within OFC is mapped onto cortex. Some similarity in the structure of representation within OFC is likely, however the number of voxels significantly predicted by the CSVA model in OFC is minimal. As in Fig. 3-51, a RGB color space is used to map PC scores onto cortex (red = scores on PC1, green = scores on PC2, blue = scores on PC3), and PC scores are thresholded at 6 standard deviations above and below 0.

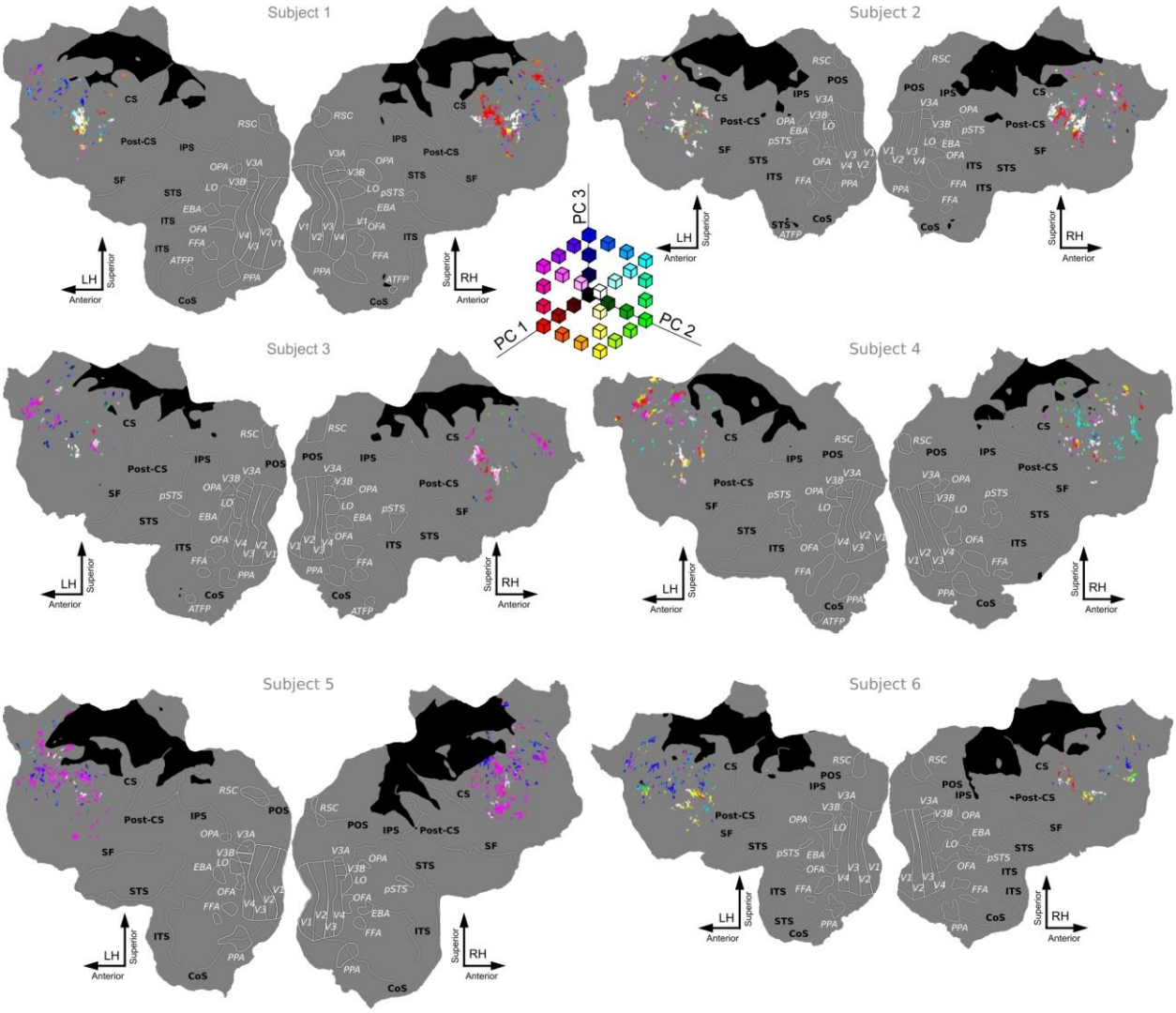


**Figure 3-56. Top 3 Group PC scores within non-OFC Frontal Cortex.**

PC scores from the top three PCs from PCA within Frontal Cortex excluding voxels within OFC is mapped onto cortex. Little similarity in the structure of representation across non-OFC Frontal Cortex even though large swaths are significantly predicted by the CSVA model. As in Fig. 3-51, a RGB color space is used to map PC scores onto cortex (red = scores on PC1, green = scores on PC2, blue = scores on PC3), and PC scores are thresholded at 6 standard deviations above and below 0.

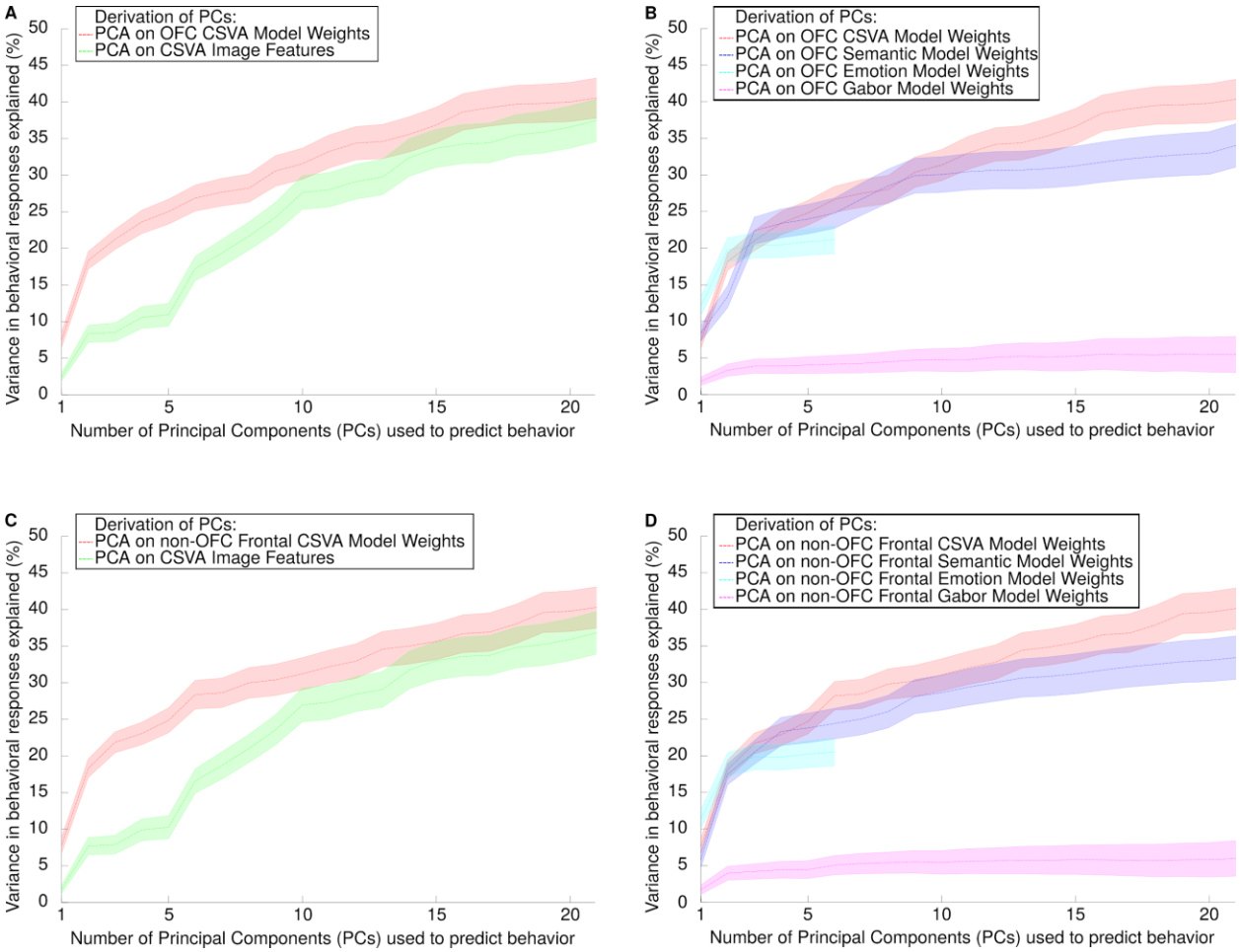


**Figure 3-57. Explained variance in behavioral responses by PCs within Frontal Cortex.**
(A) OFC tuning to CSVA features predicted behavioral responses significantly better than components from PCA conducted directly on the features themselves for all dimensions considered (all panels here use the same format as described in Fig. 3-53). (B). Both the Semantic Only and Emotion Only models outperform the Gabor model in predicting behavior. While the Emotion Only model performs better than the CSVA model at low dimensionality, both the Emotion Only and Semantic Only Models predict significantly less variance at their maximal levels of prediction (at n=6 and n=21 PCs respectively). (C-D) Tuning within non-OFC regions of Frontal Cortex were likewise examined. (C) Tuning in non-OFC Frontal regions also significantly better predicted behavioral responses than did PCs of stimulus features themselves across all 21 dimensions. (D) As seen in OFC, CSVA PCs from non-OFC Frontal Cortex predicted significantly more variance than the Emotion Only and Semantic Only models at their respective maximal dimensionality, and significantly more than the Gabors model in all dimensions.

# 8. Assessing Stimulus Correlations between Gabor Model Features and CSVA Model Features

## Introduction

The principal components analysis (PCA) on "Combined Semantic Valence & Arousal" (CSVA) feature weights reported in chapter 2 aimed to determine primary dimensions of representation related to semantic and emotional content within occipital-temporal cortex (OTC). It is also known that a Gabor wavelet model (see supplemental materials from chapter 2 for details) can significantly predict BOLD activity within some of the same regions of OTC as the CSVA model. This raises the possibility that representation within OTC attributed to the primary dimensions of the CSVA feature weights should actually be attributed to shared variance, or correlations, between the CSVA and Gabor wavelet models. To determine whether this was the case in these data, we removed any variance explained by the Gabor wavelet model from the BOLD data, refit the CSVA model to those residuals, and then re-calculated the PC scores from the top 3 PCs of the original CSVA model. We then compared the resulting PC score pattern across OTC of these with the original PC score pattern to determine the extent that Gabor wavelet feature correlations with CSVA features affected OTC tuning.

## Methods

In order to assess the impact of possible stimulus correlations, we removed variance in the BOLD signal explained by the Gabor wavelet model, refit the CSVA model to those residuals, and re-calculated CSVA PC scores on the resultant feature weights. Specifically, the following procedure was used independently on every cortical voxel. 10-fold cross-validation on the 30 estimation runs was used to estimate out-of-sample BOLD data predictions for the Gabor wavelet model. During each fold, the Gabor wavelet model was fit to 27 of the 30 estimation runs, and the resulting feature weights were then used to predict the BOLD data of the remaining 3 runs. This was repeated for all 10 non-overlapping groups of 3 estimation runs and the results were concatenated, resulting in an out-of-sample predicted BOLD time-series for all 30 estimation runs. This predicted BOLD time-series was then used as a "nuisance regressor", and was regressed out of the original BOLD data time-series. The CSVA model was then re-fit to this new BOLD time-series, resulting in a new set of voxel-wise feature weights with Gabor wavelet model stimulus correlations controlled for. Finally, these new feature weights were multiplied by the top 3 original CSVA PC loadings, resulting in Gabor-controlled voxel-wise PC scores. These were plotted on flatmaps. In addition, to assess the degree to which these Gabor-controlled PC scores differed from the original CSVA PC scores, the Gabor-controlled PC scores were subtracted from the original CSVA PC scores for every voxel, and the resultant differences were plotted on a flatmap.

## Results

Figure 3-58 shows the results of this analysis. Column A of Figure 3-58 shows the difference between the original PC scores and PC scores created by multiplying voxel feature weights from a CSVA model with Gabor wavelet model variance removed, and the original CSVA PC loadings. Thus, these maps show the effects of stimulus correlations between the Gabor wavelet and CSVA models on the CSVA model PC scores. Column B shows maps similar to those in column, except that the out-of-sample CSVA model explained variance has been removed before refitting the CSVA model. This was done as a sanity check of the method, and we expected these maps to be nearly identical to the original CSVA PS score maps shown in column D, which they are. Column C shows the difference between columns B and A, showing voxel-wise CSVA PC scores that cannot be explained by the Gabor wavelet model. Comparing the maps in column C with those in D we see a high degree of similarity across all subjects, although they are not identical. Across all subjects, PC scores within EVC shifted to blue and magenta hues, indicating high scores on PC 3 and low on PCs 1 and 2. Additionally, voxels in the region between LO and pSTS that were largely red (high on PC 1 and low on PCs 2 and 3) have become largely magenta (high on PC 1 and 3, low on PC 2), indicating a higher loading on PC 3. Finally, many voxels between FFA and pSTS that had been colored yellow (high on PCs 1 and 2, low on 3) are now colored white, again indicating an increase in scores on PC 3. Thus, it seems that removing variance explained by the Gabor wavelet model caused PC 3 scores to increase in several regions of OTC, while largely keeping the spatial structure in tact. This indicates that the effect of correlations between Gabor and CSVA features on OTC tuning is minimal.

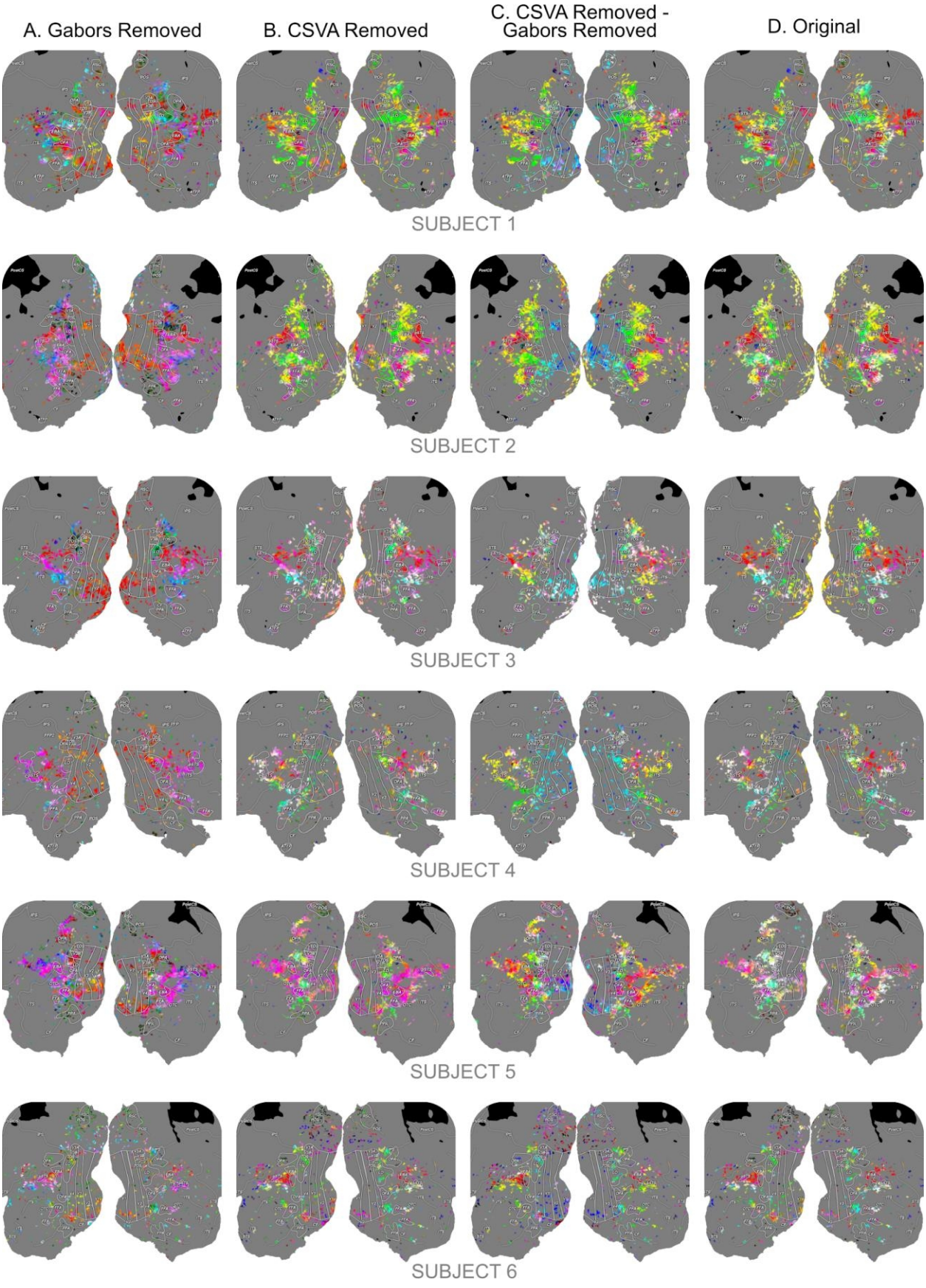
A. Gabors Removed
B. CSVA Removed
C. CSVA Removed - Gabors Removed
D. Original
SUBJECT 1
SUBJECT 2
SUBJECT 3
SUBJECT 4
SUBJECT 5
SUBJECT 6

**Figure 3-58. Correlations between Gabor and CSVA model features do not drive the variance captured by the CSVA model PCs in OTC.**

(A) Flatmaps depicting the impact of stimulus feature correlations between the Gabor wavelet model and the CSVA model on CSVA model PC scores (PC scores scaled by standard deviation and colored same as in Fig. 2-5 for all flatmaps here). Out-of-sample variance explained by the Gabor wavelet model was regressed out of the estimation BOLD data, and the residuals were used to re-fit the CSVA model resulting in new feature weights. Voxel-wise PC scores accounting for Gabor and CSVA feature correlations were then calculated by multiplying these new feature weights with the original top 3 CSVA model PC loadings. Finally, these new PC scores were subtracted from the original CSVA PC scores, and that difference visualized on these flatmaps. These differences can be interpreted as showing the effects of stimulus correlations between the Gabor wavelet and CSVA models on the CSVA model PC scores. (B) 10-fold cross-validation was used to estimate the out-of-sample variance explained by the Gabor wavelet model, and thus only 90% of the data was used to estimate the feature weights for each fold. To ensure that this reduced amount of estimation data did not hinder our findings, we repeated the procedure using the CSVA model itself. The flatmaps in panel B show the subtraction between the new PC scores and the original CSVA model PC scores. The pattern of scores across OTC in these flatmaps remains similar to the pattern of the original PC scores (panel D), although the magnitude of the scores is less (as can be seen by the less vibrant colors, likely due to random noise in the estimated feature weights being subtracted out. This confirms that our cross-validation procedure for model estimation is effective. (C) Flatmaps depicting scores from panel B minus those shown in A. These maps show voxel-wise CSVA PC scores that cannot be explained by the Gabor wavelet model. The similarity between these maps and the original CSVA PC score maps indicates that voxel-wise CSVA model PC scores cannot be explained by tuning to Gabor wavelet model features. (D) Original CSVA model PC scores for comparison.

# 9. Controlling for Physiological Noise

## Introduction

Self-reported emotional arousal, as quantified in many of our models, is known to correlate with physiological arousal, as quantified by respiration rate and pulse-oximetry measures. Thus it is desirable to control for any affect that physiological arousal may have had on the BOLD signal. Respiration rate and pulse oximetry data were collected during fMRI scanning sessions for all subjects. These data were then preprocessed using the PhLEM Matlab Toolbox (Verstynen & Deshpande, 2011) to create nuisance regressors whose explained variance was removed from the BOLD signal prior to re-estimation of the CSVA model feature weights and prediction of the validation BOLD data.

## Methods

To control for effects of physiological noise on the BOLD signal, respiration rate and pulse-oximetry data were collected on all subjects during fMRI data collection, as noted in chapter 2. These time series data were then preprocessed using the Physiological Log Extraction for Modeling (PhLEM) Matlab Toolbox (Verstynen & Deshpande, 2011), along with custom scripts. This toolbox implements the RETROICOR method (Glover, Li & Ress, 2000) for creating nuisance regressors. As Glover, Li & Ress note: "It is assumed that the physiological processes are quasi-periodic so that cardiac and respiratory phases can be uniquely assigned for each image in the time series. Accordingly, the physiological noise component y(t) can be expressed as a low order Fourier series expanded in terms of these phases". The toolbox generates a sine/cosine phase time-series for the respiration signal, and two cosine/sine phase time-series for the pulse-oximetry signal, resulting in 6 time-series whose explained variance in the BOLD data is removed. This is done by fitting a general linear model (GLM) using these 6 physiological nuisance regressors to the BOLD data across the estimation data and validation data separately. The residuals from this model were then used to estimate the CSVA feature weights exactly as described in chapter 2. The feature weights from this model with physiological noise controlled for were then subjected to a PCA as done on the CSVA model in chapter 2. The same voxels used on the PCA of the original CSVA model were used in this new PCA. Pearson correlation coefficients were then estimated between the top 3 PCs from the original CSVA model, and the new CSVA model controlling for physiological noise.

## Results

Comparing the prediction accuracy (correlation between predicted and recorded validation BOLD data) of the original CSVA model with this "physio-removed" CSVA model, we can see that prediction accuracies did not change in any meaningful way (Fig. 3-59). Additionally, PCA was done on the voxel feature weights from the "physio-removed" CSVA

model. Correlations between the top 3 “physio-removed’ CSVA PCs and the original CSVA PCs (Table 3-4) was extremely high (rs>0.99), indicating very little effect of physiological noise on the dimensionality of tuning within OTC. Taken together, these results show that physiological noise, possibly attributable to emotional arousal, was not underlying the results of our VWM analysis of the CSVA model from chapter 2.

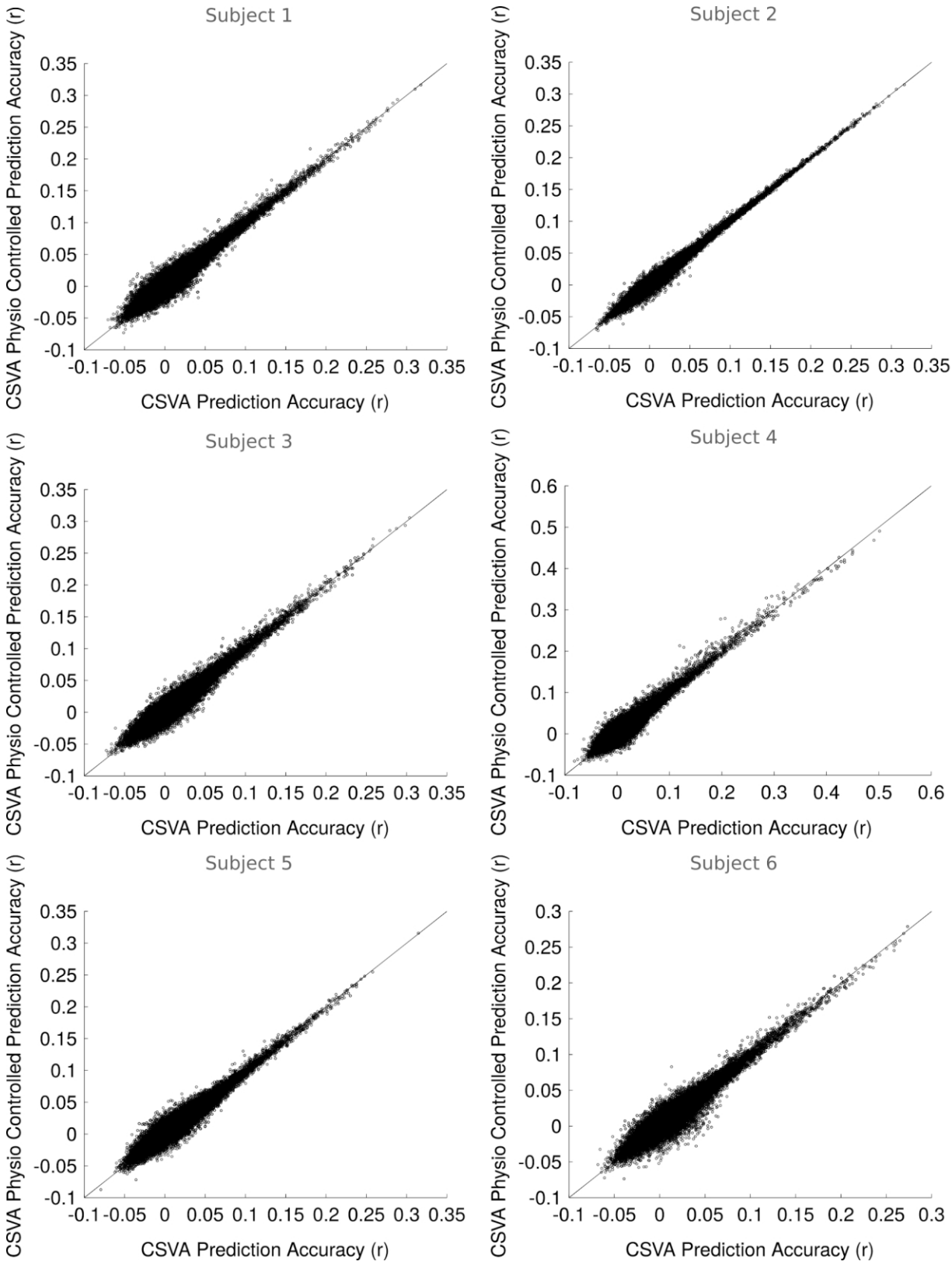


**Figure 3-59. Similarity in Prediction Accuracy when Controlling for Physiological Noise.**

The PhLEM Matlab Toolbox was used to calculate nuisance regressors which capture low-frequency phase information contained in respiration and pulse-oximetry signals. Variance explained in the BOLD data by these nuisance regressors was estimated using linear regression, and the residuals were then used to re-fit the CSVA model, resulting in feature weights that have been estimated while controlling for the effects of physiological noise ("CSVA Physio Controlled"). Scatter plots comparing the voxel-wise prediction accuracy (correlation) values from the Original CSVA model and the Physio Controlled CSVA model are shown here. Points on the x=y line show equal performance for both models; points above the line indicate voxels where the model on the y-axis (CSVA Physio Controlled) performed better, and points below the line indicate voxels where the model on the x-axis (Original CSVA) performed better. Controlling for contributions of physiological noise to the BOLD signal does not change CSVA model fit across the 6 subjects.

| | **CSVA Physio Controlled** | | | |
|---|---|---|---|---|
| **Original CSVA** | | PC1 | PC2 | PC3 |
| | PC1 | 0.9997 | 0.0021 | 0.0034 |
| | PC2 | 0.0172 | 0.9997 | -0.0010 |
| | PC3 | 0.0024 | -0.0000 | 0.9988 |

**Table 3-4. Similarity of Top 3 Principal Component Loadings when Controlling for Physiological Noise.**

Pearson correlation coefficients between the top 3 Principal Components (PCs) from the Original CSVA model and the Physio Controlled CSVA model. Controlling for the effect of physiological noise on the BOLD data had very little influence on the top 3 CSVA model PC loadings, as the high correlations (>.99) between corresponding PCs show.

## From Cortical to Behavioral Responses of Naturalistic Emotional Images

The literature reviews and supplemental analyses described in this chapter provide additional evidence, along with the study described in chapter 2, that emotional valence and arousal of animate stimuli are represented within OTC. As discussed previously, neuroscientists debate the location and nature of the neural mechanisms that subserve the extraction of emotional content that confers emotional stimuli with prioritized processing. While these findings are consistent with the "two-stage" or "multi-wave" hypotheses (see section 1 of this chapter for details) which postulate that a fast, feed-forward sweep of processing within OTC is responsible for quick extraction of emotional information, the relatively low temporal resolution of fMRI prevents any strong conclusions regarding temporal processes from being inferred from these studies. While fMRI temporal resolution is too slow to draw meaningful temporal inferences, a rich history of psychophysical experiments has provided temporal constraints on the ways in which the visual system processes information. In particular, clever studies utilizing

ultra-rapid presentation of naturalistic images and EEG have inferred a 150ms upper bound on the processing needed to extract the semantic category of visual stimuli (Thorpe, Fixe, & Marlot, 1996). The addition of backwards masking to these paradigms have resulted in studies which found a 12ms lower bound on the length of processing time that occurs within a given stage of the visual feed-forward hierarchy (Bacon-Mace et al., 2005). Several other studies have found that humans can extract a wide range of information from a very brief presentation of visual stimuli, from object category (Fei-Fei et al., 2007), to emotional valence (Nummenmaa, Hyona, & Calvo, 2010).

Building on these findings, in Chapter 4 we present a study using ultra-rapid presentation of naturalistic emotional images and multiple forced-choice (mAFC) categorization of valence and semantic category to answer several questions related to human perception of visual images. Proponents of the "two-route" hypothesis, which postulates that a subcortical neural circuit (e.g. the "low-road" or "fast-route") subserves the fast, automatic detection of threatening stimuli point to the prioritized processing of emotional stimuli as evidence that a dedicated subcortical circuit must exist due to the fast temporal nature of the observed behavioral outcomes in response to emotional stimuli. This prioritized processing is thought to come from evolved mechanisms which detect, and respond, to negative, threatening, animate stimuli. Thus the first questions chapter 4 addresses is whether we find evidence of this prioritized processing for negative, animate stimuli. Lack of evidence for prioritized processing for negative, animate stimuli would be inconsistent with the "two-route" hypothesis. Additionally, many proponents of the "two-route" hypothesis also argue that affective information is extracted from images before semantic information is extracted, a hypothesis called affective primacy. They argue that the subcortical fast route subserves this fast extraction of emotional information in parallel to, and before, semantic information is extracted within OTC. In contrast to the affective primacy hypothesis, the cognitive primacy hypothesis postulates that semantic information must be extracted before emotional information can be extracted, which is consistent with the "two-stage" or "multi-wave" hypotheses. Thus, in chapter 4 we describe data that in which competing patterns of conditional categorization performance support either the cognitive or affective primacy hypothesis. These experiments shed light on the related questions around the nature of prioritized processing of emotional stimuli and whether the affective or cognitive primacy hypothesis is correct, both of which are critical to the arguments made for the existence of a subcortical fast route for processing of emotional stimuli.

# Chapter 4. Ultra-Rapid Affective and Semantic Categorization of Naturalistic Images: Support for the Cognitive Primacy Hypothesis

## Introduction

### Ultra-Rapid Presentation of Visual Stimuli

A brief glance can provide information about the surrounding environment which allows for adaptive behavior that might mean the difference between life and death. Quickly extracting the relevant semantic and emotional information from an information dense sensory world is one of the human brain's most adaptive computational feats. Early studies using rapid serial visual presentation (RSVP) of novel images demonstrated that humans can accurately identify objects presented for as little as 50ms (Intraub, 1981; Potter, 1975; Potter & Levy, 1969), although the novel images are soon forgotten. More recent studies using RSVP have shown evidence for accurate object identification with presentation times as short as 15ms (Potter, Staub, Rado & O'Connor (2002); Keysers & Perrett, 2002). Additionally, accurate classification of numerous image properties under ultra-rapid presentation (<50ms) of single images is well within achievable human performance. Observers can determine whether a scene is natural or urban (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007), detect the presence of large objects such as vehicles (Van Rullen & Thorpe, 2001) or animals (Thorpe, Fize, & Marlot, 1996), determine the basic-level or superordinate-level categories of a scene (Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005), and even determine the pleasantness of a scene (Kaplan, 1992). While visual iconic memory could have helped mediate performance in these tasks, a number of more recent studies have also found accurate classification on similar tasks using backwards masking to control for any effects of iconic memory (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei et al., 2007; Greene & Oliva, 2009a, 2009b; Grill-Spector & Kanwisher, 2005; Maljkovic & Martini, 2005; Nummenmaa, Hyönä, & Calvo, 2010). While many of these studies motivate their experiments by invoking the survival advantage of quick scene processing, surprisingly few have used the paradigms here mentioned to study naturalistic emotional scenes which contain scenarios where adaptive responses are actually necessary. So what is known concerning visual processing of emotional scenes?

### Biological Preparedness

Arguing from an evolutionary perspective, theorists have proposed that some stimuli are "biologically prepared" by our evolutionary past to afford prioritized processing (Ohman, Erixon, & Lofberg, 1975; Ohman, Flykt, & Esteves, 2001; Seligman, 1971). Biologically prepared stimuli were originally conceived of as threatening, animate stimuli such as spiders, snakes and angry faces for which quick responses are highly adaptive. Since then other theorists have expanded this conceptualization to include other categories of emotional stimuli (Anderson, 1990; Murphy & Zajonc, 1993). Numerous studies have found evidence for this prioritized processing of emotional stimuli. For example, subjects can identify fear-relevant stimuli faster than fear-irrelevant stimuli in a visual search task (Ohman, Flykt, & Esteves, 2001), and in

continuous flash suppression studies which render stimuli non-conscious, faces showing fearful expressions break into consciousness faster than do faces showing happy expressions (Yang, Zald, & Blake, 2007). Recently, several studies using ultra-rapid scene presentations of naturalistic emotional stimuli have found evidence that observers can accurately categorize pleasant vs. unpleasant images presented to extrafoveal vision (Calvo & Nummenmaa, 2007), and that accurate semantic categorization (animal vs. human) is faster for emotional than non-emotional images, also presented extrafoveally (Calvo & Lang, 2005; Calvo, Nummenmaa, & Hyöna, 2008; Nummenmaa, Hyöna, & Calvo, 2010). Together these findings seem to support the theory that emotional stimuli benefit from prioritized visual processing. Other studies investigating the effects of emotional stimuli on selective attention have found some contradictory findings, however.

## Attentional Narrowing

In a now seminal work, Easterbrook (1959) hypothesized that an emotional stimulus serves to narrow the focus of attention onto that stimulus in order to facilitate appropriate behavioral responses. As van Steenbergen and colleagues argue (2011), Easterbrook's original hypothesis only related to unpleasant (i.e. negative) situations, although some authors suggest any increase of arousal serves to narrow the focus of attention (Anderson, 1990). Using a visual anti-saccade task which compares reaction times of pro- vs. anti-saccades, van Steenbergen showed that attentional narrowing was indeed only present for negative, and not positive, stimuli. Crucially though, their findings were in the opposite direction than expected. They expected that the negative stimulus would increase cognitive control, and thus reaction times on the anti-saccade task would decrease. What they found is an increase in reaction times in the pro-saccade task however. In other words, subjects were slower to saccade towards the negative stimuli than neutral stimuli. In another study investigating the effects of emotion induction on attention, Cohen, Henik & Mor (2011) found that negative task cues impaired task performance in the absence of executive process activation. Findings from these studies, and other showing that processing of negative events may slow down performance of a subsequent task (Gehring et al., 1993; Notebaert et al., 2009; Pessoa, 2009; Cohen et al., 2011), might then predict that semantic categorization of ultra-rapidly presented images would be slower for negative images relative to neutral or positive images. In addition to separately considering the performance of affective and semantic categorization tasks under ultra-rapid presentation, comparing the performance of both tasks can help shed light on which process occurs first, and potentially illuminate whether one is dependent on the other.

## Cognitive Primacy Hypothesis vs. Affective Primacy Hypothesis

Most theories of emotional processing assume that affective processing is automatic, fast, and effortless (Bargh, 1997). This assumption has led some theorists to argue that the affective information of a scene can be extracted prior to the extraction of semantic information, a hypothesis known as the affective primacy hypothesis (S. T. Murphy & Zajonc, 1993; Zajonc, 1980). This hypothesis postulates that recognition of the affective information of a scene occurs prior to, and independently of, the perception of objects and scene information. Support for this hypothesis comes from several lines of research. First, psychophysiological studies presenting subliminal backwards-masked emotional images have found increases in both peripheral physiology (facial electromyographic responses: Dimberg, Thunberg, & Elmehed, 2000;

electrodermal responses: Glascher & Adolphs, 2003; Ohman & Soares, 1998) and central nervous system activity (Morris et al., 1998b), which suggests affective processing can occur without conscious awareness. Second, emotional priming studies have shown that task-irrelevant emotional images, relative to neutral images, influence the performance of subsequent categorization tasks, which has been taken to suggest that emotional processing is automatic and mandatory (Klauer & Musch, 2003). In addition, central to neuroscientific theories accounting for the prioritized processing of biologically prepared stimuli is the assertion that a dedicated neural system subserves the fast and automatic processing of emotional visual stimuli (Ohman & Mineka, 2001; Tamietto & de Gelder, 2010). A so-called "subcortical fast route" originating in the retina, continuing to the superior colliculus, on to the pulvinar and ending in the amygdala is presumed to serve the function of a dedicated processing unit for extraction of emotional information from visual stimuli (LeDoux, 1994).

In contrast to the affective primacy hypothesis, proponents of the cognitive primacy hypothesis argue that a scene's semantic information must first be extracted before affective meaning can be attributed to the scene (Nummenmaa, Hyona, & Calvo, 2010; Lazarus, 1984; Rolls, 1999; Storbeck & Clore, 2007). Evidence for the cognitive primacy hypothesis comes from several lines of research. First, studies of neurological patients with damaged or missing striate cortex (V1) have found that, while some patients can show somewhat accurate classification of facial expressions presented to their blind visual fields (de Gelder, Vroomen, Pourtois, & Weiskrantz, 1999), a phenomenon known as affective blindsight, this capacity does not generalize to emotional scenes (de Gelder, Pourtois, & Weiskrantz, 2002). Second, more recent affective priming studies have called into question whether subliminal affective priming can indeed occur with novel stimuli, as the effect was only seen with previously viewed images (Calvo & Nummenmaa, 2007) or with conscious recognition of facial identity (Nummenmaa, Peets, & Salmivalli, 2008). Third, more recent visual search studies have failed to establish pop-out effects for emotional relative to neutral targets (Batty, Cave, & Pauli, 2005; Calvo, Nummenmaa, & Avero, 2008; Fox et al., 2000; Horstmann & Bauland, 2006). Additionally, several recent findings in the affective neuroscience domain have called into question the plausibility of a "subcortical fast route" dedicated to extracting emotional information from visual stimuli (Pessoa & Adolphs, 2010; Pourtois, Schettino, & Vuilleumier, 2013). Finally, in a series of 6 experiments using brief presentations of emotional stimuli, a recent study by Nummenmaa, Hyona, & Calvo (2010) made a strong case that semantic categorization can be done faster than affective categorization, and that accurate semantic categorization must happen before affective categorization can successfully occur. Using both eye saccade and button presses subjects categorized the semantic category (animal vs. human) and valence (either negative vs. neutral or negative vs. positive) of emotional scene images briefly presented both foveally and extrafoveally, across the six experiments. They found that response times for semantic categorization were faster than for valence categorization using both eye saccades and button presses and when presented both foveally and extrafoveally. In different experiments, they also found that semantic category could be accurately categorized (at 20ms) for images presented at shorter presentation times than could valence category (at 40ms). Considered as a whole, this body of research provides strong evidence in support of the cognitive primacy hypothesis as opposed to the affective primacy hypothesis, although several limitations to these findings constrain the conclusions that can be drawn from them.

## The Current Study

While the studies conducted by Nummenmaa and colleagues (2011) provided perhaps the best evidence to date that the human visual system confers prioritized processing of emotional stimuli to briefly presented images, and that semantic information must be processed before affective information is processed (e.g the cognitive primacy hypothesis), they have several limitations. First, Nummenmaa, Hyona, & Calvo (2010) used all animate stimuli (animals vs. humans) in their experiments, and as such could not address the issue of whether prioritized processing occurs for only animate stimuli, as the biological preparedness hypothesis might suggest, or if it generalizes to inanimate stimuli as well. Second, Two alternative forced choice tasks (2AFC) were used in all of these studies, thus limiting the conclusions that can be drawn to a differentiation of only two classes (pleasant vs. unpleasant, emotional vs. non-emotional, animal vs. human) and not a differentiation between, for example the three valence levels of negative, neutral and positive or four semantic categories of animal, human, object and building. As the biological preparedness hypothesis predicts prioritized processing for negative, threatening, and animate stimuli, these task design limitations prevent the authors from differentiating task performance of negative, animate stimuli from any other type of stimulus, for example. Additionally, it is known that any response bias in categorization tasks can artificially inflate %-correct scores, and thus many researchers use a detection theory metric called d-prime which accounts for both false positive and hit rates to quantify classification performance. As the four previously mentioned ultra-rapid presentations studies looking at naturalistic emotional stimuli simply used %-correct scores to quantify task performance, their findings may have been mediated by response bias.

In order to address these limitations, we conducted two studies utilizing a large stimulus set of naturalistic emotional images covering multiple animate and inanimate semantic categories, with images being chosen such that each of these semantic categories contained equal numbers of exemplars categorised by independent raters as negative, neutral or positive. . Subjects viewed these stimuli at several brief presentation speeds (17ms, 33ms, 50ms in experiment 1 and 17ms & 33ms in experiment 2) and categorized each image for its semantic category (animal, human, object, building) and i emotional valence (negative, neutral, positive). By asking subjects to categorize the images into 3 valence categories we were able to not only address the question of whether emotional images are conferred with prioritized processing relative to neutral images, but whether there are differences between processing for negative and positive images relative to each others. Furthermore, using both animate and inanimate stimuli we were able to test competing predictions made by the biological preparedness and attentional narrowing theories. The biological preparedness theory predicts that prioritized processing, and thus higher categorization performance of both valence and semantic category, should be present for animate negative stimuli relative to all other semantic by valence image categories. Conversely, the attentional narrowing theory predicts that all negative stimuli (or even all arousing stimuli) would slow down processing of subsequent tasks and thus result in lower valence and semantic categorization performance for negative (and perhaps positive) stimuli. Finally, we addressed the debate between the cognitive and affective primacy hypotheses by comparing the conditional probabilities of getting the semantic task correct conditioned on getting the valence task correct, and vice versa. These conditional probabilities indicate if one process' success is necessary for the other's success. If the cognitive primacy hypothesis is true, then semantic categorization must occur before valence categorization, and thus it would predict

that the conditional probability of getting the valence task correct would only be above chance when the semantic task is done correctly. Conversely, the affective primacy hypothesis would predict that the conditional probability of getting the semantic categorization task correct would only be above chance when the valence categorization task was done correctly.

# Experiment 1

## Method

### Participants.

Nineteen volunteers (11 females, 8 males, mean age = 20.78, age range = 18-28) from the University of California, Berkeley's research subject participation pool participated in Experiment 1. Participants were compensated $20/1.5 hours. 1 participant was excluded from the analysis because of technical issues during the experiment that led to incomplete data. Written, informed consent was obtained for all participants before conducting the experiment. The experimental protocol for this study was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley.

### Materials.

1080 images were selected for use in the experiment. These each belonged to one of four semantic categories (animal, human, object, and buildings) and one of three valence categories (negative, neutral and positive). The image set comprised images from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and was supplemented with images found on the internet in order to have an equal number of images (90) within each of the 12 semantic by valence categories (e.g. positive-animals, negative-buildings, etc.). These four semantic categories were chosen to balance the images in terms of animacy/inanimacy, as animacy is known to be a major dimension of representation within visual cortex (Shai et al, 2013). Image semantic and valence categories was determined by three raters. Where there was a difference of opinion, the majority categorization was used. If all three raters differed, the image was removed from the stimulus set.

### Procedure.

Participants attended two experimental sessions, each approximately 1.5 hours in duration. During each session, participants were seated in a dark room in Tolman Hall on the UC Berkeley campus specifically used for psychological experiments. Images were presented at 12 x 12 degrees of visual angle on a CRT monitor with a refresh rate of 60Hz. Eighteen blocks, each consisting of 30 trials, were presented during each session. In total, each participant viewed 30 images from each semantic by valence category at each of the three presentation times used (17 ms, 33 ms, or 50 ms). The total image set was divided into two equal sets (540 images each - one for each session), and blocks were counterbalanced across participants to account for order effects. On each trial, the stimulus image was presented for a duration of either 17 ms, 33 ms, or 50 ms, after which four mask images were presented (Fig. 4-1). Mask images were generated using an algorithm described in Portilla & Simoncelli (2000) which use pairwise statistics of a wavelet pyramid to retain local image structure while breaking global image structure. In each

trial, the first mask image presented was generated from that trial's stimulus image. The remaining mask images were generated from one random stimulus image from each of the three semantic categories other than the trial image's category. Within each block, trial order was pseudo randomized such that images belonging to the same condition (semantic x valence category x stimulus presentation) were not presented sequentially. After stimulus and mask presentation, participants were instructed to respond as quickly as possible to a sequence of two forced choice categorization tasks, presented in the following order: (a) four-point semantic category (animal, human, object, or building), (b) three-point valence category (negative, neutral, or positive), followed by a nine-point arousal (emotional intensity) rating task.

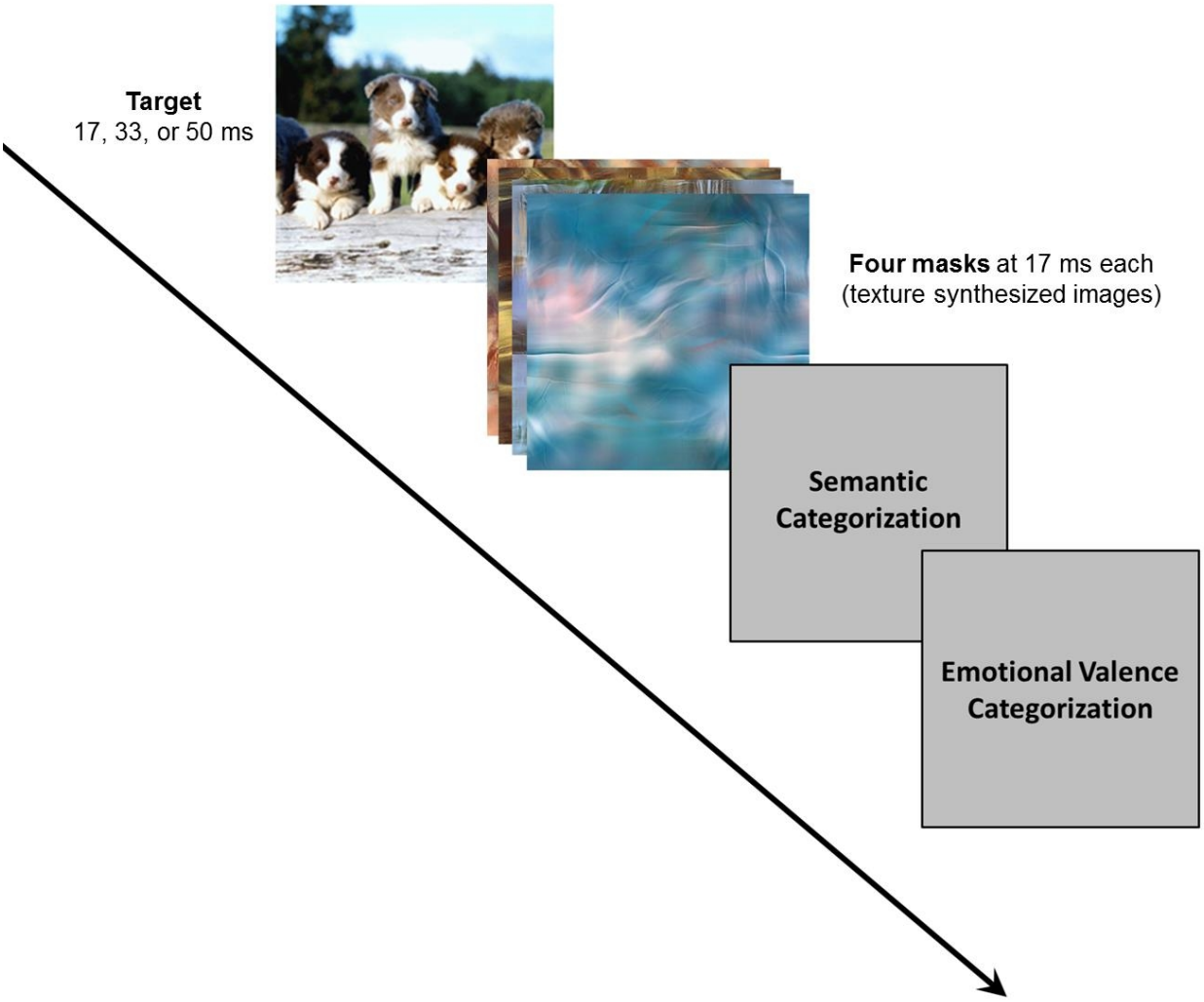


**Figure 4-1. Experimental Design of Experiment 1.**
Schematic showing the experimental design of experiment 1. Stimulus was shown for one of 3 presentation times (17ms, 33ms, 50ms) followed by presentation of 4 texture synthesized masks (17ms presentation time each) matched on low-level structural features to the stimulus just presented. A 4AFC semantic categorization task (animal, human, object, building) then a 3AFC valence categorization task (negative, neutral, positive) followed.

**Data Analysis.**

Task performance on both the semantic and valence categorization tasks were quantified per participant using a percentage of trials correctly categorized (%-correct), a pairwise d-prime (pairwise-d') score, and 3 (or 4) multiple alternative forced choice d-prime (mAFC-d') score. The %-correct scores were calculated for each presentation speed by semantic by valence condition (e.g. 17ms-positive-human), resulting in 36 %-correct scores per participant. Aggregate scores were also calculated by collapsed across either semantic, valence or semantic and valence categories resulting in 9, 12 and 3 scores respectively. These scores were calculated by dividing the number of correct trials by the total number of trials for each condition (or aggregate condition).

While informative, the %-correct score can be biased, and so to account for response bias d-prime scores were calculated. D-prime scores come from signal detection theory, and quantify the difference of Gaussian distributions which model the signal plus noise level and the noise level, resulting in an estimate of the signal level alone. Historically d-prime scores are calculated on 2 alternative forced choice (2AFC) tasks, however our task consisted of 4 choices and calculating pairwise-d' accounts for bias between the two pairs of categories included, but cannot account for bias in the categories excluded. Therefore, we used a technique for estimating d' on mAFC (4AFC in this case) tasks developed by Decarlo (2012) which uses maximum likelihood estimation and Gauss-Hermite quadrature for parameter estimation (R code generously supplied by Dr. Siegfried Macho, 2014). A single mAFC-d' score was calculated per participant for the semantic categorization task at each speed by valence condition (12 scores), and likewise a mAFC-d' score was calculated per participant for the valence categorization task at each speed by semantic condition (9 scores). Aggregate scores across all semantic by valence conditions at each speed were also calculated.

We additionally calculated the pairwise d' scores on the semantic categorization task for all possible pairs of semantic categories (6 pairs by valence by speed = 54 scores), and likewise on the valence categorization task on all possible pairs of valence categories (3 pairs by semantic by speed = 36 scores). Aggregate scores were also calculated as explained above. Only trials were the participant chose one of the two categories of the pairwise d' were used in the calculations. A hit rate was calculated as the proportion of trials answered correctly, and a false alarm rate was calculated as the proportion of trials answered incorrectly. The means of standard normal distributions for the hit rate (signal plus noise) and the false alarm (noise) were then calculated by using those proportions as p-values and entering them into a reverse standard normal cumulative distribution function. The false alarm mean was then subtracted from the hit rate mean to calculate the pairwise-d'.

To test whether %-correct scores were above chance at the group level, and whether both mAFC-d' and pairwise-d' were above 0 at the group level, t-tests were calculated across participant scores. Bonferroni correction for multiple comparisons were done, where results were grouped into families that accounted for the entire data set. For example, %-correct scores calculated for each speed (3 scores) were grouped together, resulting in a significance threshold of $\alpha/3$.

To detect group-level differences in %-correct scores for the semantic categorization task ANOVA was used. Factors of valence, presentation speed and semantic category were used in these ANOVAs. Tukey's post-hoc correction was used to determine which valence category comparisons differed significantly. Likewise, ANOVA was used to detect group-level

differences in %-correct scores of the valence categorization task with factors of semantic category, presentation speed and valence. Tukey's post-hoc correction was also used here to determine which semantic category comparisons differed significantly. To detect differences in group-level semantic task mAFC-d' scores, ANOVA was used with factors of presentation speed and valence. Tukey's post-hoc tests were also used here to determine which presentation speeds and valence categories differed significantly. ANOVA was also used to detect differences in valence task mAFC-d' scores for factors of presentation speed and semantic category. Tukey's post-hoc tests were also used here to determine which presentation speeds and semantic categories differed significantly.

To test for evidence supporting either the cognitive primacy or affective primacy hypothesis, data from both the semantic and valence categorization tasks were used in conjunction to calculate conditional probabilities for each participant. The conditional probability of getting the semantic task correct when getting the valence task correct, P(Semantic=correct | Valence=correct) or P(S|V), and the conditional probability of getting the semantic task correct when getting the valence task wrong P(S|~V) were calculated. Additionally, the conditional probability of getting the valence task correct when getting the semantic task correct, P(V|S), and getting the valence task correct when getting the semantic task wrong P(V|~S) were calculated. T-tests were used to determine significance of these probabilities at the group-level, and Bonferroni corrections for multiple comparisons were applied using the same families as described above. We looked for two alternate patterns of data that would arguably be more supportive of either (i) the cognitive primacy hypothesis or (ii) the affective primacy hypothesis. The first pattern comprised both group-level P(S|V) and group-level P(S|~V) being greater than chance, (indicating that the semantic task was done correctly regardless of whether the valence task was done correctly) and the group-level P(V|S) being greater than chance but the group-level P(V|~S) being at or below chance, . This pattern of data is consistent with the cognitive primacy hypothesis. Conversely, when both the group-level P(V|S) and P(V|~S) were significantly greater than chance, and the group-level P(S|V) was significantly greater than chance but the P(S|~V) was not significantly greater than chance, the pattern of data was labelled as consistent with the affective primacy hypothesis. We explored the presence of either of these patterns at each speed by semantic by valence condition (36 data cells). Only those conditions were the marginal probability of getting both semantic P(S) and valence P(V) correct were significantly above chance were considered. All analyses apart from the calculation of the mAFC-d' scores were done using custom scripts written in Matlab 2017a (The Mathworks, Inc).

## Results

### Semantic Categorization Task.

Descriptive statistics of group-level %-correct scores for the semantic categorization task are shown in Table 4-1. Scores are shown for each speed by semantic by valence condition, as well as for speed by semantic, speed by valence and just speed aggregate conditions. To determine which scores were significantly above chance performance (25% for the semantic task), one-tailed t-tests were used across single subject scores. Results indicated that performance in all conditions and aggregate conditions were significantly above chance ($ps < .05$) after Bonferroni correction for multiple comparisons (Table 4-2 and Fig. 4-2A-C). Near ceiling

performance was seen for many categories at the 33ms and 50ms presentation times. Using %-correct accuracy scores is vulnerable to biases, so to check whether participants showed bias we investigating the distribution of possible responses on incorrect trials. Participants selected object more than 50% of the time, indicating a strong bias to guess object under uncertainty (Fig. 4-2D). Thus, mAFC-d' scores were calculated to account for bias, resulting in a single score at each speed by valence level, for each participant.

**Table 4-1: Semantic Categorization Task %-Correct Descriptive Statistics.**

| Experiment 1 17ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 48.3 | 21.6 | 39.5 | 84.1 | 16.2 | 40.6 | 84.3 | 18.8 | 49.9 | 81.4 | 19.2 | 46.7 |
| | Human | 56.9 | 14.1 | 51.1 | 87.2 | 11.6 | 61.5 | 93.7 | 10.7 | 74.1 | 86.4 | 15.0 | 63.8 |
| | Object | 64.1 | 15.0 | 57.9 | 90.6 | 15.8 | 65.7 | 86.1 | 15.3 | 68.6 | 85.2 | 15.8 | 66.8 |
| | Building | 36.3 | 15.0 | 30.20 | 82.8 | 13.4 | 54.1 | 83.0 | 16.4 | 53.3 | 75.6 | 18.5 | 47.8 |
| | TOTAL | 51.4 | 19.4 | 47.6 | 86.2 | 16.9 | 58.0 | 86.8 | 17.8 | 64.2 | 82.2 | 19.2 | 58.00 |

| Experiment 1 33ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 75.9 | 16.6 | 69.1 | 84.1 | 12.6 | 78.9 | 84.3 | 13.2 | 78.8 | 81.4 | 14.5 | 78.1 |
| | Human | 78.3 | 10.2 | 74.1 | 87.2 | 9.0 | 83.5 | 93.7 | 5.0 | 91.7 | 86.4 | 10.4 | 84.1 |
| | Object | 79.1 | 11.9 | 74.2 | 90.6 | 7.9 | 87.3 | 86.1 | 7.3 | 83.1 | 85.2 | 10.3 | 71.7 |
| | Building | 61.1 | 16.2 | 54.4 | 82.8 | 12.2 | 77.8 | 83.0 | 12.7 | 77.8 | 75.6 | 17.1 | 71.7 |
| | TOTAL | 73.6 | 15.6 | 70.6 | 86.2 | 10.8 | 84.0 | 86.8 | 10.8 | 84.6 | 82.2 | 13.9 | 80.6 |

| Experiment 1 50ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 84.4 | 11.4 | 79.8 | 88.1 | 7.7 | 85 | 92.0 | 10.3 | 87.8 | 88.2 | 10.2 | 85.9 |
| | Human | 85.6 | 7.3 | 82.6 | 94.3 | 4.5 | 92.4 | 94.6 | 4.6 | 92.7 | 91.5 | 7.0 | 89.9 |
| | Object | 85.7 | 7.7 | 82.6 | 85.9 | 5.5 | 93.7 | 93.1 | 6.7 | 90.4 | 91.6 | 7.9 | 89.8 |
| | Building | 71.9 | 13.4 | 66.4 | 86.9 | 9.9 | 82.8 | 88.7 | 8.8 | 85.1 | 82.5 | 13.1 | 79.5 |
| | TOTAL | 81.9 | 11.6 | 79.6 | 91.3 | 8.1 | 89.7 | 92.1 | 8.0 | 90.6 | 88.4 | 10.5 | 87.3 |

| Experiment 2 17ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 50.3 | 14.3 | 46.1 | 49.2 | 17.6 | 44.0 | 57.5 | 15.2 | 53.0 | 52.3 | 16.0 | 49.6 |
| | Human | 52.2 | 13.0 | 48.4 | 62.5 | 14.3 | 58.3 | 84.2 | 10.0 | 81.3 | 66.3 | 18.3 | 63.3 |
| | Object | 62.4 | 14.4 | 58.2 | 66.2 | 15.1 | 61.7 | 73.4 | 10.2 | 70.4 | 67.3 | 14.1 | 65.0 |
| | Building | 38.3 | 14.1 | 34.1 | 66.0 | 13.1 | 62.1 | 79.6 | 12.3 | 76.0 | 61.3 | 21.7 | 57.7 |
| | TOTAL | 50.8 | 16.3 | 48.5 | 61.0 | 16.5 | 58.6 | 73.7 | 15.7 | 71.4 | 61.8 | 18.6 | 60.3 |

| Experiment 2 33ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| | Animal | 72.1 | 14.0 | 68.0 | 78.9 | 10.1 | 75.9 | 86.1 | 11.3 | 82.7 | 79.0 | 13.1 | 76.8 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Semantic** | **Human** | 73.5 | 10.2 | 70.5 | 88.6 | 6.6 | 86.6 | 96.1 | 6.5 | 94.1 | 86.1 | 12.3 | 84.0 |
| | **Object** | 77.1 | 9.0 | 74.4 | 88.2 | 6.6 | 86.2 | 86.6 | 8.7 | 84.0 | 83.9 | 9.5 | 82.4 |
| | **Building** | 66.6 | 13.0 | 62.7 | 84.8 | 10.1 | 81.9 | 91.8 | 7.8 | 89.5 | 81.1 | 14.9 | 78.6 |
| | **TOTAL** | 72.3 | 12.2 | 70.6 | 85.1 | 9.3 | 83.8 | 90.1 | 9.6 | 88.7 | 82.5 | 12.8 | 81.5 |

**Table 4-2: Semantic Categorization Task %-Correct t-test Results.**

| Experiment 1 17ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 4.6 (17) | <.01 | 1.1 | 5.8 (17) | <.001 | 1.4 | 7.3 (17) | <.001 | 1.7 | 9.9 (53) | <.001 | 1.4 |
| | Human | 9.6 (17) | <.001 | 2.3 | 15.1 (17) | <.001 | 3.6 | 21.3 (17) | <.001 | 5.0 | 20.7 (53) | <.001 | 2.8 |
| | Object | 11.1 (17) | <.001 | 2.6 | 12.7 (17) | <.001 | 3. | 13.8 (17) | <.001 | 3.3 | 21.2 (53) | <.001 | 2.9 |
| | Building | 3.2 (18) | <.05 | 0.8 | 10.9 (17) | <.001 | 2.6 | 9.0 (17) | <.001 | 2.1 | 10.7 (53) | <.001 | 1.5 |
| | TOTAL | 11.6 (71) | <.001 | 1.4 | 18.3 (71) | <.001 | 2.2 | 20.4 (71) | <.001 | 2.4 | 26.9 (215) | <.001 | 1.8 |

| Experiment 1 33ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 13.1 (17) | <.001 | 3.1 | 20.0 (17) | <.001 | 4.7 | 19.0 (17) | <.001 | 4.5 | 28.6 (53) | <.001 | 3.9 |
| | Human | 22.1 (17) | <.001 | 5.2 | 29.3 (17) | <.001 | 6.9 | 58.7 (17) | <.001 | 13.8 | 43.4 (53) | <.001 | 5.9 |
| | Object | 19.3 (17) | <.001 | 4.6 | 35.4 (17) | <.001 | 8.3 | 35.3 (17) | <.001 | 8.3 | 43.2 (53) | <.001 | 5.9 |
| | Building | 9.4 (17) | <.001 | 2.2 | 20.1 (17) | <.001 | 4.7 | 19.4 (17) | <.001 | 4.6 | 21.8 (53) | <.001 | 3.0 |
| | TOTAL | 26.5 (71) | <.001 | 3.1 | 48.1 (71) | <.001 | 5.7 | 48.5 (71) | <.001 | 5.7 | 60.4 (215) | <.001 | 4.1 |

| Experiment 1 50ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 22.1 (17) | <.001 | 5.2 | 34.8 (17) | <.001 | 8.2 | 27.6 (17) | <.001 | 6.5 | 45.4 (53) | <.001 | 6.2 |
| | Human | 35.1 (17) | <.001 | 8.3 | 64.6 (17) | <.001 | 15.2 | 64.3 (17) | <.001 | 15.2 | 70.2 (53) | <.001 | 9.5 |
| | Object | 33.7 (17) | <.001 | 7.9 | 54.2 (17) | <.001 | 12.8 | 43.1 (17) | <.001 | 10.2 | 62.2 (53) | <.001 | 8.5 |
| | Building | 14.8 (17) | <.001 | 3.5 | 26.4 (17) | <.001 | 6.2 | 30.7 (17) | <.001 | 7.2 | 32.2 (53) | <.001 | 4.4 |
| | TOTAL | 41.5 (71) | <.001 | 4.9 | 69.6 (71) | <.001 | 8.2 | 70.9 (71) | <.001 | 8.4 | 89.2 (215) | <.001 | 6.1 |

| Experiment 2 17ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 10.1 (32) | <.001 | 1.8 | 7.9 (32) | <.001 | 1.4 | 12.3 (32) | <.001 | 2.1 | 17.0 (98) | <.001 | 1.7 |
| | Human | 12.0 (32) | <.001 | 2.1 | 15.0 (32) | <.001 | 2.6 | 34.1 (32) | <.001 | 5.9 | 22.5 (98) | <.001 | 2.3 |
| | Object | 14.9 (32) | <.001 | 2.6 | 15.6 (32) | <.001 | 2.7 | 27.3 (32) | <.001 | 4.8 | 30.0 (98) | <.001 | 3 |
| | Building | 5.4 (32) | <.001 | 0.9 | 17.9 (32) | <.001 | 3.1 | 25.4 (32) | <.001 | 4.4 | 16.7 (98) | <.001 | 1.7 |
| | TOTAL | 18.2 (131) | <.001 | 1.6 | 25.0 (131) | <.001 | 2.2 | 35.6 (131) | <.001 | 3.1 | 39.3 (397) | <.001 | 2 |

| Experiment 2 33ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 19.4 (32) | <.001 | 3.4 | 30.7 (32) | <.001 | 5.3 | 31.0 (32) | <.001 | 5.4 | 41.0 (98) | <.001 | 4.1 |
| | Human | 27.2 (32) | <.001 | 4.7 | 55.2 (32) | <.001 | 9.6 | 62.5 (32) | <.001 | 10.9 | 49.4 (98) | <.001 | 5.0 |
| | Object | 33.4 (32) | <.001 | 5.8 | 54.8 (32) | <.001 | 9.5 | 40.7 (32) | <.001 | 7.1 | 62.0 (98) | <.001 | 6.2 |
| | Building | 18.4 (32) | <.001 | 3.2 | 33.9 (32) | <.001 | 5.9 | 49.4 (32) | <.001 | 8.6 | 37.4 (98) | <.001 | 3.8 |
| | TOTAL | 44.6 (131) | <.001 | 3.9 | 74.2 (131) | <.001 | 6.5 | 78.1 (131) | <.001 | 6.8 | 89.2 (397) | <.001 | 4.5 |

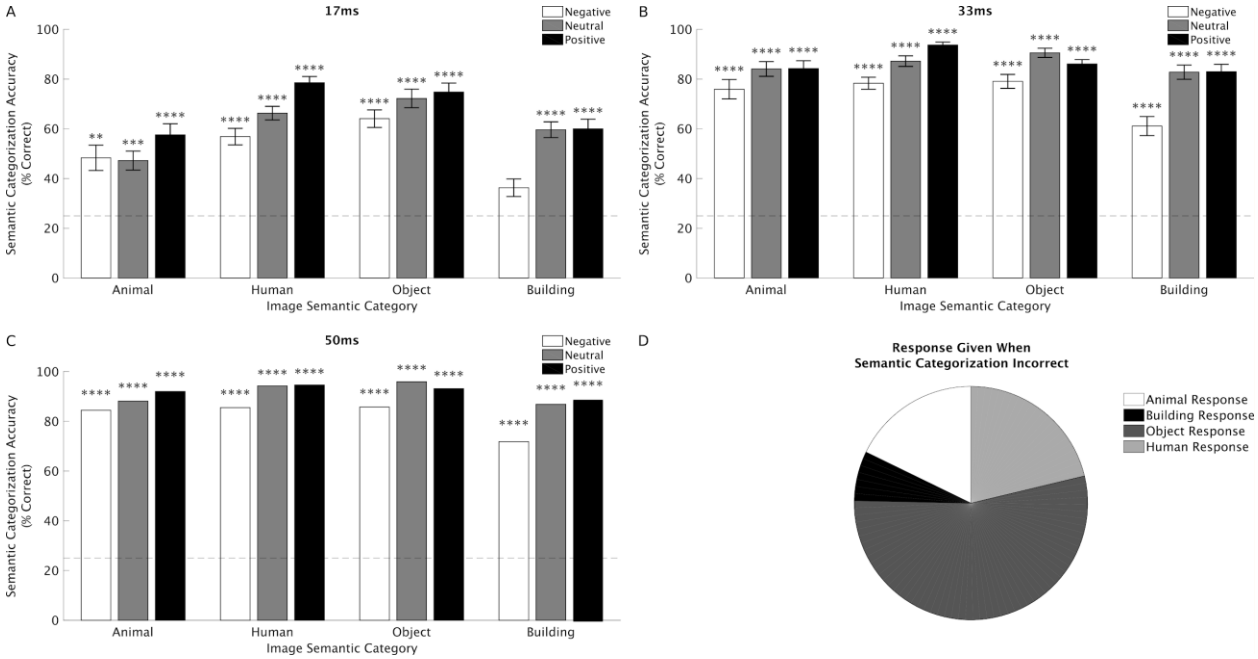


**Figure 4-2. Semantic Categorization Task Performance for Experiment 1.**
(A-C) Bar plots showing group-averaged percentage of semantic categorization task trials classified correctly in experiment 1. Each plot shows results broken down by semantic and valence categories from trials at one of the 3 presentation speeds (17ms in A, 33ms in B, and 50ms in C). Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (25%), and stars above each bar indicate which presentation speed by semantic by valence conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Group level task performance at all semantic by valence conditions was significantly above chance at all three presentation speeds. (D) Pie chart showing proportion of incorrectly classified semantic categorization task responses for each semantic category, across all subjects. As objects were chosen for more than 50% of incorrect responses this figure shows a clear bias to select object when the semantic category was unknown.

Descriptive statistics of group-level mAFC-d' scores are shown in Table 4-3 (Fig. 4-3, and Table 4-4). mAFC-d' scores revealed that semantic categorization performance was above chance for each valence condition at each duration of presentation (ps < .001) and for each duration (ps < .001, collapsing across valence levels). To test for differences in semantic categorization task performance as a function of image valence and presentation duration, a two-way ANOVA was conducted using factors of presentation speed (17, 33, & 50 ms) and valence (negative, neutral, & positive), and dependent variable: semantic categorization mAFC-d' scores There were significant main effects of both image presentation speed $F(2,161)=16.1$, $p<.001$, partial $\eta^2=0.17$, and valence $F(2,161)=5.2$, $p<.01$, partial $\eta^2=0.06$ were significant, but the interaction was not ($p>.1$). As the main effects of presentation speed and valence were both significant, one way ANOVAs collapsing across levels of valence and speed, respectively, were conducted. The one-way ANOVA addressing the influence of presentation speed (17, 33, or 50ms) on semantic categorization mAFC-d' scores was significant $F(2,159)=15.0$, $p<.001$, partial $\eta^2=0.16$. Tukey's post-hoc tests revealed that mAFC-d' scores at from images presented at 17ms were significantly smaller than those presented at 50ms, M difference=-2.65, $p<.001$, 95% CI=(-3.81, -1.50), and mAFC-d' scores were significantly smaller for when presented at 33ms versus 50ms, M difference=-1.72, $p<.01$, 95% CI=(-2.87, -0.56), however the difference

between 17ms and 33ms was not significant (p>.1). The one-way ANOVA examining the influence of valence category (negative, neutral, positive) on mAFC-d' scores , collapsing across speed of presentation, was also significant f(2,159)=4.3, p<.05, partial $\eta^2$=0.05. Tukey's post-hoc test revealed that only the difference between negative and positive stimuli was significant, M difference=-1.48, p<.05, 95% CI=(-2.70, -0.25), while the differences between negative and neutral (p>,1) and neutral and positive (p>.1) were not. mAFC-d' scores were significantly lower for negative than positive stimuli.

Even though no significant interaction was detected, one-way ANOVAs were conducted at each presentation speed, given that this study is investigating ultra-fast perception, and hence exploring performance at the different presentation speeds was of interest. These ANOVAs revealed significant main effects of valence at all three presentation speeds: at 17ms f(2,53)=14.2, p<.001, $\eta^2$=0.36, at 33ms f(2,53)=10.8, p<.001, $\eta^2$=0.30, and 50ms f(2,53)=9.6, p<.001, $\eta^2$=0.27. Tukey's post-hoc test was then used to find which levels of valence differed significantly at each presentation speed. Critically, at all 3 speeds negative stimuli were semantically categorized more poorly than both neutral and positive stimuli; at 17ms negative vs neutral (M difference=-0.35), p<.01, 95% CI=(-0.63,-0.08), at 17ms negative vs positive (M difference=-0.61), p<.001, 95% CI=(-0.88,-0.33), at 33ms negative vs neutral (M difference=-0.63), p<.001, 95% CI=(-1.00,-0.27), at 33ms negative vs positive (M difference=-0.58), p=.001, 95% CI=(-0.94,-0.21), at 50ms negative vs neutral (M difference=-1.12), p<.001, 95% CI=(-1.78,-0.44), and at 50ms negative vs positive (M difference=-1.00), p=.002, 95% CI=(-1.67,-0.33). Neutral and positive stimuli did not differ in semantic categorization mAFC-d' scores at any presentation speed (ps > .1).

**Table 4-3: Semantic Categorization Task mAFC-d' Descriptive Statistics.**

| Experiment 1 mAFC-d' | | Speed | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 17ms | | | 33ms | | | 50ms | | |
| | | M d' | SD | 95% CI | M d' | SD | 95% CI | M d' | SD | 95% CI |
| Valence | Negative | 1.1 | 0.3 | 0.9 | 1.9 | 0.5 | 1.7 | 2.2 | .3 | 2.1 |
| | Neutral | 1.4 | 0.3 | 1.3 | 2.6 | 0.5 | 2.4 | 3.3 | 1.0 | 2.9 |
| | Positive | 1.7 | 0.4 | 1.5 | 2.5 | 0.3 | 2.4 | 3.2 | 1.0 | 2.8 |
| | TOTAL | 1.4 | 0.4 | 1.3 | 2.3 | 0.5 | 2.2 | 2.9 | 1.0 | 2.7 |

| Experiment 2 mAFC-d' | | Speed | | | | | |
|---|---|---|---|---|---|---|---|
| | | 17ms | | | 33ms | | |
| | | M d' | SD | 95% CI | M d' | SD | 95% CI |
| Valence | Negative | 1.0 | 0.3 | 0.9 | 1.7 | 0.3 | 1.8 |
| | Neutral | 1.3 | 0.3 | 1.2 | 2.3 | 0.3 | 2.2 |
| | Positive | 1.8 | 0.3 | 1.7 | 2.8 | 0.6 | 2.6 |
| | TOTAL | 1.4 | 0.5 | 1.3 | 2.3 | 0.6 | 2.2 |

**Table 4-4: Semantic Categorization Task mAFC-d' t-test Results.**

| Experiment 1 mAFC-d' | | Speed | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 17ms | | | 33ms | | | 50ms | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Valence | Negative | 14 (32) | <.001 | 3.3 | 15.7 (32) | <.001 | 3.7 | 32.6 (32) | <.001 | 7.7 |
| | Neutral | 18.8 (32) | <.001 | 4.4 | 22.4 (32) | <.001 | 5.3 | 14.1 (32) | <.001 | 3.3 |
| | Positive | 19.1 (32) | <.001 | 4.5 | 33.5 (32) | <.001 | 7.9 | 13.7 (32) | <.001 | 3.2 |
| | TOTAL | 24.5 (98) | <.001 | 3.3 | 32.5 (98) | <.001 | 4.4 | 22.4 (98) | <.001 | 3.1 |

| Experiment 2 mAFC-d' | | Speed | | | | | |
|---|---|---|---|---|---|---|---|
| | | 17ms | | | 33ms | | |
| | | t (df) | p | d | t (df) | p | d |
| Valence | Negative | 20.9 (32) | <.001 | 3.6 | 35.1 (32) | <.001 | 6.1 |
| | Neutral | 24.0 (32) | <.001 | 4.2 | 40.4 (32) | <.001 | 7.0 |
| | Positive | 29.7 (32) | <.001 | 5.2 | 25.7 (32) | <.001 | 4.5 |
| | TOTAL | 29.9 (98) | <.001 | 3.0 | 36.5 (98) | <.001 | 3.7 |

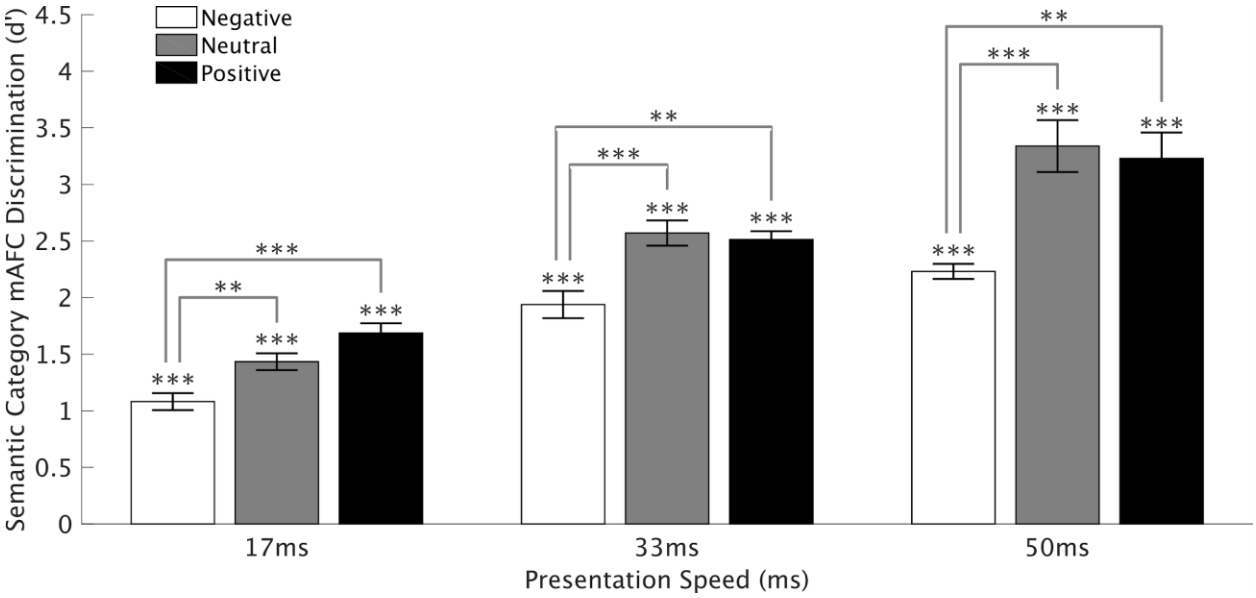


**Figure 4-3. Semantic Categorization task mAFC-d' Scores for Experiment 1.**
To account for the observed response bias in the semantic categorization task from experiment 1 (shown in Fig. 4-2) multiple alternative forced choice (mAFC) d' scores were calculated which account for both false positive and hit rates. A score of 0 indicates chance performance. Results are broken down by presentation speed and valence category (negative, neutral, positive). Error bars show standard error (S.E) across subjects, stars above each bar indicate which presentation speed by valence conditions were classified with above chance performance, while lines between bars indicate which presentation speed by valence conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Performance was significantly above chance at all presentation speed by valence conditions. Critically, negative stimuli were less accurately classified than both neutral and positive stimuli were at all presentation speeds.

### Valence Categorization Task.

Descriptive statistics for the valence categorization task are shown in Table 4-5 (and Table 4-6 and Fig. 4-4). As with the semantic categorization task, % correct valence categorization scores are shown for each speed by semantics by valence condition, as well as aggregate conditions of speed by semantics, speed by valence and speed only. One-tailed t-tests comparing %-correct scores with chance level (33.3%) revealed that at the fastest presentation speed of 17ms, all neutral categories were significantly identified above chance (ps < .001), however this is likely due to a participant bias to respond with neutral when guessing. Approximately 63% of incorrect trials were responded to with neutral, as can be seen in Figure 4-4D. At 33ms, positive human images were correctly categorized for valence significantly above chance, M=61.9%, t(17)=6.8, p<.001, d=1.6. This was also true for all neutral categories (ps < .05), except neutral humans. At 50ms, positive human images were again correctly categorized at an above chance level M=63.1%, t(17)=6.6, p<.001, d=1.6. At 50ms, categorization performance of three negative conditions was significantly above chance, namely: negative animals M=49.1%, t(17)=4.7, p<.01, d=1.1, negative humans M=54.4%, t(17)=6.4, p<.001, d=1.5, and negative objects M=58.7%, t(17)=6.5, p<.001, d=1.5. As was seen at 33ms, all neutral categories were categorized correctly at above chance performance (ps < .05), except neutral human categories (p>.1).

**Table 4-5: Valence Categorization Task %-Correct Descriptive Statistics.**

| Experiment 1 17ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 22.4 | 13.5 | 16.9 | 72.4 | 21.3 | 63.7, | 21.5 | 23.8 | 11.7 | 38.8 | 31.0 | 31.7 |
| | Human | 23.7 | 14.4 | 17.8 | 62.0 | 19.5 | 54.0 | 41.5 | 18.7 | 33.8 | 42.4 | 23.5 | 37.1 |
| | Object | 23.7 | 14.4 | 17.8 | 73.7 | 21.2 | 65,.0 | 19.4 | 16.6 | 12.6 | 39.0 | 30.3 | 32.1 |
| | Building | 13.3 | 12.5 | 8.2 | 75.6 | 19.1 | 67.7, | 22.4 | 20.4 | 14.1 | 37.1 | 32.7 | 29.7 |
| | TOTAL | 20.8 | 14.1 | 18.0 | 70.9 | 20.5 | 66.9 | 26.2 | 21.6 | 22.0 | 39.3 | 29.4 | 36.0 |

| Experiment 1 33ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 38.9 | 13.1 | 33.5 | 60.9 | 32.4 | 47.7 | 38.5 | 29.7 | 26.4 | 46.1 | 28.0 | 39.7 |
| | Human | 41.3 | 17.0 | 34.3 | 50.9 | 25.1 | 40.6 | 61.9 | 17.9 | 54.5 | 51.4 | 21.6 | 46.4 |
| | Object | 45.9 | 16.9 | 39.0 | 72.4 | 20.3 | 64.1 | 25.9 | 18.3 | 18.4 | 48.1 | 26.5 | 42.1 |
| | Building | 20.0 | 13.6 | 14.4 | 64.4 | 29.0 | 52.6 | 39.3 | 29.6 | 27.1 | 41.2 | 30.8 | 34.2 |
| | TOTAL | 36.5 | 17.9 | 33.0 | 62.2 | 27.6 | 56.8 | 41.4 | 27.3 | 36.0 | 46.7 | 27.0 | 43.7 |

| Experiment 1 50ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 49.1 | 14.2 | 43.2 | 59.8 | 31.5 | 46.9 | 43.7 | 32.3 | 30.5 | 50.9 | 27.6 | 44.6 |
| | Human | 54.4 | 14.0 | 48.7 | 46.7 | 25.4 | 36.2 | 63.1 | 19.1 | 55.3 | 54.8 | 20.8 | 50.0 |
| | Object | 58.7 | 16.6 | 51.9 | 68.9 | 26.1 | 58.2 | 37.0 | 20.9 | 28.5 | 54.9 | 25.0 | 49.2 |
| | Building | 26.5 | 16.8 | 19.6 | 66.3 | 31.0 | 53.6 | 45.9 | 32.1 | 32.8 | 46.2 | 31.6 | 39.0 |
| | TOTAL | 47.2 | 19.6 | 43.3 | 60.4 | 29.3 | 54.7 | 47.5 | 28.0 | 42.0 | 51.7 | 26.6 | 48.7 |

| Experiment 2 17ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 42.2 | 16 | 37.5 | 65.3 | 13.2 | 61.4 | 26.9 | 10.9 | 23.6 | 44.8 | 20.8 | 41.3 |
| | Human | 32.5 | 11.9 | 29 | 56.3 | 16 | 51.5 | 47.4 | 15.7 | 42.8 | 45.4 | 17.5 | 42.5 |
| | Object | 32 | 13.7 | 28 | 70.2 | 14.5 | 65.9 | 30.2 | 14.5 | 25.9 | 44.1 | 23.3 | 40.3 |
| | Building | 20.5 | 13 | 16.7 | 73.4 | 17.8 | 68.2 | 28.5 | 22 | 22 | 40.8 | 29.4 | 35.9 |
| | TOTAL | 31.8 | 15.6 | 29.6 | 66.3 | 16.6 | 63.9 | 33.2 | 18.1 | 30.6 | 43.8 | 23.1 | 41.9 |

| Experiment 2 33ms | | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI | M % | SD | 95% CI |
| Semantic | Animal | 62.9 | 18.4 | 57.5 | 55.5 | 18.9 | 49.9 | 51.4 | 16 | 46.7 | 56.6 | 18.3 | 53.5 |
| | Human | 52.0 | 14.5 | 47.7 | 46.6 | 18 | 41.2 | 71.4 | 13 | 67.6 | 56.7 | 18.6 | 53.6 |
| | Object | 49.7 | 15.1 | 45.2 | 69.1 | 14.5 | 64.8 | 56.4 | 15.9 | 51.7 | 58.4 | 17.1 | 55.5 |
| | Building | 28.9 | 13.3 | 25 | 66.9 | 23.8 | 59.8 | 46.7 | 26.9 | 38.7 | 47.5 | 26.9 | 43 |
| | TOTAL | 48.4 | 19.7 | 45.5 | 59.5 | 21 | 56.5 | 56.5 | 20.7 | 53.5 | 54.8 | 20.9 | 53 |

**Table 4-6: Valence Categorization Task %-Correct t-test Results.**

| Experiment 1 17ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | -3.4 (17) | >.1 | -0.8 | 7.8 (17) | <.001 | 1.8 | -2.1 (17) | >.1 | -0.5 | 1.3 (53) | >.1 | 0.2 |
| | Human | -2.8 (17) | >.1 | -0.7 | 6.2 (17) | <.001 | 1.5 | 1.8 (17) | >.1 | 0.4 | 2.8 (53) | <.05 | 0.4 |
| | Object | -2.8 (17) | >.1 | -0.7 | 8.1 (17) | <.001 | 1.9 | -3.5 (17) | >.1 | -0.8 | 1.4 (53) | >.1 | 0.2 |
| | Building | -6.8 (17) | >.1 | -1.6 | 9.4 (17) | <.001 | 2.2 | -2.3 (17) | >.1 | -0.5 | 0.8 (53) | >.1 | 0.1 |
| | TOTAL | -7.5 (71) | >.1 | -0.9 | 15.5 (71) | <.001 | 1.8 | -2.8 (71) | >.1 | -0.3 | 3.0 (215) | <.01 | 0.2 |

| Experiment 1 33ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 1.8 (17) | >.1 | 0.4 | 3.6 (17) | <.05 | 0.9 | 0.7 (17) | >.1 | 0.2 | 3.3 (53) | <.01 | 0.5 |
| | Human | 2.0 (17) | >.1 | 0.5 | 3.0 (17) | >.1 | 0.7 | 6.8 (17) | <.001 | 1.6 | 6.1 (53) | <.001 | 0.8 |
| | Object | 3.2 (17) | >.1 | 0.7 | 8.2 (17) | <.001 | 1.9 | -1.7 (17) | >.1 | -0.4 | 4.1 (53) | <.001 | 0.6 |
| | Building | -4.2 (17) | >.1 | -1 | 4.6 (17) | <.01 | 1.1 | 0.8 (17) | >.1 | 0.2 | 1.9 (53) | >.1 | 0.3 |
| | TOTAL | 1.5 (71) | >.1 | 0.2 | 8.9 (71) | <.001 | 1.0 | 2.5 (71) | =.07 | 0.3 | 7.3 (215) | <.001 | 0.5 |

| Experiment 1 50ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 4.7 (17) | <.01 | 1.1 | 3.6 (17) | <.05 | 0.8 | 1.4 (17) | >.1 | 0.3 | 4.7 (53) | <.001 | 0.6 |
| | Human | 6.4 (17) | <.001 | 1.5 | 2.2 (17) | >.1 | 0.5 | 6.6 (17) | <.001 | 1.6 | 7.6 (53) | <.001 | 1.0 |
| | Object | 6.5 (17) | <.001 | 1.5 | 5.8 (17) | <.001 | 1.4 | 0.8 (17) | >.1 | 0.2 | 6.3 (53) | <.001 | 0.9 |
| | Building | -1.7 (17) | >.1 | -0.4 | 4.5 (17) | <.01 | 1.1 | 1.7 (17) | >.1 | 0.4 | 3.0 (53) | <.05 | 0.4 |
| | TOTAL | .06 (71) | <.001 | 0.7 | 7.8 (71) | <.001 | 0.9 | 4.3 (71) | <.001 | 0.5 | 10.1 (215) | <.001 | 0.7 |

| Experiment 2 17ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 3.2 (32) | <.05 | 0.6 | 13.9 (32) | <.001 | 2.4 | -3.4 (32) | >.1 | -0.6 | 5.5 (98) | <.001 | 0.6 |
| | Human | -0.4 (32) | >.1 | -0.1 | 8.2 (32) | <.001 | 1.4 | 5.2 (32) | <.001 | 0.9 | 6.8 (98) | <.001 | 0.7 |
| | Object | -0.5 (32) | >.1 | -0.1 | 14.6 (32) | <.001 | 2.5 | -1.2 (32) | >.1 | -0.2 | 4.6 (98) | <.001 | 0.5 |
| | Building | -5.7 (32) | >.1 | -1 | 12.9 (32) | <.001 | 2.3 | -1.3 (32) | >.1 | -0.2 | 2.5 (98) | =.05 | 0.3 |
| | TOTAL | -1.1 (131 | >.1 | -0.1 | 22.8 (131) | <.001 | 2 | -0.1 (131) | >.1 | 0.0 | 9.0 (397) | <.001 | 0.5 |

| Experiment 33ms | Valence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Negative | | | Neutral | | | Positive | | | TOTAL | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| Semantic | Animal | 9.2 (32) | <.001 | 1.6 | 6.7 (32) | <.001 | 1.2 | 6.5 (32) | <.001 | 1.1 | 12.7 (98) | <.001 | 1.3 |
| | Human | 7.4 (32) | <.001 | 1.3 | 4.2 (32) | <.01 | 0.7 | 16.9 (32) | <.001 | 2.9 | 12.5 (98) | <.001 | 1.3 |
| | Object | 6.2 (32) | <.001 | 1.1 | 14.2 (32) | <.001 | 2.5 | 8.3 (32) | <.001 | 1.5 | 14.6 (98) | <.001 | 1.5 |
| | Building | -1.9 (32) | >.1 | -0.3 | 8.1 (32) | <.001 | 1.4 | 2.8 (32) | =.09 | 0.5 | 5.2 (98) | <.001 | 0.5 |
| | TOTAL | 8.8 (131) | <.001 | 0.8 | 14.3 (131) | <.001 | 1.2 | 12.8 (131) | <.001 | 1.1 | 20.4 (397) | <.001 | 1 |

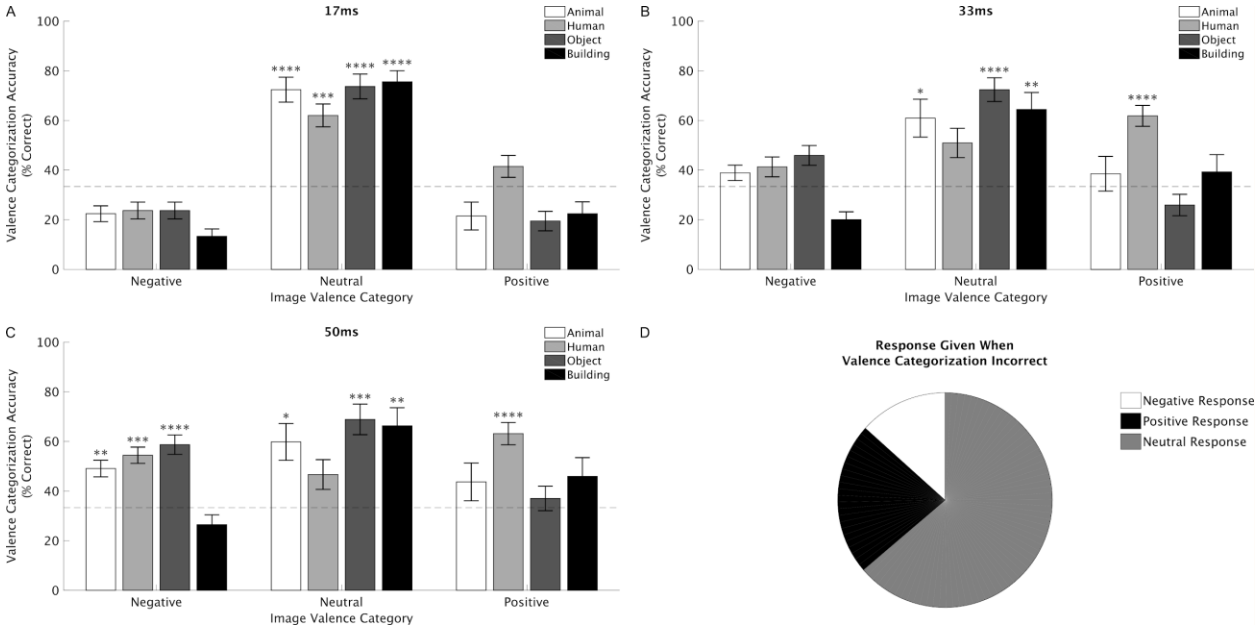


**Figure 4-4. Valence Categorization Task Performance for Experiment 1.**
(A-C) Bar plots showing group-averaged percentage of valence categorization task trials classified correctly in experiment 1. Each plot shows results broken down by valence and semantic categories from trials at one of the 3 presentation speeds (17ms in A, 33ms in B, and 50ms in C). Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (33%), and stars above each bar indicate which presentation speed by valence by semantic conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). At 33ms and 50ms, images of positive humans were classified with above chance performance, while at 50ms negative images of animals, humans and objects were all classified with above chance performance. (D) Pie chart showing proportion of incorrectly classified valence categorization task responses for each valence category, across all subjects. As neutral were clearly chosen for more than 50% of incorrect responses this figure shows a clear bias to select neutral when the valence category was unknown.

To account for the observed neutral response bias in the valence categorization task, mAFC-d' scores were also calculated. Table 4-7 contains descriptive statistics for the mAFC-d' scores. One tailed t-tests were conducted to determine whether group-level mean mAFC-d' scores were above zero (Table 4-8). Valence categorization performance was significantly above zero for all semantic conditions at each presentation speed (ps < .001), indicating that once the neutral bias was accounted for, participants could categorize valence significantly above chance. To determine whether significant differences in valence categorization performance exist across semantic categories a two-way ANOVA was conducted on valence categorization mAFC-d' scores using factors of presentation speed (17, 30 and 50 ms) and semantic category (animal, human, object, building) (Fig. 4-5). Main effects of both presentation speed $f(2,215)=53.8$, $p<.001$, partial $\eta^2=0.35$ and semantic category were significant $f(3,215)=2.94$, $p=.034$, partial $\eta^2=0.04$, but the interaction was not ($p>.1$). As the main effects of presentation speed and semantic category were both significant, one way ANOVAs collapsing across semantic category and presentation speed, respectively, were performed. The one-way ANOVA examining the effect of speed level (17, 33, or 50ms) on valence categorization mAFC-d' scores was significant $f(2,213)=52.6$, $p<.001$, partial $\eta^2=0.33$. Tukey's post-hoc tests revealed that differences in mAFC-d' scores between all speed levels were significant, where images presented at the shorter presentation speed showed smaller d' scores, meaning they were

categorized less well: 17ms vs 33ms M difference=-0.31, p<.001, 95% CI=(-0.44, -0.19), 17ms vs 50ms, M difference=-0.53, p<.001, 95% CI=(-0.65, -0.40), and 33ms vs 50ms, M difference=-0.21, p<.001, 95% CI=(-0.33, -0.09). The one-way ANOVA examining the effect of semantic category (animal, human, object, building) on valence categorization mAFC-d' scores was not significant however (p>.1).

Although no significant interaction between presentation speed and semantic category was detected in mAFC-d' scores for the valence rating task, further investigation using one-way ANOVAs at each speed level was conducted given our interest in ultra-rapid categorization of emotional stimuli. These one-way ANOVAs of semantic category at each presentation speed revealed that semantic category was only significant at the 33ms speed, f(3,215)=3.4, p<.023, $\eta^2$=0.13. Tukey's post-hoc test was then used on the one-way ANOVA of semantic category at the 33ms presentation speed to test for differences in mAFC=d' across levels of semantic category. At 33ms, semantic categorization performance (mAFC-d') of object images was significantly greater than that for building images (M difference=-0.2641), p=.047, 95% CI=(0.0022, 0.5259), and semantic categorization performance (mAFC-d') of human images was significantly greater than that for building images (M difference=0.2724), p=.038, 95% CI=(0.0106, 0.5342).

**Table 4-7: Valence Categorization Task mAFC-d' Descriptive Statistics.**

| **Experiment 1 mAFC-d'** | | **Speed** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **17ms** | | | **33ms** | | | **50ms** | | |
| | | M d' | SD | 95% CI | M d' | SD | 95% CI | M d' | SD | 95% CI |
| **Semantic** | **Animal** | 0.3 | 0.3 | 0.2 | 0.7 | 0.3 | 0.6 | 0.9 | 0.3 | 0.7 |
| | **Human** | 0.4 | 0.2 | 0.3 | 0.7 | 0.3 | 0.6 | 0.8 | 0.3 | 0.7 |
| | **Object** | 0.3 | 0.3 | 0.2 | 0.7 | 0.3 | 0.6 | 1 | 0.3 | 0.8 |
| | **Building** | 0.3 | 0.3 | 0.2 | 0.5 | 0.3 | 0.3 | 0.8 | 0.4 | 0.6 |
| | **TOTAL** | 0.3 | 0.3 | 0.3 | 0.6 | 0.3 | 0.6 | 0.9 | 0.3 | 0.8 |

| **Experiment 2 mAFC-d'** | | **Speed** | | | | | |
|---|---|---|---|---|---|---|---|
| | | **17ms** | | | **33ms** | | |
| | | M d' | SD | 95% CI | M d' | SD | 95% CI |
| **Semantic** | **Animal** | 0.5 | 0.2 | 0.4 | 0.9 | 0.3 | 0.8 |
| | **Human** | 0.5 | 0.3 | 0.4 | 0.8 | 0.2 | 0.8 |
| | **Object** | 0.5 | 0.4 | 0.4 | 1 | 0.3 | 0.9 |
| | **Building** | 0.5 | 0.9 | 0.3 | 0.7 | 0.4 | 0.6 |
| | **TOTAL** | 0.5 | 0.5 | 0.4 | 0.9 | 0.3 | 0.8 |

**Table 4-8: Valence Categorization Task mAFC-d' t-test Results.**

| **Experiment 1 mAFC-d'** | | **Speed** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **17ms** | | | **33ms** | | | **50ms** | | |
| | | t (df) | p | d | t (df) | p | d | t (df) | p | d |
| **Semantic** | **Animal** | 5.2 (17) | <.001 | 1.2 | 8.6 (17) | <.001 | 2.0 | 10.9 (17) | <.001 | 2.6 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Human** | 6.9 (17) | <.001 | 1.6 | 12.1 (17) | <.001 | 2.8 | 13.5 (17) | <.001 | 3.2 |
| | **Object** | 4.6 (17) | <.001 | 1.1 | 11.0 (17) | <.001 | 2.6 | 12.3 (17) | <.001 | 2.9 |
| | **Building** | 4.0 (17) | <.001 | 0.9 | 6.3 (17) | <.001 | 1.5 | 8.4 (17) | <.001 | 2.0 |
| | **TOTAL** | 10.2 (71) | <.001 | 1.2 | 17.6 (71) | <.001 | 2.1 | 21.8 (71) | <.001 | 2.6 |

| **Experiment 2 mAFC-d'** | | **Speed** | | | | | |
|---|---|---|---|---|---|---|---|
| | | **17ms** | | | **33ms** | | |
| | | t (df) | p | d | t (df) | p | d |
| **Semantic** | **Animal** | 10.8 (32) | <.001 | 1.9 | 19.4 (32) | <.001 | 3.4 |
| | **Human** | 9.4 (32) | <.001 | 1.6 | 23.2 (32) | <.001 | 4 |
| | **Object** | 8.2 (32) | <.001 | 1.4 | 18.4 (32) | <.001 | 3.2 |
| | **Building** | 3.4 (32) | <.001 | 0.6 | 11.5 (32) | <.001 | 2 |
| | **TOTAL** | 11.1 (131) | <.001 | 1 | 32.8 (131) | <.001 | 2.9 |

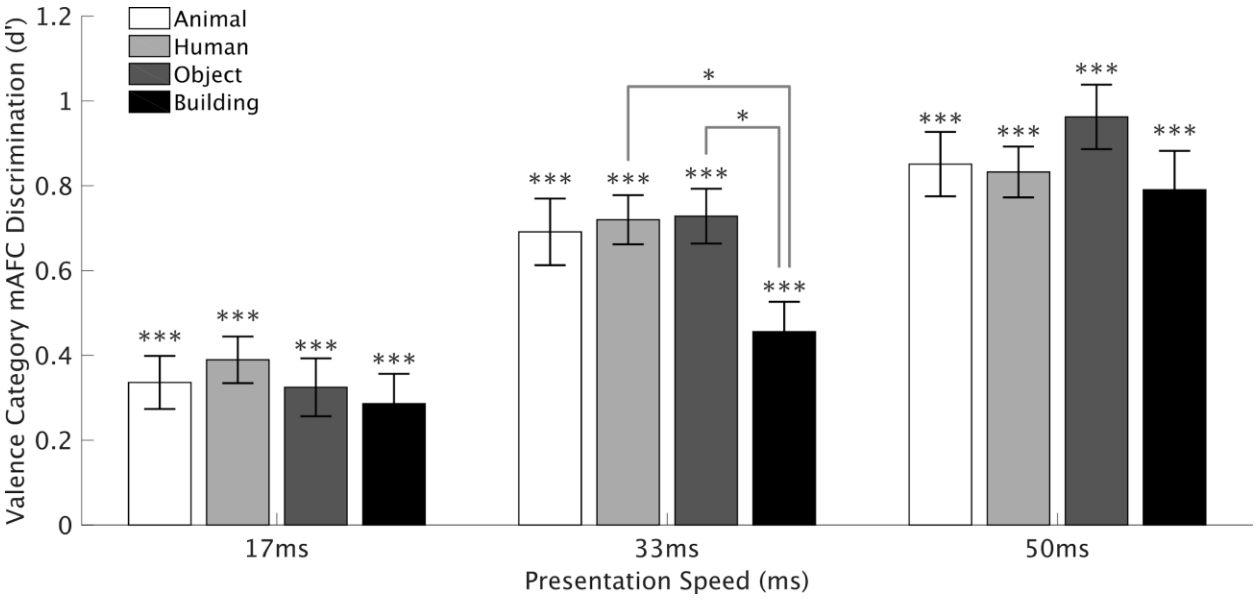


**Figure 4-5. Valence Categorization Task mAFC-d' Scores for Experiment 1.**
To account for the observed response bias in the valence categorization task from experiment 1 (shown in Fig. 4-4), multiple alternative forced choice (mAFC) d' scores were calculated which account for both false positive and hit rates. A score of 0 indicates chance performance. Results are broken down by presentation speed and semantic category (animal, human, object, building). Error bars show standard error (S.E) across subjects, stars above each bar indicate which presentation speed by semantic conditions were classified with above chance performance, while lines between bars indicate which presentation speed by valence conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Unlike the percent-correct scores shown in Figure 4-4, once response bias was accounted for, performance was significantly above chance at all presentation speed by semantic conditions. At 33ms, performance for valence categorization of building images was significantly worse than that of human and object categories (ps<.05).

**Cognitive Primacy Hypothesis Vs. Affective Primacy Hypothesis.**

To find evidence for either the semantic or affective primacy hypothesis, four conditional probabilities across the semantic and valence tasks were calculated: the probability of getting the semantic task correct on trials where the valence task was correctly answered, denoted as the probability of semantic task correct (S) given valence task correct (V) (P(S|V), the probability of semantic task correct given valence task incorrect (~V) P(S|~V), the probability of valence task correct given semantic task correct P(V|S), and the probability of valence task correct given semantic task incorrect (~S), P(V|~S)). T-tests were then used to determine whether probabilites where significantly above chance at the group level (see the Data Analysis part of the Methods section for more information). These conditional probabilities were calculated for all speed by semantic by valence conditions (Figs. 4-6, 4-S3, & 4-S4). Only those conditions with above chance performance in both the semantic and valence tasks were considered. Additionally, all neutral valence conditions were excluded from the analysis due to the neutral response bias in the valence categorization task. We note that, although conditions with the object semantic level also might be influenced by response bias, they were kept in because we expected to see evidence for the cognitive primacy hypothesis, and the object semantic response bias only made that more unlikely. This is because semantic categorization accuracy will likely be higher than chance for objects due to the selection bias, even though the subject did not accurately perceive the category correctly. If classification of semantic category is indeed necessary for valence categorization, as the cognitive primacy hypothesis postulates, then those stimuli correctly labelled as objects simply due to the response bias should not also get the valence categorization task correct, and would then count as evidence against the cognitive primacy hypothesis as we quantify it.

At 17ms, none of the conditions had both significant semantic and valence %-correct scores (ps > .1 for valence %-correct). At 33ms, conditional probabilities for positive humans were P(S|V)=0.97, $t(17)=83.6$, $p<.001$, P(S|~V)=0.87, $t(17)=26.0$, $p<.001$, P(V|S)=0.64, $t(17)=7.2$, $p<.001$, P(V|~S)=0.20, $t(17)=-2.4$, $p>.1$, showing support for the cognitive as opposed to the emotion, primacy hypothesis (Fig. 4-6). At 50ms positive humans also showed support for the cognitive, versus the emotion, primacy hypothesis (Fig. 4-6), with conditional probabilities P(S|V)=0.97, $t(17)=82.9$, $p<.001$, P(S|~V)=0.88, $t(17)=20.5$, $p<.001$, P(V|S)=0.65, $t(17)=7.1$, $p<.001$, P(V|~S)=0.26, $t(17)=-1.0$, $p>.1$. Additionally, at 50 ms responses to images of negative animals, humans and objects provided evidence for the cognitive, but not emotion, primacy hypothesis (Fig. 4-6). Conditional probabilities for negative animals were P(S|V)=0.91, $t(17)=27.3$, $p<.001$, P(S|~V)=0.82, $t(17)=13.9$, $p<.001$, P(V|S)=0.52, $t(17)=6.9$, $p<.001$, P(V|~S)=0.45, $t(17)=1.2$, $p>.1$, for negative humans they were P(S|V)=0.90, $t(17)=34.4$, $p<.001$, P(S|~V)=0.8, $t(17)=17.3$, $p<.001$, P(V|S)=0.58, $t(17)=7.4$, $p<.001$, P(V|~S)=0.40, $t(17)=0.9$, $p>.1$, and for negative objects they were P(S|V)=0.88, $t(17)=29.2$, $p<.001$, P(S|~V)=0.81, $t(17)=21.9$, $p<.001$, P(V|S)=0.61, $t(17)=6.7$, $p<.001$, P(V|~S)=0.45, $t(17)=1.7$, $p>.1$. Importantly, in none of the presentation speed by valence by arousal conditions did the patterns of conditional probabilities provide support for the affective primacy hypothesis.

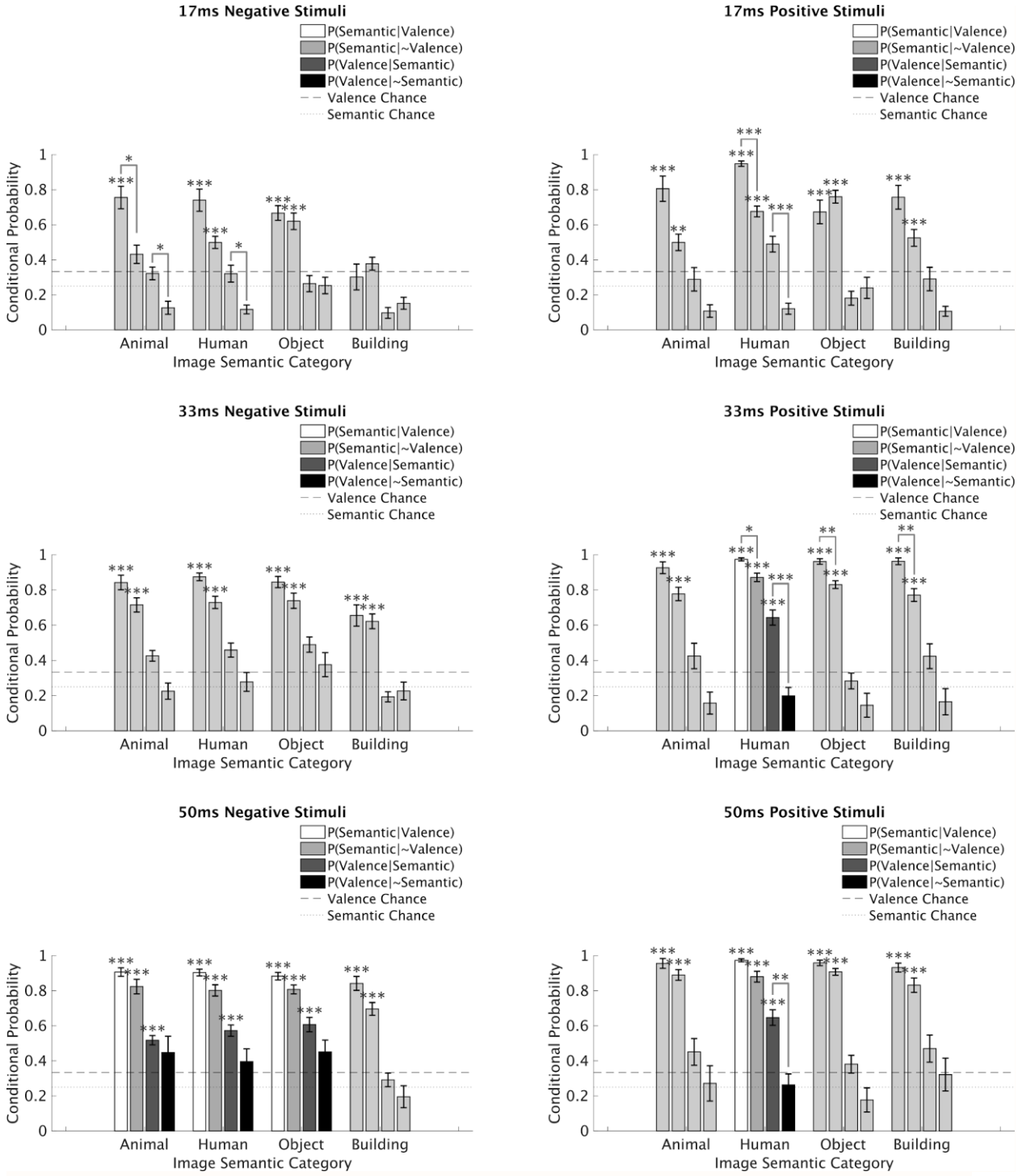


**Figure 4-6. Task Performance Conditional Probabilities Showing Evidence of Cognitive Primacy Hypothesis for Experiment 1.**

To quantify the co-occurrence of successful categorization of valence and semantic category in experiment 1, conditional probabilities for each task were calculated, broken down by presentation speed by valence by semantic category conditions (see methods for details). Due to response bias in the valence categorization of neutral images, they were excluded from consideration (see methods for reasoning). As the cognitive primacy hypothesis predicts valence information cannot be extracted before semantic information, we looked for supporting evidence of the cognitive primacy hypothesis under the following circumstances: (i) above chance conditional probability of getting

the valence correct when getting the semantic category correct (P(V|S) above chance), (ii) conditional probability of getting the valence correct when not getting the semantic category correct (P(V|~S not above chance), (iii) overall (marginal) probability of getting both the semantic and valence categories correct was above chance (P(S) and P(V) above chance). Bars were colored for those experimental conditions where these constraints were met showing evidence of the cognitive primacy hypothesis, and greyed out for those experimental conditions were these constraints were not met. None of the experimental conditions showed evidence for the affective primacy hypothesis (see methods for description of constraints). Note that the positive human experimental condition at 17ms is greyed out because the marginal probability of correct valence categorization (P(V)) was not significant, even though the P(S|V) was significant and the P(S|~V) was not. Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (25%), and stars above each bar indicate which presentation speed by semantic by valence conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

### Discussion

The results from experiment 1 showed that, when controlling for response bias using mAFC-d' scores, humans can accurately categorize both the semantic and valence content of naturalistic emotional images that are presented as briefly as 17ms and followed by backwards masking. Additionally, categorization of semantic content is influenced by the emotional valence of the images. Negative images were categorized significantly worse than were both neutral and positive images, findings that are inline with findings from the attentional narrowing literature that stimulus negative valence can impede cognitive task performance, such as the Attention Network Test (Cohen, Henik, & Mor, 2011) and anti-saccade tasks (van Steenbergen, Band, & Hommel, 2011). In contrast, valence categorization performance was not influenced by the semantic category of the images. Taken together, these findings contradict predictions made by the biological preparedness hypothesis, which would predict improved categorization performance for negative, animate stimuli. Finally, when both the semantic and valence categorization tasks were performed with above chance accuracy, the conditional probability of getting the valence categorization task correct was only above chance when the semantic task was also done correctly, whereas the reverse did not hold true. This is consistent with the proposal that semantic information must first be extracted before affective information can be determined, which supports the cognitive primacy hypothesis as opposed to the affective primacy hypothesis.

## Experiment 2

In order to replicate our findings and address several limitations of experiment 1, a second experiment was conducted. First, the valence ratings used as ground truth for the valence categorization task in experiment 1 had been determined by the majority rating of three raters. We were thus concerned that performance on the valence judgment in Experiment 1 might have been underestimated as a result of disagreement between our in-house raters and participants. To better ensure that each image used was consistently allocated to a given valence category, all images used in experiment 2 were independently rated by 18 subjects on Amazon's mechanical turk website, in addition to the original three in-house raters. Only those images where the modal response of mechanical turk workers agreed with all three in-house raters were used for experiment 2. Second, a number of the object category images from Experiment 1 depicted the object against a plain white background, which led to the concern that they may have been easier to identify. This might alter their identifiability relative to images from other semantic categories

given that briefly presented images of scenes located against inconsistent or uncommon background colors are more difficult to identify (Boyce, Pollatsek, and Rayner, 1989; Goffaux, et al., 2005). Thus, all images of objects with a white background were replaced with images of objects in their natural environments. Third, in order to control for any confounds of task order, trials in experiment 2 were counterbalanced so that the semantic categorization task came first on half of the trials and the valence categorization task came first on the other half of trials, per subject.

## Method

### Participants.

Forty-three volunteers (22 females, mean age = 19.43, age range = 18-25) from the University of California, Berkeley subject population participated in Experiment 2. Participants were compensated $20/1.5 hours. 10 subjects were excluded for not completing the full set of sessions, for expressing that they misunderstood the task after having completed the experiment, if they finished the experiment extremely quickly or slowly, or if they were found not paying attention to the task. Written, informed consent was obtained for all participants before conducting the experiment. The experimental protocol for this study was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley.

### Materials.

720 images were selected across the same four semantic categories (animal, human, object, and building) and valence categories (negative, neutral, and positive) as were used in Experiment 1. Only 720 images were required here, as opposed to the 1080 used in experiment 1, as only two of the three original presentation times were used (17ms and 33ms). The majority of images were a subset of those used in experiment 1. Ratings of semantics, valence and arousal for all images used in Experiment 1 without a white background were rated by 18 participants on Amazon Mechanical Turk. Only those images where the modal response of the Mechanical Turk participants agreed with the in-house raters' categorization, on both semantic and valence categories, were retained for use in experiment 2. Where necessary to keep an even balance of images across semantic by valence conditions, new replacement images were obtained from the internet. These images were rated by both the in-house raters and 18 Amazon Mechanical Turk participants, and the same criteria was used to determine eligibility for inclusion in the final stimulus set. Eight randomizations of the stimulus set were created such that each image was seen at either the 17ms or 33ms presentation speed, had either the semantic or valence categorization task come first, and was presented in one of two orders. Equal numbers of both males and females were assigned to each randomization.

### Procedure.

The experimental procedure for Experiment 2 was identical to that of Experiment 1,with two exceptions. First, only two stimulus presentation speeds were used, namely 17ms and 33ms. The 50ms presentation speed was excluded because subjects were at near ceiling performance on the semantic categorization task in Experiment 1. Thus, the stimulus set was broken down into 24 blocks of 30 stimuli each. Twelve blocks were shown in the first session, and twelve in the second session. Second, to control for possible task order effects, task order was counterbalanced

such that within each block equal numbers of trials had the semantic task first and the valence task first.

### Data Analysis.

Analysis of data from Experiment 2 followed in exactly the same manner as was done in Experiment 1. One additional analysis was done to determine whether task order had any affect on either the semantic or valence categorization task performance. A 4-way ANOVA was conducted on both the semantic and emotion % correct scores, with factors for presentation speed, semantic category, valence category and task order. An additional 3-way ANOVA was done on the mAFC-d' scores from the semantic categorization task with factors for presentation speed, valence category and task order. Finally, a 3-way ANOVA was done on the mAFC-d' scores from the valence categorization task with factors for presentation speed, semantic category and task order.

## Results

### Semantic Categorization Task.

Descriptive statistics of group-level %-correct scores for the semantic categorization task from Experiment 2 are shown below those for Experiment 1 in Table 4-1. Scores are shown in the same manner as for Experiment 1. Results of one-tailed t-tests indicated that semantic categorization was significantly above chance ($ps < .01$) in all conditions and aggregate conditions after Bonferroni correction for multiple comparisons (Table 4-2 and Fig. 4-7A-B). As with Experiment 1, near ceiling performance was seen for many categories at the 33ms presentation times. The response bias to select the object semantic category when guessing (Fig. 4-7C) was also seen here in experiment 2, as it was in experiment 1 (as indicated by disproportionate selection of the object category when incorrect). Thus, mAFC-d' scores were again calculated in the same manner as in Experiment 1 to account for this bias.

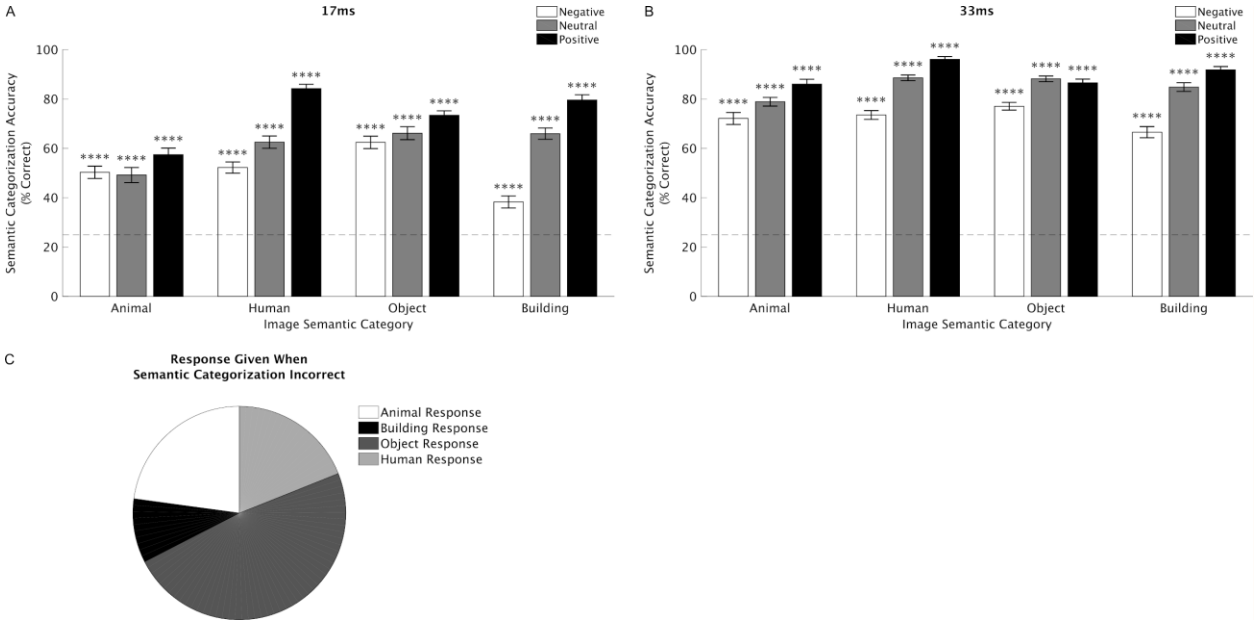


**Figure 4-7. Semantic Categorization Task Performance for Experiment 2.**
(A-B) Bar plots showing group-averaged percentage of semantic categorization task trials classified correctly in experiment 2. Each plot shows results broken down by semantic and valence categories from trials at one of the 2 presentation speeds (17ms in A, and 33ms in B). Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (25%), and stars above each bar indicate which presentation speed by semantic by valence conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Group level task performance at all semantic by valence conditions was significantly above chance at all three presentation speeds. (C) Pie chart showing proportion of incorrectly classified semantic categorization task responses for each semantic category, across all subjects. As objects were chosen for nearly 50% of incorrect responses this figure shows a clear bias to select object when the semantic category was unknown.

Descriptive statistics of group-level mAFC-d' scores are shown in Table 4-3 (see also Fig. 4-8). Reproducing our findings from experiment 1, semantic task performance as quantified by mAFC-d' scores was found to be significantly above chance for all speed by valence conditions, as well as aggregate valence conditions, (ps < .001, Table 4-4). A two-way ANOVA was conducted on the mAFC-d' scores using factors of presentation speed and valence to test for differences in semantic categorization task performance. Both main effects of presentation speed $f(2,197)=54.5$, $p<.001$, partial $\eta^2=0.17$, and valence $f(2,197)=43.1$, $p<.001$, partial $\eta^2=0.14$ were significant, but there was no significant interaction ($p>.1$). As the both the main effects of presentation speed and valence were significant, as they were in experiment 1, one way ANOVAs were conducted by collapsing across levels of speed and valence. The one-way ANOVA on speed level (17, 33) was significant $f(1,196)=34.1$, $p<.001$, partial $\eta2=0.15$, revealing that semantic task performance is significantly better at presentation speeds of 33ms than at presentation speeds at17ms. The one-way ANOVA of semantic task mAFC-d' scores on valence category (negative, neutral, positive) was also significant $f(1,195)=12.9$, $p<.001$, partial $\eta2=0.12$. Tukey's post-hoc test revealed that, in addition to a significantly worse semantic task performance for negative versus positive stimuli as was seen in experiment 1, M difference=-1.14, $p<.05$, 95% CI=(-1.66, -0.61), semantic task mAFC-d' scores for neutral images were significantly smaller than for positive images, M difference=-0.66, $p<.01$, 95% CI=(-1.19, -0.13). The difference in semantic task mAFC-d' scores between negative and neutral images was not

significant however(p=.09), with negative stimuli trending towards worse performance than neutral stimuli.

As in experiment 1, although no significant interaction was detected, one-way ANOVAs were conducted at both presentation speeds because this study is investigating ultra-fast perception. Theses one-way ANOVAs revealed significant differences across levels of valence in semantic task mAFC-d' scores; at 17ms $f(2,53)=55.5$, $p<.001$, $\eta^2=0.54$, and at 33ms $f(2,53)=50.6$, $p<.001$, $\eta^2=0.51$. Tukey's post-hoc test was then used to find which levels of valence differed significantly at each presentation speed. For 17ms, as seen in Experiment 1, semantic categorization was significantly poorer both for neutral images than for negative images, (M difference=-0.31, p<.001, 95% CI=(-0.50,-0.13)) and for positive images than for negative images (M difference=-0.80), p<.001, 95% CI=(-0.98,-0.62). Additionally, in Experiment 2 semantic categorization performance was significantly better for positive images than for neutral images (M difference=-0.48), p<.001, 95% CI=(-0.67, -0.30)). Likewise at 33ms, semantic categorization performance of negative images was significantly greater than for neutral images (M difference=-0.64, p<.001, 95% CI=(-0.89,-0.38)), and for positive images (M difference=-1.07, p=.001, 95% CI=(-1.33,-0.82)). Additionally, at 33ms semantic task mAFC-d' scores were greater for positive images than they were for neutral images (M difference=-0.44, p=.001, 95% CI=(-0.69,-0.18)).

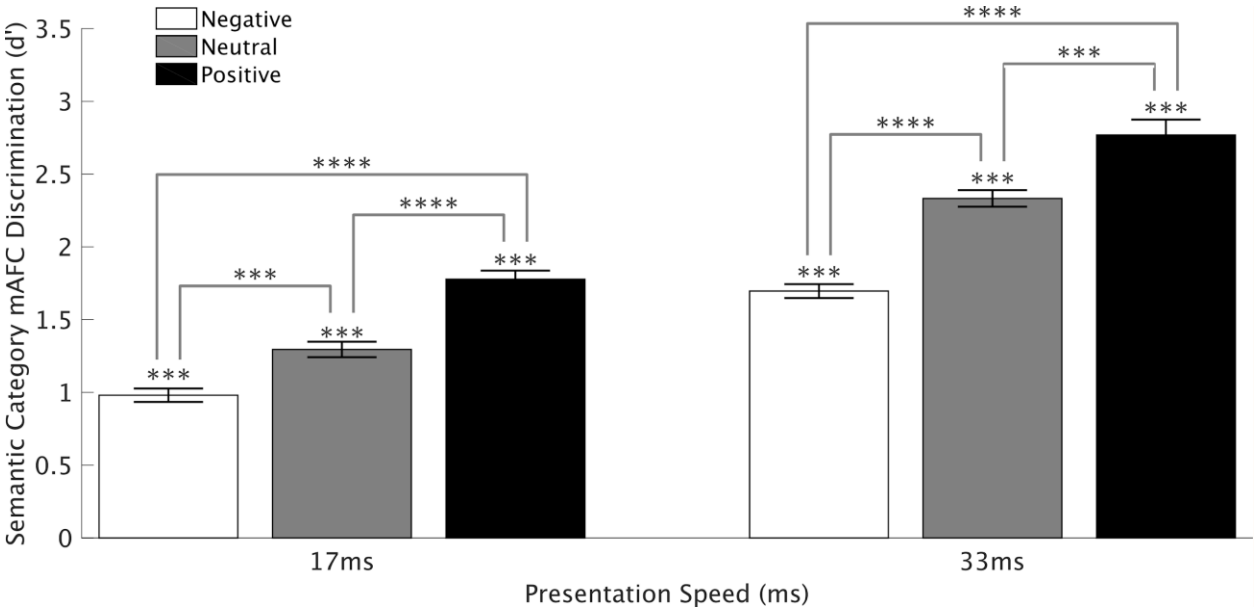


**Figure 4-8. Semantic Categorization task mAFC-d' Scores for Experiment 2.**
To account for the observed response bias in the semantic categorization task from experiment 2 (shown in Fig. 4-9) multiple alternative forced choice (mAFC) d' scores were calculated which account for both false positive and hit rates. A score of 0 indicates chance performance. Results are broken down by presentation speed and valence category (negative, neutral, positive). Error bars show standard error (S.E) across subjects, stars above each bar indicate which presentation speed by valence conditions were classified with above chance performance, while lines between bars indicate which presentation speed by valence conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Performance was significantly above chance at all presentation speed by valence conditions. Critically, negative stimuli were less accurately classified than both neutral and positive stimuli were at both presentation speeds, and positive stimuli were more accurately classified than both negative and neutral stimuli were at both presentation speeds.

**Valence Categorization Task.**

Descriptive statistics of the %-correct scores for the valence categorization task in Experiment 2 are shown below those for Experiment 4-1 in Table 4-5 (and Table 4-6 and Fig. 4-9). At 17ms (Fig. 4-9A), one-tailed t-tests revealed above chance valence categorization accuracy for negative animals M=42.2%, t(32)=3.2, p<.05, d=0.6, and positive humans M=47.4%, t(32)=5.2, p<.001, d=0.9. Additionally, as in Experiment 1, all neutral categories were significantly identified above chance (ps < .001). At 33ms (Fig. 9B), valence categorization of positive human images was again significantly above chance, M=71.4%, t(32)=16.5, p<.001, d=2.9, as was also true for images of positive animals M=51.4%, t(32)=6.5, p<.001, d=1.1, and images of positive objects M=56.4%, t(32)=8.5, p<.001, d=1.5. Additionally valence categorization of three negative image categories was significantly above chance, namely for images of negative animals M=62.9%, t(32)=9.2, p<.001, d=1.6, images of negative humans M=52.0%, t(32)=7.4, p<.001, d=1.3, and images of negative objects M=49.7%, t(32)=6.2, p<.001, d=1.1. As seen in experiment 1, valence categorization of all neutral image categories was significantly above chance (ps<.01). Also as seen in experiment 1, a response bias in the valence categorization task to select neutral when guessing was seen in Experiment 2, with approximately 57% of incorrect trials categorized as neutral (Fig. 4-9C).

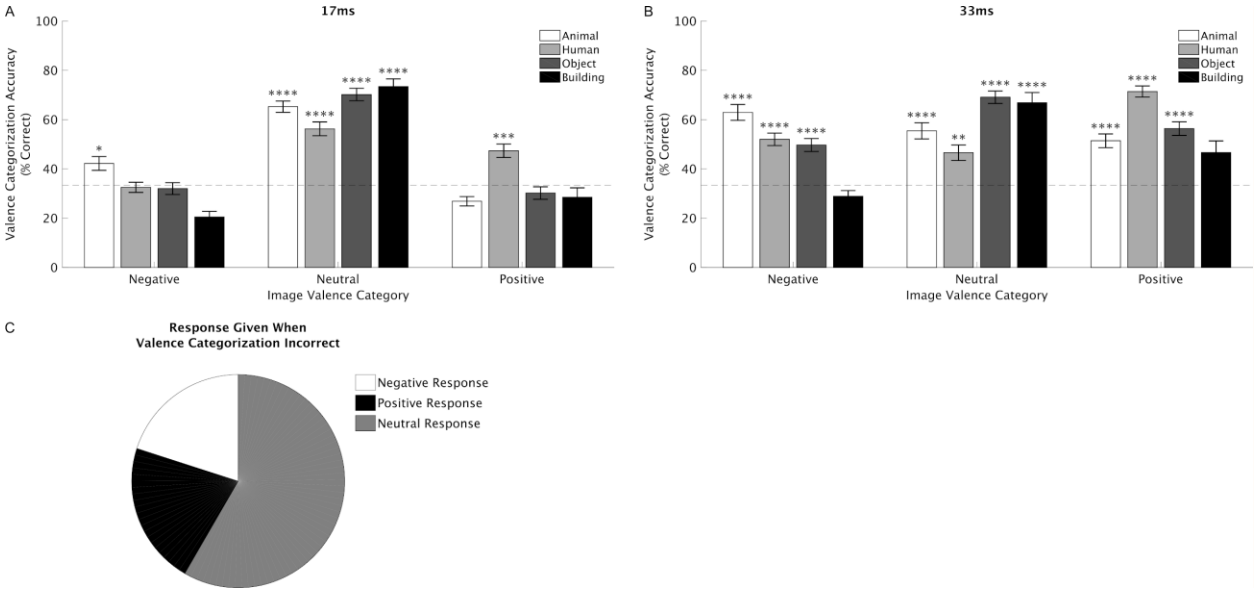


**Figure 4-9. Valence Categorization Task Performance for Experiment 2.**
(A-B) Bar plots showing group-averaged percentage of valence categorization task trials classified correctly in experiment 2. Each plot shows results broken down by valence and semantic categories from trials at one of the 2 presentation speeds (17ms in A, and 33ms in B). Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (33%), and stars above each bar indicate which presentation speed by valence by semantic conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). At 17ms, images of positive humans were classified with above chance performance, while at 33ms positive and negative images of animals, humans and objects were all classified with above chance performance. (C) Pie chart showing proportion of incorrectly classified valence categorization task responses for each valence category, across all subjects. As neutral were clearly chosen for more than 50% of incorrect responses this figure shows a clear bias to select neutral when the valence category was unknown.

To account for the observed neutral response bias in the valence categorization task of experiment 2, mAFC-d' scores were again calculated. Table 4-7 contains descriptive statistics for the mAFC-d' scores. mAFC-d' scores for the valence categorization task were significantly above zero for all speed by semantic conditions (ps < .001; Table 4-8 and Fig. 4-10). A two-way ANOVA with speed and semantic factors and dependent variable valence categorization mAFC-d' scores revealed that the main effect of speed was significant $f(1,263)=28.5$, $p<.001$, partial $\eta^2=0.10$ with valence categorization performance significantly worse at presentation speeds of 17ms than at presentation speeds of 33ms. The main effect of semantic category was not significant however, nor was the interaction (ps>.1).

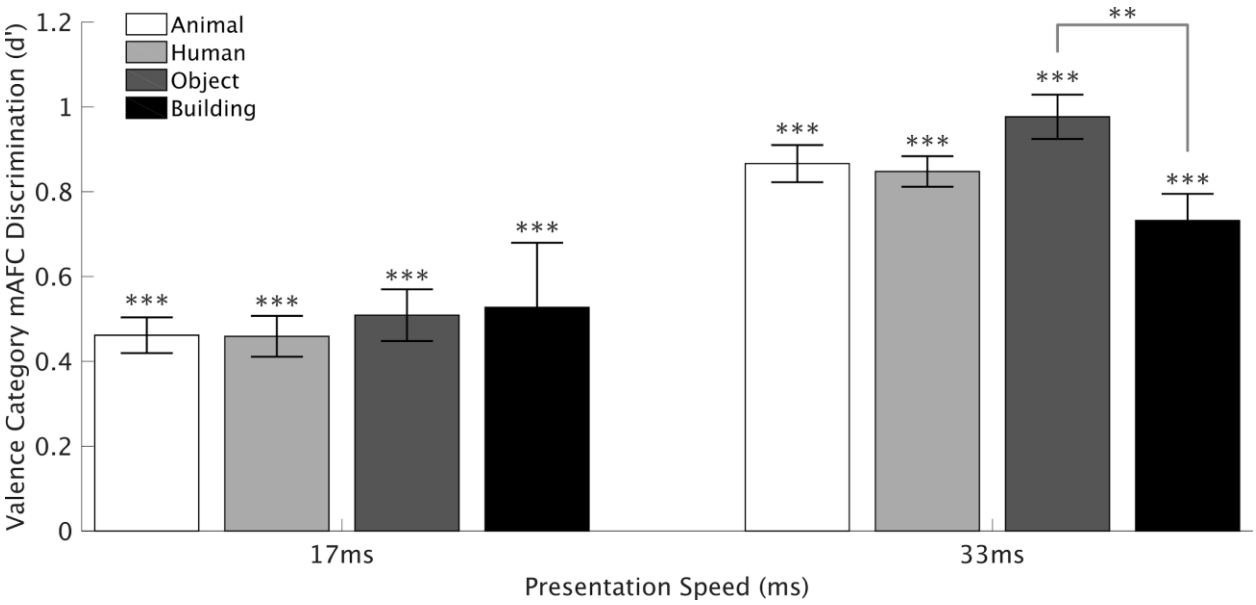


**Figure 4-10. Valence Categorization Task mAFC-d' Scores for Experiment 2.**
To account for the observed response bias in the valence categorization task from experiment 1 (shown in Fig. 4-9), multiple alternative forced choice (mAFC) d' scores were calculated which account for both false positive and hit rates. A score of 0 indicates chance performance. Results are broken down by presentation speed and semantic category (animal, human, object, building). Error bars show standard error (S.E) across subjects, stars above each bar indicate which presentation speed by semantic conditions were classified with above chance performance, while lines between bars indicate which presentation speed by valence conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001). Unlike the percent-correct scores shown in Figure 4-9, once response bias was accounted for, performance was significantly above chance at all presentation speed by semantic conditions. At 33ms, performance for valence categorization of building images was significantly worse than that of object categories (p<.05).

### Effects of Task Order.

To test for possible effects of task order, a 4-way ANOVA was conducted on both the semantic and valence %-correct scores, using factors of presentation speed (17 & 33ms), semantic category (human, animal, object, building), valence (negative, neutral, positive) and task order (semantic task first, valence task first). Main effects of task order was not significant in either analysis (ps>.1), nor were the two-way interactions between task order and semantic, task order and valence, or task order and presentation speed (ps>.1). A 3-way ANOVA using

factors of presentation speed (17 & 33ms), valence (negative, neutral, positive) and task order (semantic task first, valence task first) was conducted on the mAFC-d' semantic scores and neither the main effect of, nor two-way interactions with, task order were not significant there ($p>.1$). A second 3-way ANOVA using factors of presentation speed (17 & 33ms), semantic category (human, animal, object, building) and task order (semantic task first, valence task first) was conducted on the mAFC-d' valence scores and neither the main effect of, nor two-way interactions with, task order was not significant there either ($p>.1$).

**Cognitive Primacy Hypothesis Vs. Affective Primacy Hypothesis.**

In the same manner as in Experiment 1, conditional probabilities were calculated for all speed by semantic by valence conditions (Fig. 4-11, 4-S5 & 4-S6). All conditions showing support for the cognitive as opposed to the affective primacy hypothesis in Experiment 1 also showed support for the cognitive as opposed to the affective primacy hypothesis in Experiment 2. This was now apparent at shorter presentation times. At 17ms, conditional probabilities for positive human images were $P(S|V)=0.97$, $t(33)=53.1$, $p<.001$, $P(S|\sim V)=0.73$, $t(33)=18.1$, $p<.001$, $P(V|S)=0.54$, $t(33)=7.0$, $p<.001$, $P(V|\sim S)=0.09$, $t(33)=-6.6$, $p>.1$, showing support for the cognitive as opposed to the affective primacy hypothesis (Fig. 4-11). At 33ms, positive human images also showed support for the cognitive as opposed to the affective primacy hypothesis (Fig. 4-11), with conditional probabilities being $P(S|V)=0.99$, $t(33)=103.3$, $p<.001$, $P(S|\sim V)=0.88$, $t(33)=23.5$, $p<.001$, $P(V|S)=0.74$, $t(33)=17.1$, $p<.001$, $P(V|\sim S)=0.11$, $t(33)=-4.8$, $p>.1$. At 33ms, images of negative animals, humans and objects also showed evidence for the cognitive as opposed to the affective primacy hypothesis (Fig. 4-11). Conditional probabilities for images of negative animals were $P(S|V)=0.80$, $t(33)=24.2$, $p<.001$, $P(S|\sim V)=0.61$, $t(33)=9.4$, $p<.001$, $P(V|S)=0.70$, $t(33)=11.8$, $p<.001$, $P(V|\sim S)=0.47$, $t(33)=2.9$, $p>.1$, for images of negative humans they were $P(S|V)=0.81$, $t(33)=29.7$, $p<.001$, $P(S|\sim V)=0.66$, $t(33)=15.6$, $p<.001$, $P(V|S)=0.57$, $t(33)=8.9$, $p<.001$, $P(V|\sim S)=0.40$, $t(33)=1.9$, $p>.1$, and for images of negative objects they were $P(S|V)=0.83$, $t(33)=32.4$, $p<.001$, $P(S|\sim V)=0.71$, $t(33)=17.2$, $p<.001$, $P(V|S)=0.54$, $t(33)=6.9$, $p<.001$, $P(V|\sim S)=0.39$, $t(33)=1.5$, $p>.1$. Additionally, at 33ms images of positive animals and positive objects also showed support for the cognitive as opposed to the affective primacy hypothesis (Fig. 4-11). The conditional probabilities were: positive animals $P(S|V)=0.92$, $t(33)=27.3$, $p<.001$, $P(S|\sim V)=0.77$, $t(33)=16.9$, $p<.001$, $P(V|S)=0.56$, $t(33)=6.8$, $p<.001$, $P(V|\sim S)=0.23$, $t(33)=-2.0$, $p>.1$, and positive objects $P(S|V)=0.92$, $t(33)=49.7$, $p<.001$, $P(S|\sim V)=0.80$, $t(33)=20.0$, $p<.001$, $P(V|S)=0.60$, $t(33)=9.5$, $p<.001$, $P(V|\sim S)=0.35$, $t(33)=0.3$, $p>.1$. Replicating our findings from experiment 1, in none of the presentation speed by valence by arousal conditions did the patterns of conditional probabilities provide support for the affective primacy hypothesis.

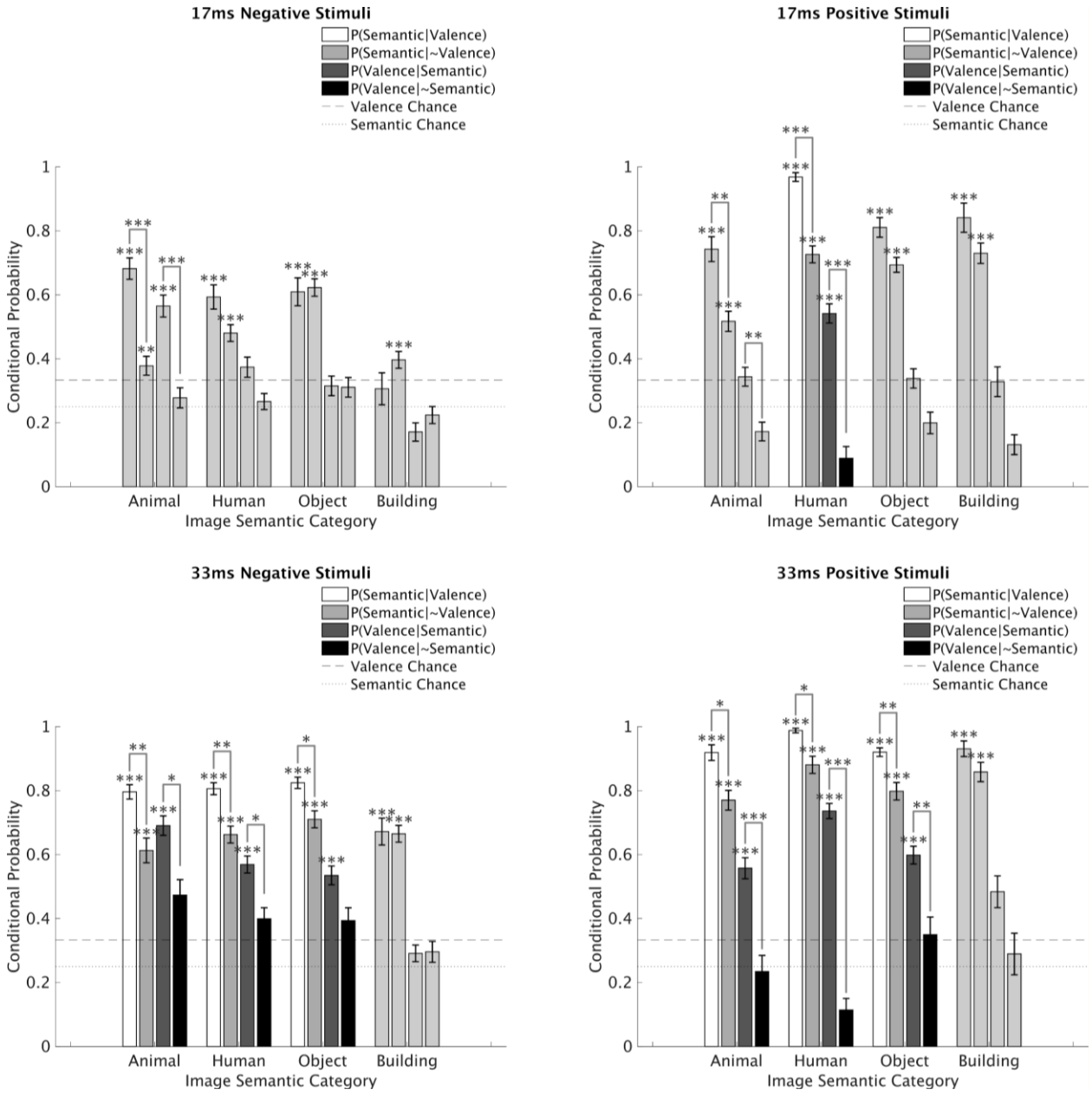


**Figure 4-11. Task Performance Conditional Probabilities Showing Evidence of Cognitive Primacy Hypothesis for Experiment 2.**

To quantify the co-occurrence of successful categorization of valence and semantic category in experiment 2, conditional probabilities for each task were calculated, broken down by presentation speed by valence by semantic category conditions (see methods for details). Due to response bias in the valence categorization of neutral images, they were excluded from consideration (see methods for reasoning). As the cognitive primacy hypothesis predicts valence information cannot be extracted before semantic information, we looked for supporting evidence of the cognitive primacy hypothesis under the following circumstances: (i) above chance conditional probability of getting the valence correct when getting the semantic category correct (P(V|S) above chance), (ii) conditional probability of getting the valence correct when not getting the semantic category correct (P(V|~S not above chance), (iii) overall (marginal) probability of getting both the semantic and valence categories correct was above chance (P(S) and P(V) above chance). Bars were colored for those experimental conditions where these constraints were met showing evidence of the cognitive primacy hypothesis, and greyed out for those experimental conditions were these

constraints were not met. None of the experimental conditions showed evidence for the affective primacy hypothesis (see methods for description of constraints). Error bars show standard error (S.E) across subjects, dotted lines indicate chance performance (25%), and stars above each bar indicate which presentation speed by semantic by valence conditions were classified with above chance performance (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

### Discussion

Experiment 2 addressed the same questions as did experiment 1, with the aim of reproducing the results while controlling for several limitations. Using mAFC-d' scores to control for response bias, revealed that both semantic and valence categorization can be performed above chance at 17ms, replicating the finding from experiment 1. The decrease in semantic categorization performance for negative, relative to neutral and positive, images was also replicated. We now also observed better semantic categorization performance for positive, relative to negative and neutral, images. This effect was not observed in experiment 1.

Turning to valence categorization task performance, this was not significantly influenced by the semantic category of the image, also replicating findings from experiment 1. Semantic and valence categorization task order was counterbalanced in experiment 2 to control for effects of task order. Analysis of variance concluded that neither task order had no significant influence upon task performance for either the semantic or valence categorization tasks, controlling for factors of presentation speed, valence and semantic category and their two-way interactions with task order. Finally, the conditional probability of accurately categorizing valence was only above chance when the semantic category was also accurately chosen for those conditions where both semantic and valence categorization tasks were done at above chance level. This again provided support for the cognitive primacy hypothesis as opposed to the affective primacy hypothesis, replicating the findings from experiment 1 here as well. Interestingly, evidence for the cognitive primacy hypothesis was observed for the same valence by semantic conditions as in experiment 1, but at a faster presentation time. One possible explanation for this increase in performance is the higher quality stimuli that were used in experiment 2. Only stimuli from experiment 1 that had a modal rating from 18 Amazon Mechanical Turk works that agreed with 3 in-house raters were used for this study, and thus images whose valence category was somewhat ambiguous were excluded from this study.

## General Discussion

In the two experiments reported here, we used rapid presentation of naturalistic emotional images to explore humans' ability to categorize the semantic category and valence of natural images at presentation times that likely curtail processing to be predominantly feed-forward. By using a stimulus set more diverse in semantic categories than previous studies, conducting multiple alternative forced choice (mAFC) tasks upon both image semantic category and valence (negative, positive and neutral), and controlling for response bias by using mAFC-d' scores, this study was able to address more questions than previous studies using a similar paradigm. In both experiments conducted, we found that semantic categorization of images presented for just 17ms could be done with better than chance performance, in line with previous findings (Greene & Oliva, 2009; Grill-Spector & Kanwisher, 2005; Nummenmaa, Hyona, & Calvo, 2010). The biological preparedness hypothesis would predict that negative stimuli might require shorter presentation durations for correct categorization than neutral stimuli (Ohman & Mineka, 2001).

Some authors might contend that this should also be true for positive stimuli (Tamietto & de Gelder, 2010; Zajonc, 1980). On the other hand, evidence from the attentional narrowing literature has shown performance costs incurred for tasks completed after viewing negative visual stimuli (Cohen, Henik, & Mor, 2011; van Steenberge, Band, & Hommel, 2011), and thus these findings might predict impaired categorization performance for negative stimuli relative and neutral and positive stimuli. In experiment 1 we found that semantic categorization was poorer for negative images relative to neutral and positive images, at all three presentation speeds. This finding was replicated in experiment 2, and additionally positive stimuli were found to be categorized significantly better than negative and neutral stimuli across both presentation speeds. Thus, these findings show that although increased performance is present for some emotional stimuli (namely positive), there is a differential effect by valence. This differential effect does not support the biological preparedness hypothesis, but rather is in line with findings from the attentional narrowing literature.

While narrowing of attention to the negative stimulus emotional content, which thus draws attention away from the semantic content of the image, could perhaps explain our semantic categorization findings, a problem for the attentional narrowing theory is our finding that valence categorization is poor, even for negative stimuli, at short presentation times. Another interesting hypothesis that might, better,explain our findings is the concept of "cognitive paralysis" (Leach, 2004, 2005). According to this hypothesis, the supervisory attentional system (SAS) becomes disabled under threat. Using the example of disasters in which many victims often perish because of a lack of adaptive behavior, Leach argues that "cognitive paralysis" occurs because the main role of the SAS is to allow survival responses to be prepared prior to facing threat, not when directly faced with a threat. If executive mechanisms are temporarily disabled when viewing negative stimuli, it could take longer to perform any explicit task regarding negative images, as we observed. If "cognitive paralysis" does indeed explain our findings, and the cognitive primacy hypothesis is true, this explanation also raises another question, namely: if semantic information is extracted before valence information, and negative stimuli cause cognitive paralysis in a semantic categorization task, how does the brain infer an image is negative without also knowing it's semantic category resulting in high semantic categorization performance for negative images (which we didn't see)? One possible explanation is that both semantic and valence information is successfully extracted during brief presentations of emotional images, which leads to temporary disabling of executive mechanisms necessary to do the semantic categorization task which do not come back online until after the visual stimulus representations in OTC are extinguished due to backwards masking.

Nummenmaa, Hyona, & Calvo (2010) also found evidence that humans can quickly categorization briefly presented emotional images, and that this categorization was faster for semantic categorization than it was for valence categorization. However, in that study only negative and positive images were used, and thus conclusions about the relative performance on negative, neutral and positive images could not be made. Furthermore, Nummenmaa and colleagues (2010) also used images of only animals and humans, and thus could not generalize their findings to inanimate stimuli. Importantly, they also restricted their analyses to percent change scores and did not calculate d-primes. Based on percent-correct scores, they concluded that categorization of pleasant vs. unpleasant images presented for 20ms and backwards masked could not be done with above chance performance. Findings from experiment 1 showed that when using percent-correct scores, valence was categorized at above chance levels for only neutral images, which was likely due simply to the large response bias to select neutral when the

valence category was unknown. However, controlling for response bias through the use of mAFC-d' scores revealed that even at 17ms, images of all semantic categories were identified at above chance performance. Similar findings emerged from experiment 2, although at 17ms even without controlling for the response bias, images of positive humans were significantly categorized with above chance performance. This relative increase in categorization performance may have been the result of the valence of the images used for experiment 2 being less ambiguous after more raters and more stringent criteria were used in determining the "ground truth" valence of those images. Thus, by using more sophisticated metrics of categorization performance than simple percent-correct scores, and optimizing image selection, we determined that affective information can be extracted from images presented as briefly as 17ms. Further exploring the valence categorization task as a function of semantic category revealed no significant differences in performance across semantic categories in either experiment 1 or 2. While the biological preparedness hypothesis would have predicted that animate stimuli should be processed sooner than inanimate stimuli, we found no evidence for this in our data, although in experiment 2 we did see that the valence of negative animals and positive humans were categorized at above chance levels before correcting for response bias suggesting those valence by semantic conditions may have some temporal processing advantage over others.

We turn finally to the question of whether our findings support either the cognitive or affective primacy hypothesis. Direct comparison of performance between the semantic and valence categorization tasks was not possible due to the differing levels of chance performance. Hence, we addressed this issue by calculating the conditional probabilities of getting one task correct conditioned on getting the other task correct or incorrect, and vice versa. These conditional probabilities provide insight into whether semantic or affective information is extracted first. Positive human images presented for 33ms in experiment 1, and 17ms in experiment 2, showed above chance performance on the valence task only when the semantic task was also done correctly ($P(V|S) > 33\%$, but $P(V|\sim S) < 33\%$), but performance on the semantic task was significant both when the valence task was done accurately ($P(S|V)$ ) or not ($P(S|\sim V)$. This was also found for negative animals, humans and objects at 50ms in experiment 1 and 17ms in experiment 2. While these findings do not establish causality, together with previous findings (Nummenmaa, Hyona, & Calvo, 2010) they are consistent with the cognitive primacy hypothesis which asserts that semantic information is extracted before affective information. Furthermore, they are also inconsistent with the affective primacy hypothesis which asserts that affective information is extracted prior to, and independently of, semantic information.

In summary our findings indicate that negative valence appears to interfere with the processing of rapidly presented stimuli required for performance of categorization tasks. Furthermore, for categorization performance at least, there is greater evidence for semantic than affective primacy and little, if any, evidence for any advantage being conferred by biological preparedness. This suggests that while stimulus valence might be extracted rapidly by subcortical systems involved in such instinctual behaviors as freezing, it does not seem to feed-forward in a manner able to inform performance of simple categorization tasks. This has important implications for the mechanisms by which we determine how best to rapidly respond to emotional stimuli in our environment.

# Supplemental Figures

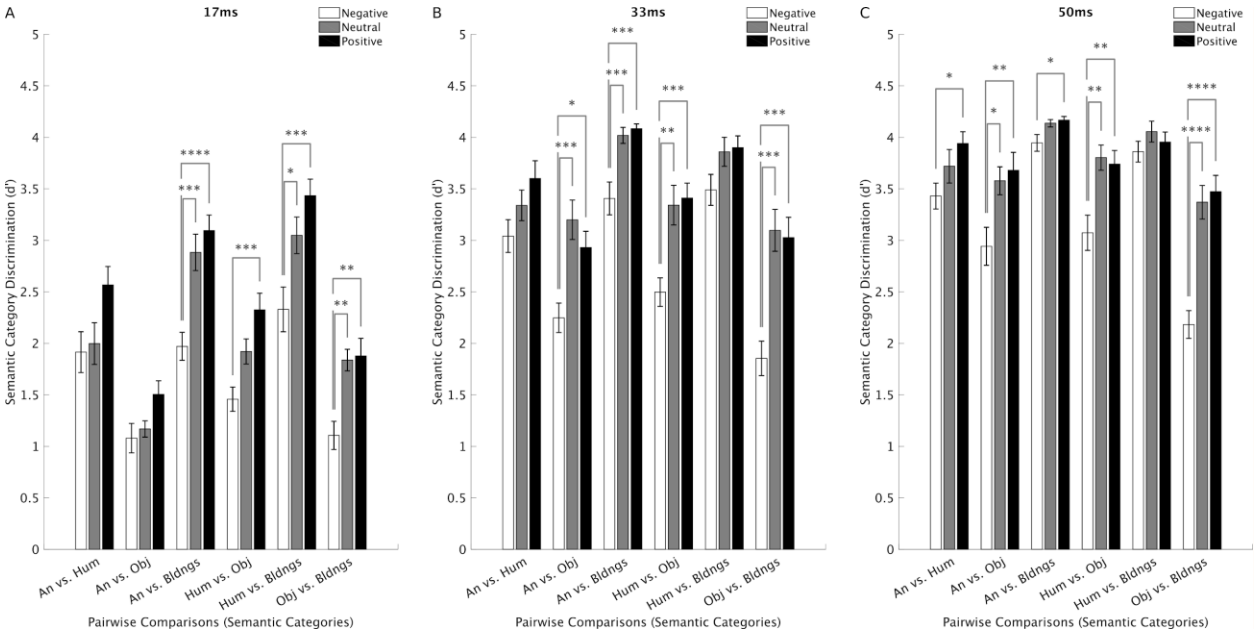


**Figure 4-S1. Semantic Categorization Task Pairwise-d' Scores for Experiment 1.**
Semantic categorization task pairwise d' scores between all six possible pairs of semantic categories from experiment 1. While more traditional, these d' scores cannot account for false positives from the two semantic categories not in the pair, unlike the mAFC-d' scores shown in Figure 4-3. A score of 0 indicates chance performance. Each plot shows results from one presentation speed (17ms, 33ms, 50ms), and scores are broken down by valence and semantic category pairs. Error bars show standard error (S.E) across subjects, stars above each bar indicate conditions that were classified with above chance performance, while lines between bars indicate which conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

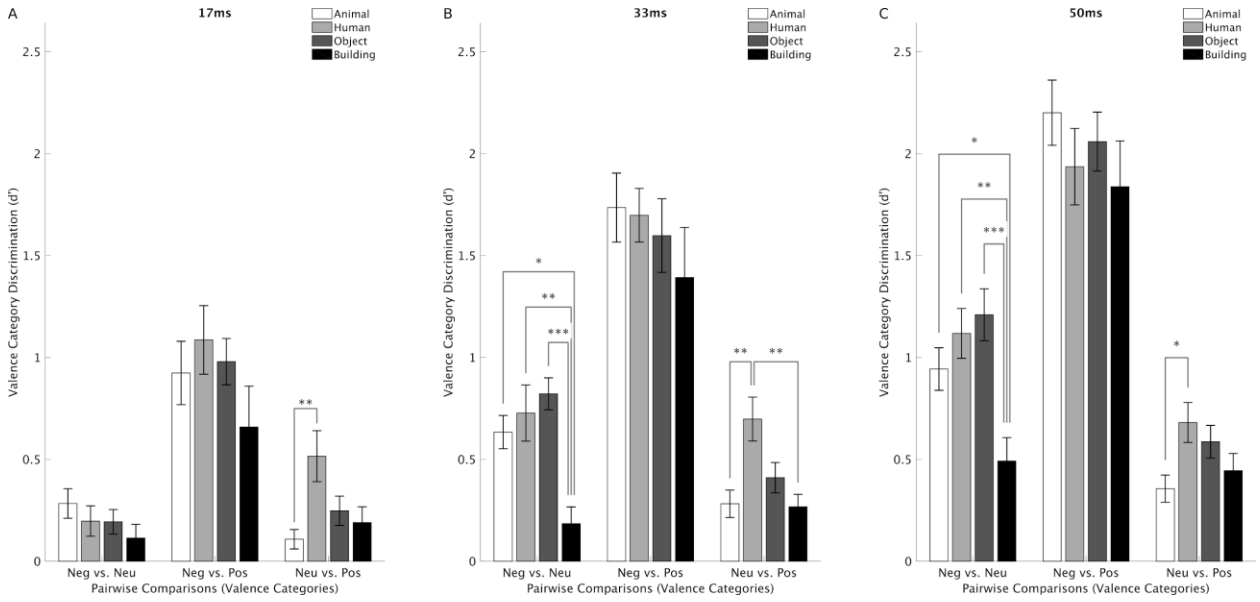


**Figure 4-S2. Valence Categorization Task Pairwise-d' Scores for Experiment 1.**
Valence categorization task pairwise d' scores between all three possible pairs of valence categories from experiment 1. While more traditional, these d' scores cannot account for false positives from the valence category not in the pair, unlike the mAFC-d' scores shown in Figure 4-5. A score of 0 indicates chance performance. Each plot shows results from one presentation speed (17ms, 33ms, 50ms), and scores are broken down by semantic category and the valence category pairs. Error bars show standard error (S.E) across subjects, stars above each bar indicate conditions that were classified with above chance performance, while lines between bars indicate which conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

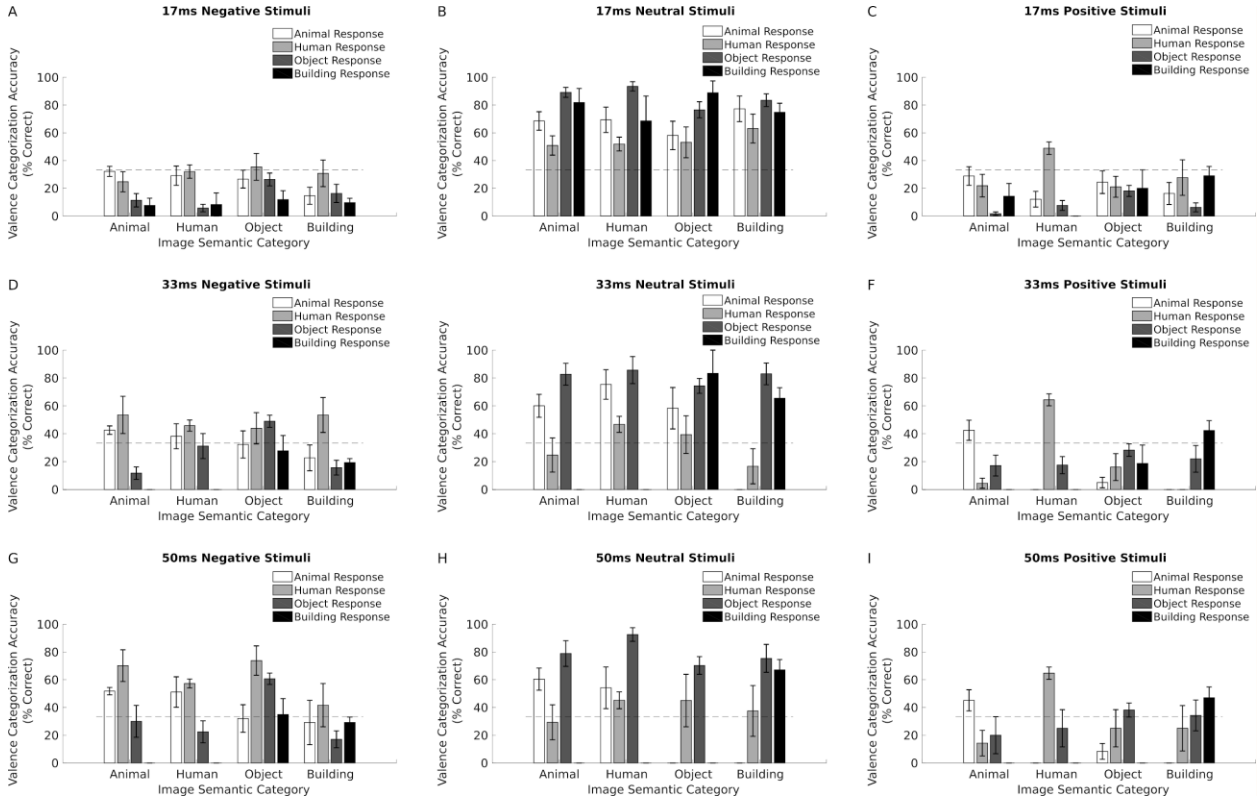


**Figure 4-S3. Valence Categorization Task Performance as a Function of Semantic Categorization Task Performance for Experiment 1.**

Bar plots showing valence categorization task performance in experiment 1 for each presentation speed by valence condition, with the four true semantic categories broken down by the four possible subject responses to the semantic categorization task. Valence categorization performance is thus plotted as a function of semantic categorization performance. Error bars show standard error (S.E) across subjects and dotted lines indicate chance performance (25%). Images of positive humans are better categorized as positive when the semantic category human is also correctly classified across all three presentation speeds. Additionally, at 50ms, images of negative animals, humans and objects are better classified as negative when their respective semantic category is correctly classified. The response bias to select neutral images when the valence category is unknown makes interpretation of this figure for neutral images difficult.

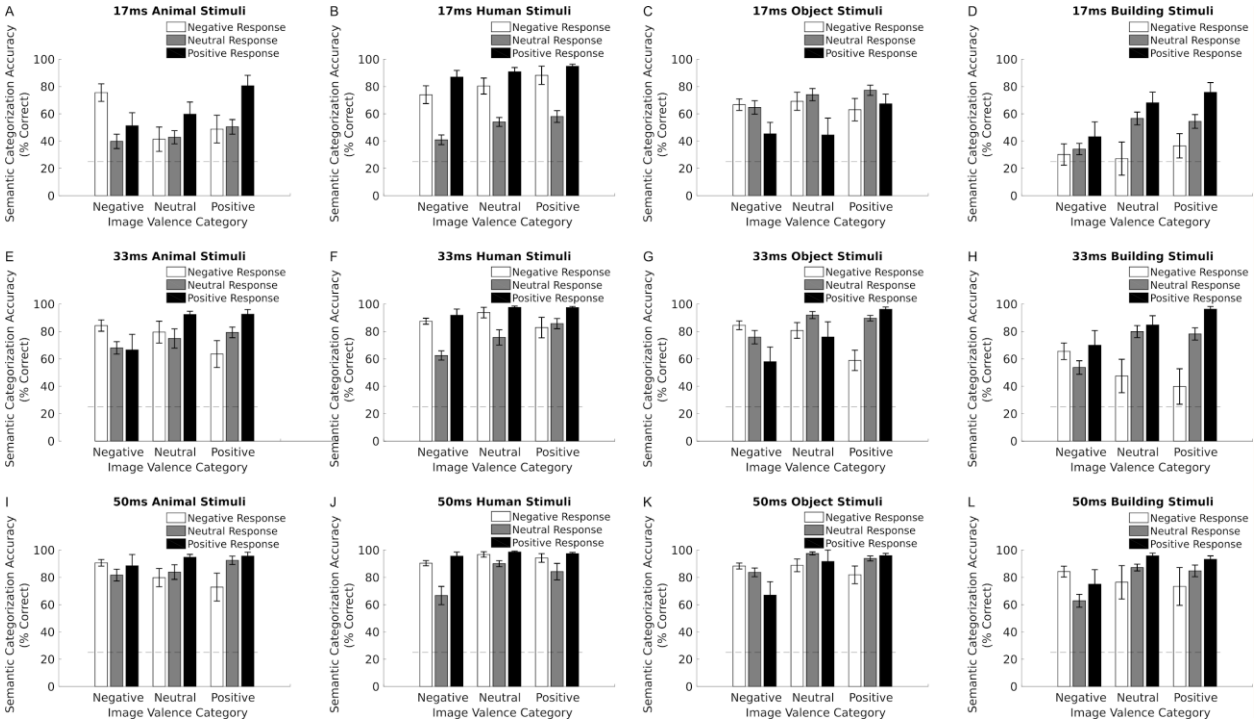


**Figure 4-S4. Semantic Categorization Task Performance as a Function of Valence Categorization Task Performance for Experiment 1.**

Bar plots showing semantic categorization task performance from experiment 1 for each presentation speed by semantic category condition, with the three true valence categories broken down by the three possible subject responses to the valence categorization task. Semantic categorization performance is thus plotted as a function of valence categorization performance. Error bars show standard error (S.E) across subjects and dotted lines indicate chance performance (33%). It does not seem that getting the valence category correct is associated with getting the semantic category correct across all speed by semantic category conditions.

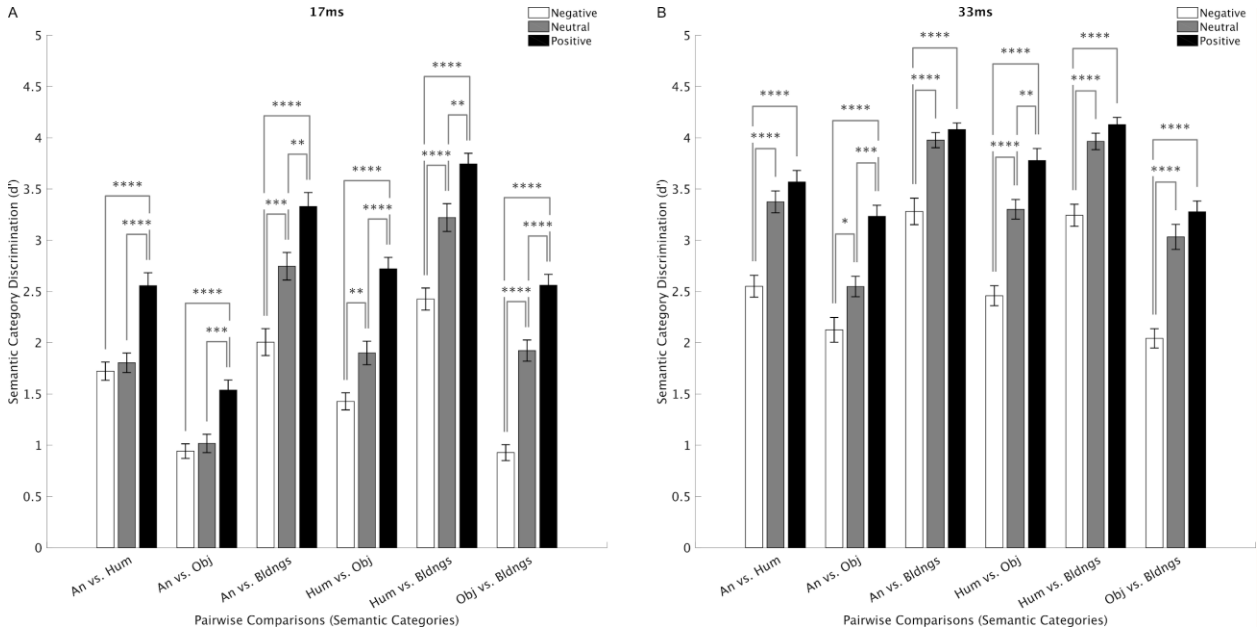


**Figure 4-S5. Semantic Categorization Task Pairwise-d' Scores for Experiment 2.**
Semantic categorization task pairwise d' scores between all six possible pairs of semantic categories from experiment 2. While more traditional, these d' scores cannot account for false positives from the two semantic categories not in the pair, unlike the mAFC-d' scores shown in Figure 4-8. A score of 0 indicates chance performance. Each plot shows results from one presentation speed (17ms, 33ms), and scores are broken down by valence and semantic category pairs. Error bars show standard error (S.E) across subjects, stars above each bar indicate conditions that were classified with above chance performance, while lines between bars indicate which conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

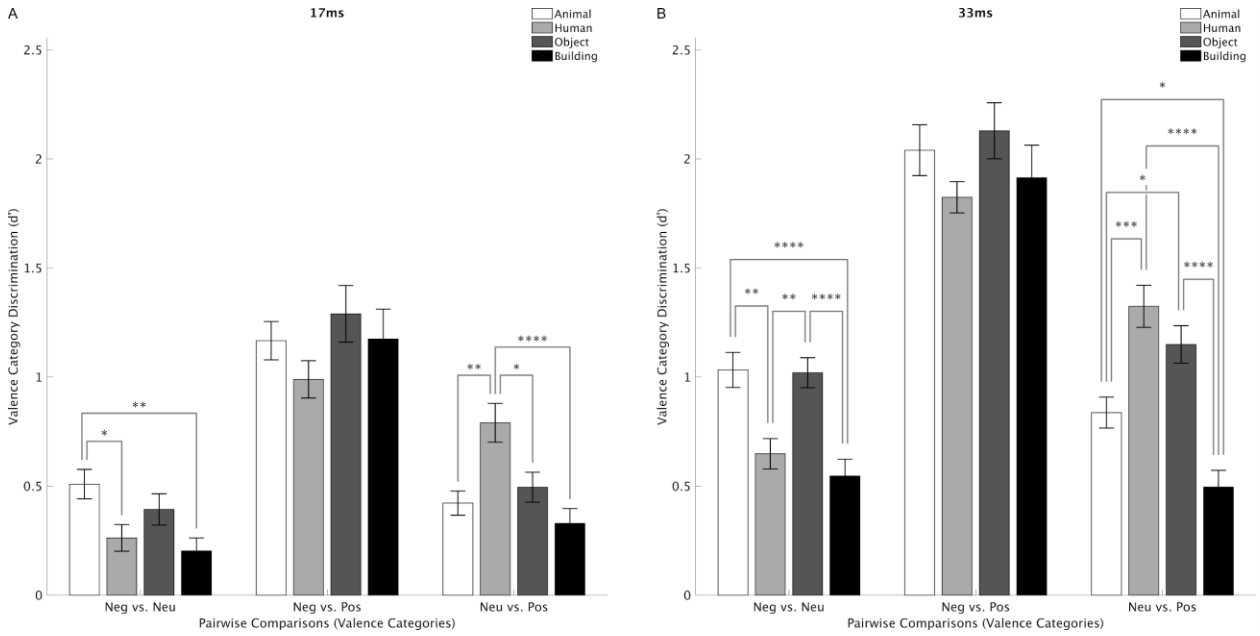


**Figure 4-S6. Valence Categorization Task Pairwise-d' Scores for Experiment 2.**
Valence categorization task pairwise d' scores between all three possible pairs of valence categories from experiment 2. While more traditional, these d' scores cannot account for false positives from the valence category not in the pair, unlike the mAFC-d' scores shown in Figure 4-10. A score of 0 indicates chance performance. Each plot shows results from one presentation speed (17ms, 33ms), and scores are broken down by semantic category and the valence category pairs. Error bars show standard error (S.E) across subjects, stars above each bar indicate conditions that were classified with above chance performance, while lines between bars indicate which conditions were significantly different from each other (t-tests, Bonferroni corrected: * p<.05, ** p<.01, *** p<.001, **** p<.0001).

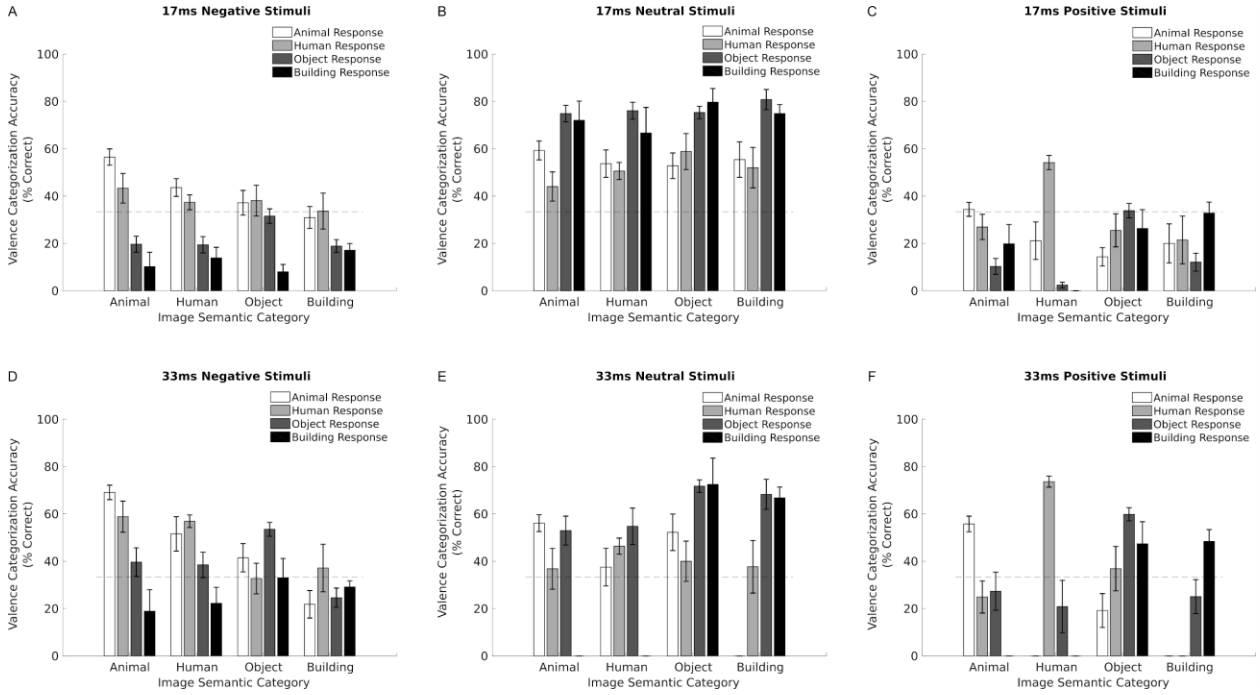


**Figure 4-S7. Valence Categorization Task Performance as a Function of Semantic Categorization Task Performance for Experiment 2.**

Bar plots showing valence categorization task performance in experiment 2 for each presentation speed by valence condition, with the four true semantic categories broken down by the four possible subject responses to the semantic categorization task. Valence categorization performance is thus plotted as a function of semantic categorization performance. Error bars show standard error (S.E) across subjects and dotted lines indicate chance performance (25%). At 17ms, images of positive humans are better categorized as positive when the semantic category human is also correctly classified, while at 33ms all positive image categories are better categorized as positive when their semantic category is correctly classified. Additionally, at 33ms, images of negative animals, humans and objects are better classified as negative when their respective semantic category is correctly classified. The response bias to select neutral images when the valence category is unknown makes interpretation of this figure for neutral images difficult.

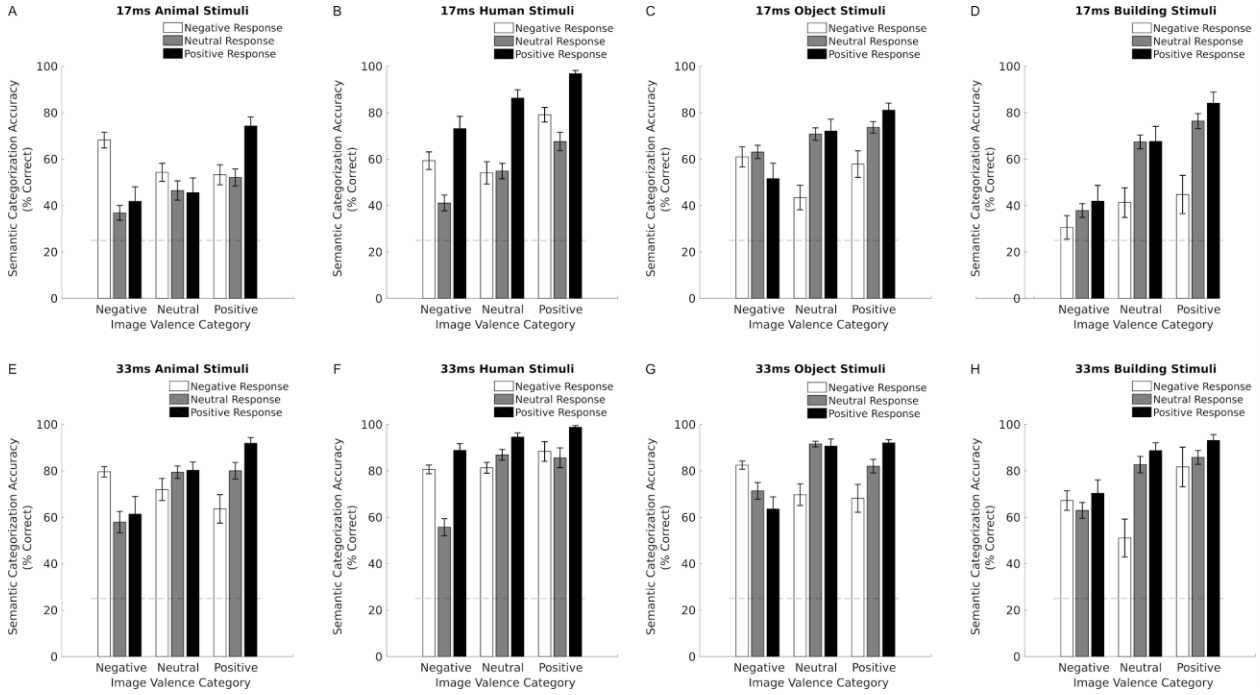


**Figure 4-S8. Semantic Categorization Task Performance as a Function of Valence Categorization Task Performance for Experiment 2.**

Bar plots showing semantic categorization task performance from experiment 1 for each presentation speed by semantic category condition, with the three true valence categories broken down by the three possible subject responses to the valence categorization task. Semantic categorization performance is thus plotted as a function of valence categorization performance. Error bars show standard error (S.E) across subjects and dotted lines indicate chance performance (33%).

# Chapter 5. Conclusions

This dissertation began with a question: how does the human visual system extract semantic and emotional information useful in determining the appropriate behavioral response to situations of high survival value? Throughout the previous three chapters I have presented evidence, and reviewed previous scientific findings, in an attempt to answer this question. In chapter 2 I reported the results of a voxel-wise modeling analysis of fMRI data collected while subjects viewed over 1600 emotional naturalistic images. Using machine learning techniques to control for model overfitting and to handle the large feature spaces modeled, I reported that a combined semantic, valence and arousal (CSVA) model was able to predict brain activity across wide swaths of OTC. These predictions were greater than predictions made by a semantics only or emotion only across a significantly greater number of voxels, across subjects. Using a group-level principal components analysis (PCA) on the CSVA model weights across voxels within OTC revealed the top 3 dimensions of this representational space accounts for more than 50% of the variance in those model weights. By comparing the top 3 PCs with several hypothetical dimensions, we showed that, in addition to animacy, valence and arousal information were contained within these dimensions, but predominantly for animate stimuli. We next aimed to determine whether OTC tuning might be useful, in a linear manner, to determine the appropriate behavioral responses to the images that subjects viewed in the scanner. To that aim, we used each image's scores on a varying number of CSVA model PCs to predict the appropriate behavioral responses to those images. The amount of variance explained by the CSVA model PCs significantly exceeded that achieved when using PCs of the stimulus features themselves, across all images viewed. Furthermore, CSVA model PCs also explained significantly more variance than that explained by OTC tuning to low-level image properties, semantic features alone, or emotional features alone. We interpreted these findings as evidence that, in addition to semantic information, the valence and arousal of predominantly animate stimuli are represented within OTC in such a fashion as to allow for a linear readout of that information useful to selection of the appropriate behavioral response to the emotional scenes viewed. Additionally, it is unlikely that these findings are simply a result of a "gain" function due to emotional attention, as a simple "gain" could not account for the observed valence representation of animate stimuli within OTC.

These findings from chapter 2 were not without limitations however. In chapter 3 I addressed theoretical limitations relating to our claim that semantic information is represented within OTC. Recent studies involving the congenitally blind have shed light on the supramodal nature of object category selectivity within OTC, indicating that this region does in fact represent abstract semantic information and not just visual feature information. Furthermore, a reassessment of the traditional role of OTC as an object detector, along with new empirical findings, suggest that the representations within OTC may be idiosyncratically structured to allow for appropriate domain-specific behavior, such as navigation when viewing landscapes, social cognition when viewing human faces, and tool manipulation when viewing inanimate artifacts. In a second literature review, I recounted behavioral and neuroscientific studies investigating the perception and attentional effects of emotional visual stimuli. It is well established that emotional stimuli (especially negative stimuli) command attention. Neuroscientific research suggests that re-entrant projections from amygdala to visual cortex increase neural firing relevant to the emotional stimulus, acting as a gain function which accounts for this attentional capture. The mechanisms and brain regions involved in signally the

amygdala to activate this gain function are debated, however. Two competing hypothesis argue that either a fast subcortical route (the "two-routes" hypothesis) or fast feed-forward sweep of OTC (the "two-stage" hypothesis) extract the emotional information necessary to alert the amygdala to initiate this gain function.

Chapter 2 also contained the results of seven empirical analyses of the fMRI dataset collected for the study in chapter 2 that address methodological limitations of that chapter. In a 2014 paper, Chikazoe and colleagues used representational similarity analysis (RSA) to argue that valence and animacy are represented within VTC, but not in the same regions. This conflicted with our findings from chapter 2 that the combination of semantics, valence and arousal are all located within overlapping regions of OTC. To resolve this discrepancy we conducted a RSA and found that indeed animacy, valence, and arousal are represented within the overlapping regions of OTC. Additionally, to test whether our findings from chapter 2 could be inferred using simpler univariate approaches, we conducted a univariate SPM analysis. We found that although representation of animacy, and valence and arousal for animate stimuli, could be found within OTC, valence of animate stimuli were not necessarily found in overlapping regions across subjects. In order to determine the contributions to model CSVA performance made by semantic vs. emotional features we partitioned the out-of-sample variance explained by the CSVA model in chapter 2. We found that the when removing variance explained by either semantic features or emotional features an impressive amount of variance was still explained by the combined representation of semantic and emotional information. Finally, several control analyses were conducted. The first of which showed that early visual cortical (EVC) voxel tuning could not account for the PCA results reported in chapter 2. Furthermore, tuning in frontal lobe regions was no better than that in OTC at explaining appropriate behavioral responses. By regressing out variance in the fMRI BOLD signal accounted for by respiration and heart rate signals, we showed that physiological noise did not meaningfully alter CSVA model performance nor dimensionality of the top 3 PCs of OTC tuning. Finally, we showed that by removing out-of-sample variance accounted for by a low-level structural model did not drastically change OTC scores on the top 2 PCs, while it had a mild effect on PC 3 scores. Taken together, these supplemental analyses suggest that, while the results from our VWM study showed a more clear picture of the role emotion plays in OTC representations, our small-N study utilized a large stimulus set which was well powered to allow even RSA and univariate SPM analyses to uncover some of the same representations as the VWM analysis of chapter 2.

Turning from cortical to behavioral representations of naturalistic emotional images, in chapter 4 we reported findings of two categorization task experiments using briefly presented emotional images. Replicated across two experiments, and using mAFC-d' scores to control for response bias, we found that both semantic (animal, human, object, buildings) and valence (negative, neutral, positive) categories can be accurately identified at presentation times as short as 17ms. In contrast to predictions made by the biological preparedness hypothesis, we found that negative stimuli were less accurately categorized than were neutral or positive stimuli across both experiments. Additionally, in experiment two, where ambiguous stimuli had been removed, the semantic category of positive stimuli was better identified than was the semantic category of negative or neutral stimuli. This suggests a penalty for the categorization of negative images, and perhaps a benefit for the categorization of positive images. By calculating the conditional probabilities of accurately completing one task (e.g semantic) when accurately completing the other (e.g. valence), we looked for support of either the affective or cognitive primacy hypotheses. At several presentation speeds and across several semantic by valence conditions,

the valence categorization task was only done accurately when the semantic categorization task was done accurately. The converse of this was not true however as the semantic categorization task done accurately regardless of whether the valence task was done accurately. This was observed in both experiments. We took these results as evidence in support of the cognitive primacy hypothesis, and found no evidence in support of the affective primacy hypothesis.

Interpreting these findings together with the previous literature reviewed in this dissertation, a picture is beginning to emerge of the possible brain regions and representations involved in the perception of naturalistic emotional images that can initiate appropriate behavioral responses to the situations of high-survival value depicted in those images. While still speculative, and in need of further research to flush out, I suggest that supramodal representations across multiple sensory cortices form a semantic "network" that represents external stimuli, which subserves synchronization of perception and attention across modalities, and which allows for the initiation of appropriate behavioral responses to those stimuli. Feed-forward passes through the sensory cortices of this network may serve to elicit rudimentary emotional information which, in concert with the pulvinar and amygdala, initiate a gain function which boosts multimodal sensory processing of those stimuli through reentrant feedback serving to focus attentional and behavioral resources towards the high survival-value stimulus.

# References

Adam, R., & Noppeney, U. (2010). Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. NeuroImage, 52(4), 1592–1602. https://doi.org/10.1016/j.neuroimage.2010.05.002

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. NeuroImage, 35(4), 1480–1494. https://doi.org/10.1016/j.neuroimage.2007.02.005

Alpers, G. W., & Gerdes, A. B. M. (2007). Here Is Looking at You: Emotional Faces Predominate in Binocular Rivalry. Emotion, 7(3), 495–506. https://doi.org/10.1037/1528-3542.7.3.495

Alpers, G. W., & Pauli, P. (2006). Emotional pictures predominate in binocular rivalry. Cognition and Emotion, 20(5), 596–607. https://doi.org/10.1080/02699930500282249

Amedi, A., Von Kriegstein, K., Van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. Experimental Brain Research, 166(3–4), 559–571. https://doi.org/10.1007/s00221-005-2396-5

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: a response-related fMRI study. Human Brain Mapping, 23(4), 200–9. https://doi.org/10.1002/hbm.20048

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. Journal of Experimental Psychology: General, 134(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. Journal of Neuroscience, 29(34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Armony, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by masked angry faces: An event-related fMRI study. NeuroImage, 13(6), 378. https://doi.org/10.1016/S1053-8119(01)91721-5

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. Vision Research, 45(11), 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), Advances in social cognition: The automaticity of everyday life (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.

Batty, M. J., Cave, K. R., & Pauli, P. (2005). Abstract stimuli associated with threat through conditioning cannot be detected preattentively. Emotion, 5(4), 418–430. https://doi.org/10.1037/1528-3542.5.4.418

Bi, Y., Wang, X., & Caramazza, A. (2016). Object Domain and Modality in the Ventral Visual Pathway. Trends in Cognitive Sciences, 20(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. Trends in Cognitive Sciences, 15(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Bocanegra, B. R., & Zeelenberg, R. (2009). Dissociating Emotion-Induced Blindness and Hypervision. Emotion, 9(6), 865–873. https://doi.org/10.1037/a0017749

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. Journal of Neurophysiology, 107(5), 1443–1456. https://doi.org/10.1152/jn.00619.2011

Bracci, S., & Op De Beeck, H. (n.d.). Is a cow mug a cow or a mug? Object Appearance, but not Semantics, is Represented in The Human Category-Selective Cortex.

Bracci, S., Ritchie, J. B., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. Neuropsychologia, 105(October 2016), 153–164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010

Bradley, M. M., Sabatinelli, D., Lang, P. J., Fitzsimmons, J. R., King, W., & Desai, P. (2003). Activation of the visual cortex in motivated attention. Behavioral Neuroscience, 117(2), 369–380. https://doi.org/10.1037/0735-7044.117.2.369

Bradley, M. M., Sabatinelli, D., Lang, P. J., Fitzsimmons, J. R., King, W., & Desai, P. (2003). Activation of the visual cortex in motivated attention. Behavioral Neuroscience, 117(2), 369–380. https://doi.org/10.1037/0735-7044.117.2.369

Brosch, T., Pourtois, G., Sander, D., & Vuilleumier, P. (2011). Additive effects of emotional, endogenous, and exogenous attention: Behavioral and electrophysiological evidence. Neuropsychologia, 49(7), 1779–1787. https://doi.org/10.1016/j.neuropsychologia.2011.02.056

Calvo, M. G., Avero, P., & Nummenmaa, L. (2011). Primacy of emotional vs . semantic scene recognition in peripheral vision. Cognition & Emotion, 25(8), 1358–1375.

Calvo, M. G., & Lang, P. J. (2005). Parafoveal Semantic Processing of Emotional Visual Scenes. Journal of Experimental Psychology: Human Perception and Performance, 31(3), 502–519. https://doi.org/10.1037/0096-1523.31.3.502

Calvo, M. G., & Nummenmaa, L. (2007). Processing of Unattended Emotional Visual Scenes. Journal of Experimental Psychology: General, 136(3), 347–369. https://doi.org/10.1037/0096-3445.136.3.347

Calvo, M. G., Nummenmaa, L., & Avero, P. (2008). Visual search of emotional faces eye-movement assessment of component processes. Experimental Psychology, 55(6), 359–370. https://doi.org/10.1027/1618-3169.55.6.359

Calvo, M. G., Nummenmaa, L., & Hyönä, J. (2008). Emotional Scenes in Peripheral Vision: Selective Orienting and Gist Processing, But Not Content Identification. Emotion, 8(1), 68–80. https://doi.org/10.1037/1528-3542.8.1.68

Carp, J. (2013). Optimizing the order of operations for movement scrubbing: Comment on Power et al. NeuroImage, 76, 436–438. https://doi.org/10.1016/j.neuroimage.2011.12.061

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. Proceedings of the National Academy of Sciences, 111(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Chavis, J. M., & Kisley, M. a. (2012). Adult Attachment and Motivated Attention to Social Images: Attachment-Based Differences in Event-Related Brain Potentials to Emotional Images. Journal of Research in Personality, 46(1), 55–62. https://doi.org/10.1016/j.jrp.2011.12.004

Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2014). Population coding of affect across stimuli, modalities and individuals. Nature Neuroscience, (June). https://doi.org/10.1038/nn.3749

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. Nature Neuroscience, 17(3), 455–462. https://doi.org/10.1038/nn.3635

Clarke, A., & Tyler, L. K. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. Journal of Neuroscience, 34(14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014

Cohen, E. H., & Tong, F. (2015). Neural Mechanisms of Object-Based Attention. Cerebral Cortex, 25(4), 1080–1092. https://doi.org/10.1093/cercor/bht303

Cohen, N., Henik, A., & Mor, N. (2011). Can emotion modulate attention? Evidence for reciprocal links in the attentional network test. Experimental Psychology, 58(3), 171–179. https://doi.org/10.1027/1618-3169/a000083

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Haxby, J. V. (2012). The representation of biological classes in the human brain. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 32(8), 2608–18. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Costantini, M., Urgesi, C., Galati, G., Romani, G. L., & Aglioti, S. M. (2011). Haptic perception and body representation in lateral and medial occipito-temporal cortices. Neuropsychologia, 49(5), 821–829. https://doi.org/10.1016/j.neuropsychologia.2011.01.034

de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. Proceedings of the National Academy of Sciences, 99(6), 4121–4126. https://doi.org/10.1073/pnas.062018499

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. NeuroReport, 10(18), 3759–3763. https://doi.org/10.1097/00001756-199912160-00007

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. The Journal of Neuroscience, 37(27), 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

de Volder, A. G., Toyama, H., Kimura, Y., Kiyosawa, M., Nakano, H., Vanlierde, A., Senda, M. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. NeuroImage, 14(1 I), 129–139. https://doi.org/10.1006/nimg.2001.0782

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Saxe, R. (2017). Organization of high-level visual cortex in human infants. Nature Communications, 8, 1–10. https://doi.org/10.1038/ncomms13995

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition, 79(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., Way, H. (2001). 2001_, 40, 1–14. https://doi.org/10.1016/S0028-3932(01)00066-5

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions To Emotional Facial Expressions. Psychological Science, 11(1), 86–89.

Dolan, R. J., & Vuilleumier, P. (2003). Amygdala automaticity in emotional processing. Annals of the New York Academy of Sciences, 985, 348–355. https://doi.org/10.1111/j.1749-6632.2003.tb07093.x

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science (New York, N.Y.), 293(5539), 2470–3. https://doi.org/10.1126/science.1063414

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. Perception and Psychophysics, 63(6), 1004–1013. https://doi.org/10.3758/BF03194519

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature, 392(6676), 598–601. https://doi.org/10.1038/33402

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cerebral Cortex, 22(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of Emotional Information in Voice-Sensitive Cortices. Current Biology, 19(12), 1028–1033. https://doi.org/10.1016/j.cub.2009.04.054

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A Limit to the Speed of Processing in Ultra-Rapid Visual Categorization of Novel Natural Scenes. Journal of Cognitive Neuroscience, 13(2), 171–180. https://doi.org/10.1162/089892901564234

Fairhall, S. L., Anzellotti, S., Ubaldi, S., & Caramazza, A. (2014). Person- and place-selective neural substrates for entity-specific semantic access. Cerebral Cortex, 24(7), 1687–1696. https://doi.org/10.1093/cercor/bht039

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? Journal of Vision, 7(1), 10. https://doi.org/10.1167/7.1.10

Ferrari, V., Codispoti, M., Cardinale, R., & Bradley, M. M. (2008). Directed and motivated attention during processing of natural scenes. Journal of Cognitive Neuroscience, 20(10), 1753–61. https://doi.org/10.1162/jocn.2008.20121

Finucane, A. M. (2011). The Effect of Fear and Anger on Selective Attention. Emotion, 11(4), 970–974. https://doi.org/10.1037/a0022574

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? Cognition and Emotion, 14(1), 61–92. https://doi.org/10.1080/026999300378996

Friston, K. J., Glaser, D. E., Henson, R. N. a, Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. NeuroImage, 16(2), 484–512. https://doi.org/10.1006/nimg.2002.1090

Gable, P., & Harmon-Jones, E. (2010). The blues broaden, but the nasty narrows: Attentional consequences of negative affects low and high in motivational intensity. Psychological Science, 21(2), 211–215. https://doi.org/10.1177/0956797609359622

Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface visualizer for fMRI. Frontiers in Neuroinformatics, 9(September), 1–12. https://doi.org/10.3389/fninf.2015.00023

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, a W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. Journal of Cognitive Neuroscience, 12(3), 495–504. https://doi.org/10.1162/089892900562165

Gehring, W. J., Goss, Coles, Meyer, & Donchin. (1993). A neural system for error detection and compensation. Psychological Science, 4(6), 385–390.

Glascher, J., & Adolphs, R. (2003). Processing of the Arousal of Subliminal and Supraliminal Emotional Stimuli by the Human Amygdala. The Journal of Neuroscience, 23(32), 10274–10282. https://doi.org/23/32/10274 [pii]

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magnetic Resonance in Medicine, 44(1), 162–167. https://doi.org/10.1002/1522-2594(200007)44:1<162::AID-MRM23>3.0.CO;2-E

Goodale, M. A., & Milner, A. D. (1992). Visual pathways to perception and action. Trends in Neuroscience, 15(1), 20–25. https://doi.org/10.1016/S0079-6123(08)60379-9

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. Nature Neuroscience, 8(2), 145–146. https://doi.org/10.1038/nn1392

Greene, M. R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. Psychological Science, 20(4), 464–72.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees Michelle. Cognitive Psychology, 58(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001.Recognition

Grice, J., Barrett, P., Cota, L., Felix, C., Taylor, Z., Garner, S., Vest, A. (2017). Four Bad Habits of Modern Psychologists. Behavioral Sciences, 7(3), 53. https://doi.org/10.3390/bs7030053

Grill-Spector, K., & Kanwisher, N. (2005). Visual Recognition. Psychological Science, 16(2), 152–160. https://doi.org/10.1111/j.0956-7976.2005.00796.x

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. Neuron, 21(1), 191–202. https://doi.org/10.1016/S0896-6273(00)80526-7

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nature Reviews. Neuroscience, 15(8), 536–548. https://doi.org/10.1038/nrn3747

Hamilton, R., Keenan, J. P., Catala, M., & Pascual-Leone, A. (2000). Alexia for Braille following bilateral occipital stroke in an early blind woman. NeuroReport, 11(2), 237–240. https://doi.org/10.1097/00001756-200002070-00003

Hansen, K. a, Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 27(44), 11896–911. https://doi.org/10.1523/JNEUROSCI.2991-07.2007

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. NeuroImage, 79, 1–9. https://doi.org/10.1016/j.neuroimage.2013.04.051

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The Weirdest People in the World? Behavioral And Brain Sciences, 33, 61–135.

Horstmann, G., & Bauland, A. (2006). Search asymmetries with real faces: Testing the anger-superiority effect. Emotion, 6(2), 193–207. https://doi.org/10.1037/1528-3542.6.2.193

Hutchison, R. M., Culham, J. C., Everling, S., Flanagan, J. R., & Gallivan, J. P. (2014). Distinct and distributed functional connectivity patterns across cortex reflect the domain-specific constraints of object, face, scene, body, and tool category-selective modules in the ventral visual pathway. NeuroImage, 96, 216–236. https://doi.org/10.1016/j.neuroimage.2014.03.068

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron, 76(6), 1210–24. https://doi.org/10.1016/j.neuron.2012.10.014

Intraub, H. (1981). Rapid Conceptual Identification of Sequentially Presented Pictures. Journal of Experimental Psychology. Human Perception and Performance, 7(3), 604–610.

Ishai, A. (2017). Jede Praxis braucht ein Kassenbuch! MMW-Fortschritte Der Medizin, 159(21–22), 38. https://doi.org/10.1007/s15006-017-0378-8

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. Neuron, 28(3), 979–990. https://doi.org/10.1016/S0896-6273(00)00168-9

Jones, E., Oliphant, E., & Peterson, P. (n.d.). SciPy: Open Source Scientific Tools for Python.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. Vision Research, 47(26), 3286–3297. https://doi.org/10.1016/j.visres.2007.09.013

Kamps, F. S., Lall, V., & Dilks, D. D. (2016). The occipital place area represents first-person perspective motion information through scenes. Cortex, 83, 17–26. https://doi.org/10.1016/j.cortex.2016.06.022

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 17(11), 4302–11.

Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In The adaptive mind.

Kawasaki, H., Adolphs, R., Kaufman, O., Damasio, H., Damasio, A. R., Granner, M., Howard, M. A. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. Nature Neuroscience, 4(1), 15–16. https://doi.org/10.1038/82850

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. Nature, 452(7185), 352–5. https://doi.org/10.1038/nature06713

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. Cerebral Cortex, 15(8), 1187–1197. https://doi.org/10.1093/cercor/bhi001

Keysers, C., & Perrett, D. I. (2002). ScienceDirect.com - Trends in Cognitive Sciences - Visual masking and RSVP reveal neural competition. Trends In Cognitive Sciences, 6(3), 120–125.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Research, 46(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Kitada, R., Yoshihara, K., Sasaki, A. T., Hashiguchi, M., Kochiyama, T., & Sadato, N. (2014). The Brain Network Underlying the Recognition of Hand Gestures in the Blind: The Supramodal Role of the Extrastriate Body Area. Journal of Neuroscience, 34(30), 10096–10108. https://doi.org/10.1523/JNEUROSCI.0500-14.2014

Kitada, R., Johnsrude, I. S., Kochiyama, T., & Lederman, S. J. (2009). Functional specialization and convergence in the occipito-temporal cortex supporting haptic and visual identification of human faces and body parts: An fMRI study. Journal of Cognitive Neuroscience, 21(10), 2027–2045. https://doi.org/10.1162/jocn.2009.21115

Kitada, R., Okamoto, Y., Sasaki, A. T., Kochiyama, T., Miyahara, M., Lederman, S. J., & Sadato, N. (2013). Early visual experience and the recognition of basic facial expressions: involvement of the middle temporal and inferior frontal gyri during haptic identification by the early blind. Frontiers in Human Neuroscience, 7(January), 1–15. https://doi.org/10.3389/fnhum.2013.00007

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), The psychology of evaluation: Affective processes in cognition and emotion. (pp. 7–49). Mahwah, NJ: Erlbaum.

Kriegeskorte, N., Mur, M., Da, R., Kiani, R., Bodurka, J., Esteky, H., Pa, B. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey [Supplemental Materials]. Neuron, 60, 1–30.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2(November), 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. a, Kiani, R., Bodurka, J., Esteky, H., Bandettini, P. a. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60(6), 1126–41. https://doi.org/10.1016/j.neuron.2008.10.043

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks Alex. Advances in Neural Information Processing Systems, 1097–1105. https://doi.org/10.2165/00129785-200404040-00005

Lacey, S., Campbell, C., & Sathian, K. (2007). Vision and touch: Multiple or multisensory representations of objects? Perception, 36(10), 1513–1521. https://doi.org/10.1068/p5850

Lai, V. T., Hagoort, P., & Casasanto, D. (2012). Affective primacy vs. cognitive primacy: Dissolving the debate. Frontiers in Psychology, 3(JUL), 1–8. https://doi.org/10.3389/fpsyg.2012.00243

Lambert, S., Sampaio, E., Mauss, Y., & Scheiber, C. (2004). Blindness and brain plasticity: Contribution of mental imagery? An fMRI study. Cognitive Brain Research, 20(1), 1–11. https://doi.org/10.1016/j.cogbrainres.2003.12.012

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neuroscience, 23(11), 571–579.

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., & Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. Neuropsychologia, 35(11), 1437–1444. https://doi.org/10.1016/S0028-3932(97)00070-5

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., & Nangia, V. (1998). Emotional arousal and activation of the visual cortex : An fMRI analysis, 199–210.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual.

Lazarus, R. S. (1984). On the primacy of Cognition. American Psychologist, 39(2), 124–129. https://doi.org/10.1037//0003-066X.39.2.124

Le Clec'H, G., Dehaene, S., Cohen, L., Mehler, J., Dupoux, E., Poline, J. B., Le Bihan, D. (2000). Distinct cortical areas for names of numbers and body parts independent of language and input modality. NeuroImage, 12(4), 381–391. https://doi.org/10.1006/nimg.2000.0627

Leach, J. (2004). Why people "freeze" in an emergency: Temporal and cognitive constraints on survival responses. Aviation Space and Environmental Medicine, 75(6), 539–542.

Leach, J. (2005). Cognitive paralysis in an emergency: The role of the supervisory attentional system. Aviation Space and Environmental Medicine, 76(2), 134–136.

LeDoux, J. E. (1994). Emotion, memory and the brain. Scientific American, June(68), 32–39. https://doi.org/10.1038/scientificamerican0694-50

LeDoux, J. (2012). Rethinking the emotional brain. Neuron, 73(4), 653–76. https://doi.org/10.1016/j.neuron.2012.02.004

Lewis, J. W. (2005). Distinct Cortical Pathways for Processing Tool versus Animal Sounds. Journal of Neuroscience, 25(21), 5148–5158. https://doi.org/10.1523/JNEUROSCI.0419-05.2005

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature, 412(12 July), 150–157.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869–878. https://doi.org/10.1038/nature06976

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream, (7). https://doi.org/10.1073/pnas.1719616115

Macho, S. (2014). SDT-Models in R : A Documentation of R-Code for Fitting Signal Detection Models.

Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory: A User's Guide. https://doi.org/10.1017/CBO9781107415324.004

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-Specific Organization in the Human Brain Does Not Require Visual Experience. Neuron, 63(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. a, Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences of the United States of America, 92(18), 8135–9.

Maljkovic, V., & Martini, P. (2005). Short-term memory for scenes with affective content. Journal of Vision, 5(3), 6. https://doi.org/10.1167/5.3.6

Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. Cognition and Emotion, 13(6), 713–740. https://doi.org/10.1080/026999399379050

Moratti, S., Keil, A., & Stolarova, M. (2004). Motivated attention in emotional picture processing is reflected by activity modulation in cortical attention networks. NeuroImage, 21(3), 954–64. https://doi.org/10.1016/j.neuroimage.2003.10.030

Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, a W., Calder, a J., & Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain : A Journal of Neurology, 121 ( Pt 1, 47–57.

Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. Nature, 393(4 June), 467–470.

Mourão-Miranda, J., Volchan, E., Moll, J., Oliveira-Souza, R. de, Oliveir, L., Bramati, I., Pessoa, L. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. NeuroImage, 20(4), 1955–1963. https://doi.org/10.1016/j.neuroimage.2003.08.011

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective Priming with optimal and suboptimal stimulus exposures. Journal of Personality and Social Psychology, 64(5), 723–739.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. NeuroImage, 56(2), 400–10. https://doi.org/10.1016/j.neuroimage.2010.07.073

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron, 63(6), 902–15. https://doi.org/10.1016/j.neuron.2009.09.006

Naselaris, T., Stansbury, D. E., & Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. Journal of Physiology - Paris, 106(5–6), 239–249. https://doi.org/10.1016/j.jphysparis.2012.02.001

Nimon, K., & Reio, T. G. (2011). Regression commonality analysis: A technique for quantitative theory building. Human Resource Development Review, 10(3), 329–340. https://doi.org/10.1177/1534484311411077

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Noppeney, U. (2007). The effects of visual deprivation on functional and structural organization of the human brain. Neuroscience and Biobehavioral Reviews, 31(8), 1169–1180. https://doi.org/10.1016/j.neubiorev.2007.04.012

Noppeney, U., Price, C. J., Penny, W. D., & Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses. Cerebral Cortex, 16(3), 437–445. https://doi.org/10.1093/cercor/bhi123

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. Frontiers in Psychology, 7(JUN), 1–4. https://doi.org/10.3389/fpsyg.2016.00934

Notebaert, W., Houtman, F., Opstal, F. Van, Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. Cognition, 111(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2010). Semantic Categorization Precedes Affective Evaluation of Visual Scenes. Journal of Experimental Psychology: General, 139(2), 222–246. https://doi.org/10.1037/a0018858

Nummenmaa, L., Peets, K., & Salmivalli, C. (208AD). Automatic Activation of Adolescents ' Peer-Relational Schemas : Evidence from Priming with Facial Identity. Child Development, 79(6), 1659–1675.

O'Craven, K. M., & Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. Journal of Cognitive Neuroscience, 12(6), 1013–1023. https://doi.org/10.1162/08989290051137549

Ohman, A., Erixon, G., & Löfberg, I. (1975). Phobias and preparedness: phobic versus neutral pictures as conditioned stimuli for human autonomic responses. Journal of Abnormal Psychology, 84(1), 41–45. https://doi.org/10.1037/h0076255

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychological Review, 108(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

Ohman, A., Carlsson, K., Lundqvist, D., & Ingvar, M. (2007). On the unconscious subcortical origin of human fear. Physiology and Behavior, 92(1–2), 180–185. https://doi.org/10.1016/j.physbeh.2007.05.057

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion Drives Attention: Detecting the Snake in the Grass. Journal of Experimental Psychology: General, 130(3), 466–478. https://doi.org/10.1037/AXJ96-3445.130.3.466

Öhman, A., & Soares, J. J. F. (1998). Emotional Conditioning to Masked Stimuli: Expectancies for Aversive Outcomes Following Nonrecognized Fear-Relevant Stimuli. Journal of Experimental Psychology: General, 127(1), 69–82. https://doi.org/10.1037/0096-3445.127.1.69

Oliva, A., & Schyns, P. G. (2000). Diagnostic Colors Mediate Scene Recognition. Cognitive Psychology, 41(2), 176–210. https://doi.org/10.1006/cogp.1999.0728

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. Biological Psychology, 77(3), 247–65. https://doi.org/10.1016/j.biopsycho.2007.11.006

Ongür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cerebral Cortex, 10(3), 206–219. https://doi.org/10.1093/cercor/10.3.206

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Frontiers in Neuroinformatics, 10(July), 1–27. https://doi.org/10.3389/fninf.2016.00027

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. NeuroImage, 47(4), 1747–56. https://doi.org/10.1016/j.neuroimage.2009.04.058

Pascual-leone, A., & Hamilton, R. (2001). The metamodal organization of the brain.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal Representations of Perceived Emotions in the Human Brain. Journal of Neuroscience, 30(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Peelen, M. V., Atkinson, A. P., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. Social Cognitive and Affective Neuroscience, 2(4), 274–283. https://doi.org/10.1093/scan/nsm023

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool Selectivity in Left Occipitotemporal Cortex. Journal of Cognitive Neuroscience, 25(8), 1225–1234. https://doi.org/10.1162/jocn

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. Neuropsychologia, 105(October), 177–183. https://doi.org/10.1016/j.neuropsychologia.2017.03.033

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. Proceedings of the National Academy of Sciences, 99(17), 11458–11463. https://doi.org/10.1073/pnas.172403899

Pessoa, L. (2009). How do emotion and motivation direct executive control? Trends in Cognitive Sciences, 13(4), 160–166. https://doi.org/10.1016/j.tics.2009.01.006

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a “low road” to “many roads” of evaluating biological significance. Nature Reviews. Neuroscience, 11(11), 773–83. https://doi.org/10.1038/nrn2920

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. NeuroImage, 16(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. C., Cohen, L., Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway Pietro. Proceedings of the National Academy of Sciences, 101(15), 5658–5663. https://doi.org/10.1073/pnas.92.18.8135

Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., De Volder, A. G. (2006). Auditory motion perception activates visual motion areas in early blind subjects. NeuroImage, 31(1), 279–285. https://doi.org/10.1016/j.neuroimage.2005.11.036

Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint . . . International Journal of Computer Vision, 40(1), 49--71.

Potter, M. C. (1975). Short-Term Conceptual Memory for Pictures. Journal of Experimental Psychology: Human Learning and Memory, 2(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., & Levy, E. I. (1969). Recognition Memory for a Rapid Sequence of Pictures. Journal of Experimental Psychology, 81(1), 10–15. https://doi.org/10.1037/h0027470

Potter, M. C., Staub, A., Rado, J., & O’Connor, D. H. (2002). Recognition Memory for Briefly Presented Pictures: The Time Course of Rapid Forgetting. Journal of Experimental Psychology: Human Perception and Performance, 28(5), 1163–1175. https://doi.org/10.1037/0096-1523.28.5.1163

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. Biological Psychology, 92(3), 492–512. https://doi.org/10.1016/j.biopsycho.2012.02.007

Pourtois, G., Spinelli, L., Seeck, M., & Vuilleumier, P. (2010). Modulation of face processing by emotional expression and gaze direction during intracranial recordings in right fusiform cortex. Journal of Cognitive Neuroscience, 22(9), 2086–2107. https://doi.org/10.1162/jocn.2009.21404

Pourtois, G., Thut, G., De Peralta, R. G., Michel, C., & Vuilleumier, P. (2005). Two electrophysiological stages of spatial orienting towards fearful faces: Early temporo-parietal activation preceding gain control in extrastriate visual cortex. NeuroImage, 26(1), 149–163. https://doi.org/10.1016/j.neuroimage.2005.01.015

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage, 59(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. Journal of Cognitive Neuroscience, 28(5), 680–692. https://doi.org/10.1162/jocn

Ptito, M., Matteau, I., Gjedde, A., & Kupers, R. (2009). Recruitment of the middle temporal area by tactile motion in congenital blindness. NeuroReport, 20(6), 543–547. https://doi.org/10.1097/WNR.0b013e3283279909

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. Current Biology, 21(5), 363–368. https://doi.org/10.1016/j.cub.2011.01.040

Rempel-clower, N. L., & Barbas, H. (2000). Rempel-clower 2000 frl-tmpl, 851–865.

Ricciardi, E., Bonino, D., Gentili, C., Sani, L., Pietrini, P., & Vecchi, T. (2006). Neural correlates of spatial working memory in humans: A functional magnetic resonance imaging study comparing visual and tactile processes. Neuroscience, 139(1), 339–349. https://doi.org/10.1016/j.neuroscience.2005.08.045

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? Neuroscience and Biobehavioral Reviews, 41, 64–77. https://doi.org/10.1016/j.neubiorev.2013.10.006

Richards, A., & Blanchette, I. (2004). Independent manipulation of emotion in an emotional stroop task using classical conditioning. Emotion, 4(3), 275–281. https://doi.org/10.1037/1528-3542.4.3.275

Rolls, E. T. (1999). The brain and emotion. Oxford, England: Oxford University.

Rolls, E. T., Tovée, M. J., & Panzeri, S. (1999). Neurophysiology of Backward Visual Masking: Information Analysis. Journal of Cognitive Neuroscience, 11(3), 300–311.

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. NeuroImage, 122, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? Visual Cognition, 12(6), 852–877. https://doi.org/10.1080/13506280444000553

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. NeuroImage, 24(4), 1265–1270. https://doi.org/10.1016/j.neuroimage.2004.12.015

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., Jeffries, J. (2011). Emotional perception: meta-analyses of face and natural scene processing. NeuroImage, 54(3), 2524–33. https://doi.org/10.1016/j.neuroimage.2010.10.011

Sabatinelli, D., Lang, P. J., Bradley, M. M., Costa, V. D., & Keil, A. (2009). The timing of emotional discrimination in human amygdala and ventral visual cortex. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 29(47), 14864–8. https://doi.org/10.1523/JNEUROSCI.3278-09.2009

Sabatinelli, D., Lang, P. J., Keil, A., & Bradley, M. M. (2007). Emotional perception: correlation of functional MRI and event-related potentials. Cerebral Cortex (New York, N.Y. : 1991), 17(5), 1085–91. https://doi.org/10.1093/cercor/bhl017

Seligman, M. E. P. (1971). Phobias and preparedness. Behavior Therapy, 2(3), 307–320. https://doi.org/10.1016/S0005-7894(71)80064-3

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2013). The Animacy Continuum in the Human Ventral Vision Pathway. Journal of Cognitive Neuroscience, 27(4), 665–678. https://doi.org/10.1162/jocn

Shenhav, A., Barrett, L. F., & Bar, M. (2013). Affective value and associative processing share a cortical substrate. Cognitive, Affective and Behavioral Neuroscience, 13(1), 46–59. https://doi.org/10.3758/s13415-012-0128-4

Simmons, W. K., & Martin, A. (2012). Spontaneous resting-state BOLD fluctuations reveal persistent domain-specific neural networks. Social Cognitive and Affective Neuroscience, 7(4), 467–475. https://doi.org/10.1093/scan/nsr018

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. Psychonomic Bulletin and Review, 1–19. https://doi.org/10.3758/s13423-018-1451-8

Storbeck, J., & Clore, G. L. (2007). On the interdependence of cognition and emotion. Cognition and Emotion, 21(6), 1212–1237. https://doi.org/10.1080/02699930701438020

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “Seeing” by using sounds. Current Biology, 24(6), 687–692. https://doi.org/10.1016/j.cub.2014.02.010

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. Neuron, 76(3), 640–652. https://doi.org/10.1016/j.neuron.2012.08.026

Striem-Amit, E., Vannuscorps, G., & Caramazza, A. (2017). Sensorimotor-independent development of hands and tools selectivity in the visual cortex. Proceedings of the National Academy of Sciences, 114(18), 4787–4792. https://doi.org/10.1073/pnas.1620289114

Tabert, M. H., Borod, J. C., Tang, C. Y., Lange, G., Wei, T. C., Johnson, R., Buchsbaum, M. S. (2001). Differential amygdala activation during emotional decision and recognition memory tasks using unpleasant words : an fMRI study. Neuropsychologia, 39, 556–573.

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. Nature Reviews. Neuroscience, 11, 697–709. https://doi.org/10.1038/nrn2889

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature. https://doi.org/10.1038/381520a0

Torry, W. I. (2011). Residence Rules Among the Gabra Nomads : Some Ecological Considerations ', 15(3), 269–285. https://doi.org/10.2307/3773135

Tovee, M. J., & Rolls, E. T. (1995). Information Encoding in Short Firing Rate Epochs by Single Neurons in the Primate Temporal Visual Cortex. Visual Cognition, 2(1), 35–58. https://doi.org/10.1080/13506289508401721

Tranel, D., Grabowski, T. J., Lyon, J., & Damasio, H. (2005). Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. Journal of Cognitive Neuroscience, 17(8), 1293–1305. https://doi.org/10.1002/ptr.2454

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage, 15(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. Proceedings of the National Academy of Sciences, 114(22), E4501–E4510. https://doi.org/10.1073/pnas.1612862114

van Steenbergen, H., Band, G. P. H., & Hommel, B. (2011). Threat but not arousal narrows attention: Evidence from pupil dilation and saccade control. Frontiers in Psychology, 2(OCT), 1–5. https://doi.org/10.3389/fpsyg.2011.00281

Vanlierde, A., De Volder, A. G., Wanet-Defalque, M. C., & Veraart, C. (2003). Occipito-parietal cortex activation during visuo-spatial imagery in early blind humans. NeuroImage, 19(3), 698–709. https://doi.org/10.1016/S1053-8119(03)00153-8

Vanni, S., Revonsuo, A., Saarinen, J., & Hari, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. NeuroReport, 8(1), 183–186. https://doi.org/10.1097/00001756-199612200-00037

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. Journal of Cognitive Neuroscience, 13(4), 454–461. https://doi.org/10.1162/08989290152001880

Verstynen, T. D., & Deshpande, V. (2011). Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. NeuroImage, 55(4), 1633–44. https://doi.org/10.1016/j.neuroimage.2010.11.090

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron, 30(3), 829–841. https://doi.org/10.1016/S0896-6273(01)00328-2

Vuilleumier, P., Armony, J. L., & Dolan, R. J. (2003). Reciprocal Links between Emotion and Attention. Human Brain Function: Second Edition, 419–444. https://doi.org/10.1016/B978-012264841-0/50022-6

Vuilleumier, P. (2015). Affective and motivational control of vision. Current Opinion in Neurology, 28, 29–35. https://doi.org/10.1097/WCO.0000000000000159

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. Trends in Cognitive Sciences, 9(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. Nature Neuroscience, 7(11), 1271–1278. https://doi.org/10.1038/nn1341

Wang, S., Tsuchiya, N., New, J., Hurlemann, R., & Adolphs, R. (2014). Preferential attention to animals and people is independent of the amygdala. Social Cognitive and Affective Neuroscience. https://doi.org/10.1093/scan/nsu065

Watson, D. M., Young, A. W., & Andrews, T. J. (2016). Spatial properties of objects predict patterns of neural response in the ventral visual pathway. NeuroImage, 126, 173–183. https://doi.org/10.1016/j.neuroimage.2015.11.043

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., Rauschecker, J. P. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. The Journal Of Neuroscience: The Official Journal Of The Society For Neuroscience, 20(7), 2664–2672.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. Current Biology, 21(11), 984–989. https://doi.org/10.1016/j.cub.2011.04.038

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. Emotion (Washington, D.C.), 7(4), 882–886. https://doi.org/10.1037/1528-3542.7.4.882

Yao, B., Yang, X., & Wu, T. (2009). Image Parsing with Stochastic Grammar: The Lotus Hill Dataset and Inference Scheme. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (p. 8). Miami, Florida.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151