Domain-Specific Tasks Differentiate Cortical Areas with Similar Sensory Responses

Logan Dowdle[1*], Geoffrey Ghose[1], Kamil Ugurbil[1], Essa Yacoub[1], Luca Vizioli[1*]

1. Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN

* Indicates Corresponding Authors

Corresponding Authors:

Logan Dowdle
Postdoctoral Associate
Center for Magnetic Resonance Research
Department of Radiology
University of Minnesota
2021 6th street SE,
55455 Minneapolis, MN,
United States
Correspondence to:
logan.dowdle@gmail.com
dowdl016@umn.edu

Luca Vizioli
Assistant Professor
Center for Magnetic Resonance Research
Department of Radiology
University of Minnesota
2021 6th street SE,
55455 Minneapolis, MN,
United States
Correspondence to:
luca.vizioli1@gmail.com
lvizioli@umn.edu

Summary

The brain is organized into distinct, but not rigid, networks. Sensory systems, for example, are enormously flexible, with cognitive variables such as attention, value, and memory able to alter sensory representations in accordance with moment-to-moment behavioral demands. Understanding these top-down modulations requires several experimental considerations. First, to understand network function, task demands must be congruent with the domain to which a network is sensitive. Second, bottom-up, sensory input must remain identical to isolate top-down effects. By combining high SNR functional imaging with the presentation of identical face stimuli under varying task demands, we have demonstrated the presence of novel face-relevant signals in regions homologous to those in non-human primate. Critically, one task was domain-specific, requiring that participants report face perception. The anterior inferior temporal region, which has historically proved difficult to resolve in humans, is both preferentially activated under challenging conditions and uniquely encodes the trial-by-trial variability of perception.

1. Introduction

Decades of human and animal research has demonstrated that cortical areas are functionally and anatomically linked to form distinct brain networks. Understanding the distinct contributions of individual areas within these networks has been challenging in part because of the rich and reciprocal pattern of connections between brain areas (Alexander et al., 1986; Felleman and Van Essen, 1991; Moeller et al., 2008). Discrimination is further confounded by the low latencies with which signals travel between areas (Laughlin and Sejnowski, 2003; Wang et al., 2008), substantial overlaps in the functional sensitivities of neurons of different areas (Arcaro and Livingstone, 2017; Haxby et al., 2001; Vogels and Orban, 1994), and the flexibility of connections necessary to maintain function across a range of tasks and environments (Bassett et al., 2011; Gonzalez-Castillo and Bandettini, 2018; Kabbara et al., 2017). A critical component of this flexibility in sensory systems is top-down control, in which cognitive variables such as attention, value, and memory can alter sensory representations in accordance with behavioral demands. Yet this flexibility also is a major challenge to experimental studies, because it implies that the nature of signal processing both within and between brain areas of a network is not fixed and very much depends on the behavioral context and goals of the subject.

While many humans and animal studies have demonstrated the ability of task manipulations to modulate sensory evoked responses, in most cases these studies have employed tasks that are not specific to the domain of the stimulus that was presented. For example, numerous electrophysiological and imaging studies suggest that attending to a particular location in visual space increases the responses of neurons whose receptive fields lie within the attended location irrespective of the selectivities of those receptive fields (Cohen and Maunsell, 2011; Liu et al., 2015; Martinez-Trujillo and Treue, 2004). Similarly, perceptual tasks involving working memory often result in a broad distribution of enhanced signals in sensory areas (Druzgal and D'Esposito, 2001, 2003; Kay and Yeatman, 2017; Leung and Alain, 2011; Pessoa et al., 2002). By contrast, top down modulation from domain-specific tasks can be far more targeted, selectively enhancing only those neurons whose receptive field responses are consistent with behavioral demands (De Martino et al., 2015; Zhang and Kay, 2018). Accordingly, the use of appropriate domain specific tasks is likely to highlight those neurons with relevant selectivities and reveal how they contribute to perception. For example, contrasting the activity elicited by identical face stimuli in

an N-back and a fixation task may reveal working memory related modulations that are unspecific to face processing and therefore fail to elucidate the changes related to the specialized processing of faces.

Establishing how different areas within a network contribute to perception requires sampling the entire network and then identifying the particular areas that can best explain task performance. However, technical limitations have precluded this approach. Invasive electrophysiological studies are necessarily highly targeted and unable to sample across an entire brain network. In addition, these techniques are frequently performed in non-human primates, with tasks that fail to capture the cognitive sophistication available to humans. Moreover, even whole brain studies, such as those relying upon functional magnetic resonance imaging, are often highly biased in their sampling because of the difficulty of imaging particular areas. For example, a number of high-level areas – crucial for human cognition – are located in cortical loci that are traditionally difficult to image (e.g. due to low SNR and their proximity to air cavities; N. K. Chen et al., 2003; Devlin et al., 2000). These limitations also constrain our ability to establish appropriate animal models for invasive measurements and manipulations because of the possibility that human areas homologous to those in animals studies cannot be sampled by fMRI (Rajimehr et al., 2009). Signal to noise issues have also limited the ability of fMRI to establish trial-to-trial behavioral covariance, which is a critical requirement for understanding the neural basis of perception (Parker and Newsome, 1998) and has been extensively employed by electrophysiology studies (Desimone et al., 1984; Hubel and Wiesel, 1959; Martinez-Trujillo and Treue, 2004; Perrett et al., 1982). This is a particularly important consideration because, by definition, all areas within a network are co-activated, but, given the richness of connectivity within the network, that co-activation does not necessarily imply perceptual relevance.

These concerns can be mitigated with ultra-high field magnets (UHF i.e. 7T and above), which yield significantly higher SNR characteristics, however researchers generally trade these SNR gains for higher resolutions (e.g. submillimeter). In this paper we instead select a lower resolution (i.e. 1.6 mm^3) and leverage the improved SNR available in UHF fMRI to study the entire ensemble of brain areas associated with a behaviorally important high-level perceptual task, namely the detection of faces in poor visibility. We choose faces as these are a well-studied, cross species visual category that is highly meaningful and therefore ideal for investigating high-level, top-down neural modulations. The human face processing network has however proved challenging to study (Kanwisher and Yovel, 2006; Kanwisher et al., 1997; McCarthy et al., 1997)(Review: Grill-Spector & Weiner, 2014), because many of its areas, due to their proximity to air cavities, are heavily affected

by the aforementioned fMRI signal difficulties. Moreover, domain-specific attentional tasks, such as those focused on high-level face analysis, which are most likely to highlight the distinctive role of particular neuronal populations have seldom been employed. Specifically, we varied the visibility of face stimuli by modulating their phase coherence and instructed the subjects to perform 2 tasks in the scanner: a *domain-specific, stimulus relevant* task Involving perceptual judgment of the visual stimuli (i.e. face detection); *and, stimulus irrelevant* fixation task.

Combining high SNR fMRI with domain-specific attention, we demonstrate that the presence of novel face-relevant signals in regions homologous to those seen in non-human primates. Consistent with the importance of task, we find that functional connectivity within this expanded network changes when subjects are performing a domain-specific task. Associated with those changes in functional connectivity, a particular area, with strong homologies to the monkey face network, but which has not previously been resolvable in humans, is both preferentially activated under challenging conditions and predictive of the subjects' detection judgments.

2. Materials and Methods

2.1 Participants

10 (5 females) healthy right-handed subjects (age range: 18-31) participated in the study. Of these, 1 participant was re-scanned due to excessive motion during scanning. All subjects had normal, or corrected vision and provided written informed consent. The local IRB at the University of Minnesota approved the experiments.

2.2 Stimuli and procedure

The experimental procedure consisted of a standard block design face localizer and a fast event-related face paradigm. For both experiments, the stimuli were centered on background of average luminance (25.4 cd/m2, 23.5-30.1). Stimuli were presented on a Cambridge Research Systems BOLDscreen 32 LCD monitor positioned at the head of the 7T scanner bed (resolution 1920, 1080 at 120 Hz; viewing distance ~89.5 cm) using a Mac Pro computer. Stimulus presentation was controlled using Psychophysics Toolbox (3.0.15) based code. Participants viewed the images though a mirror placed in the head coil. Behavioral responses were recorded using Cambridge Research Systems button box and Psychophysics Toolbox.

2.3 Face localizer

2.3.1 Stimuli.

All visual stimuli used for the face localizer consisted of grayscale photographs depicting 20 different faces (10 identities × 2 genders) and objects (both taken from Stigliani et al., 2015) and textures of noise.

The experiment consisted of the presentation of grayscale stimuli drawn from different stimulus categories. There were 11 categories, grouped into 3 stimulus domains: faces (adult, child), no face objects (including characters (word, number), body parts (body, limb), places (corridor, house), and objects (car, instrument)) and phase scrambled faces, computed by randomly scrambling the phase coherence of the face stimuli. Each stimulus was presented on a scrambled background (different backgrounds for different stimuli) and occupied a square region with dimensions 10° × 10°. Noise texture stimuli were created by randomly scrambling the phase of the face images (i.e. 0% phase coherence).

The stimuli were equated in terms of luminance, contrast and spatial frequency content by taking the Fourier spectra across stimuli and ensuring that the rotational average amplitudes for a given spatial

frequency were equated across images while preserving the amplitude distribution across orientations (Willenbockel et al., 2010). The root mean square contrast (i.e. the standard deviation of pixel intensity) was also kept constant across stimuli.

2.3.2 Visual presentation paradigm.

Face localizer runs involved presentation of blocks of faces, objects and noise textures. Each run began with presentation of a black fixation cross displayed on a grey background for 12 sec and consisted of 9 randomly presented blocks of images. Each block (3 blocks/category; separated by a 12 sec fixation) involved presentation of 10 different stimuli randomly presented for 800 ms, separated by a 400 ms interstimulus interval (ISI). To ensure that participants' attention was maintained throughout the localizer we implemented a 1-back task, where subjects were instructed to respond to the repetition of 2 identical stimuli (representing about 10% of the trials), by pressing a button on a response pad held in their right hand. All participants completed two runs of the block design face localizer, where each block occurred 3 times within a run, for a total run duration of 228 seconds.

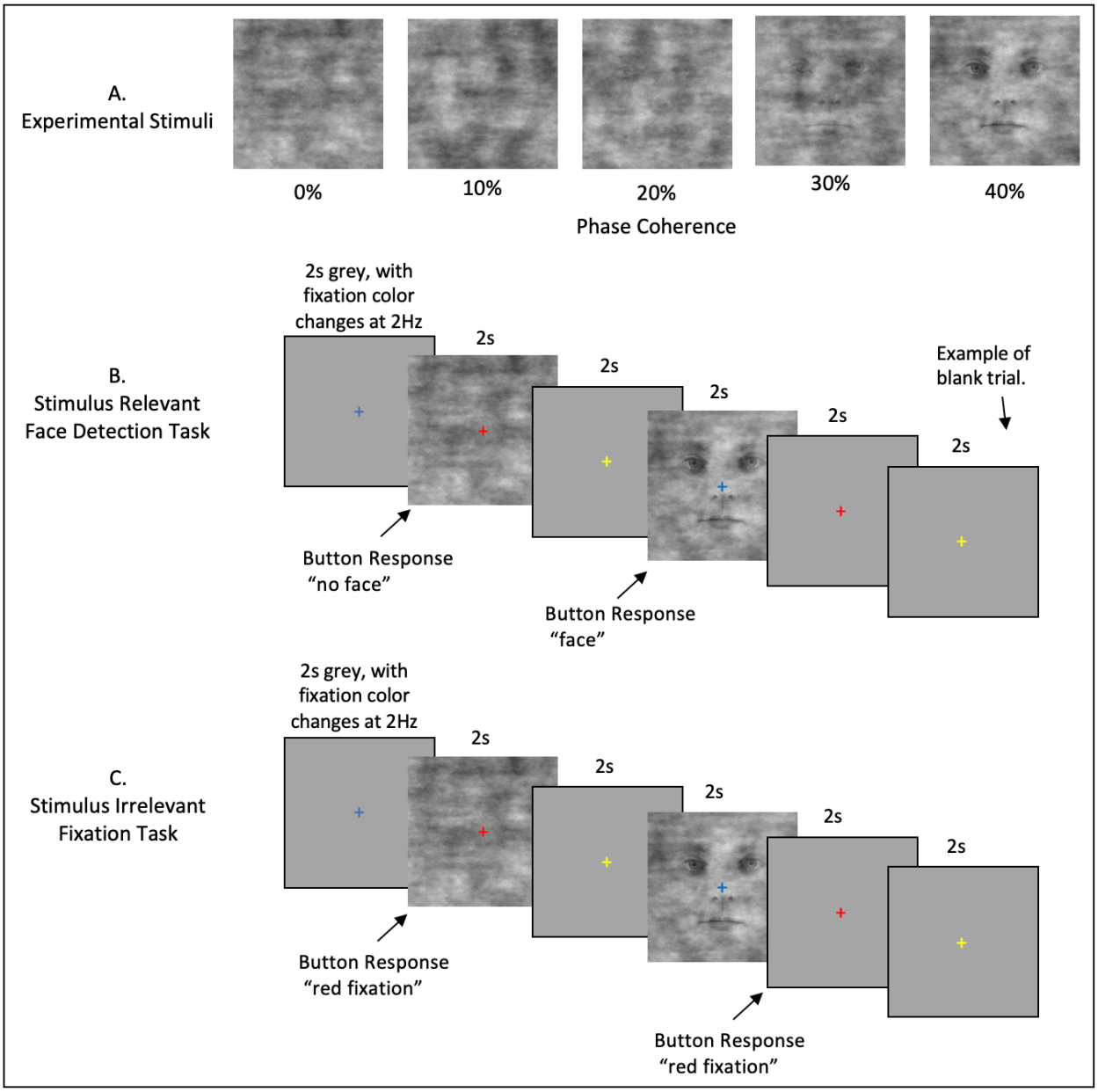
2.4 Event related experiment.

2.4.1 Stimuli.

We used grayscale images of faces (20 male and 20 female). We manipulated the phase coherence of each face, from 0% to 40% in steps of 10%, resulting in 200 images (5 visual conditions x 20 identities x 2 genders). We equated the amplitude spectrum across all images. Stimuli approximately subtended 9° of visual angle. Faces were cropped to remove external features by centering an elliptical window with uniform gray background to the original images. The y diameter of the ellipse spanned the full vertical extent of the face stimuli and the x diameter spanned 80% of the horizontal extent. Before applying the elliptical window to all face images, we smoothed the edge of the ellipse by convolving with an average filter (constructed using the "fspecial" function with "average" option in MATLAB – see Figure 1A). This procedure was implemented to prevent participants from performing edge detection, rather than the face detection task, by reacting to the easily identifiable presence of hard edges in the face images.

Like for the localizer experiment, here too we equated amplitude spectrum across the whole stimulus set. We controlled the Fourier spectra across stimuli, ensuring that the rotational average amplitudes for a given spatial frequency were equated across images while preserving the amplitude distribution across orientations

(Willenbockel et al., 2010). The root mean square contrast (i.e. the standard deviation of pixel intensity) was also kept constant across stimuli.



Caption Figure 1. Event Related Stimuli and Tasks. **A)** Example stimuli associated with the 5 phase coherence levels used. **B)** The Stimulus Relevant Face detection task. Stimuli appeared for 2 seconds, with an intertrial interval of 2 seconds. The fixation cross changed color with a frequency of 2Hz and was visible throughout all experimental procedures. Participants indicated "face" or "no face" using an MRI response hand pad. Blank trials were randomly inserted during each fMRI run. **C)** The Stimulus Irrelevant Fixation task. Timing is identical to B. Participants indicated when the fixation cross changed to the color red.

2.4.2 Tasks.

In the scanner, participants were instructed to maintain fixation on a central cross throughout the run and to perform one of 2 tasks: one domain-specific face detection task, involving perceptual judgment of the visual stimuli; and a second, difficult, non-specific attention fixation task that required responding to a specific color change (i.e. red) of the fixation cross.

In the former, participants were instructed to respond as quickly as possible by pressing one of 2 buttons on their button box to indicate whether they perceived a face. Subjects' instructions were carefully delivered to indicate that there were no correct answers and that we were instead interested in the subjects' perception only (Figure 1B).

The latter was designed to isolate bottom-up stimulus-driven responses. In order to ensure this, we piloted the fixation task to maximize task difficulty and direct the attention away from the face stimuli on 4 participants that were not included in the experimental subject pool. Based on participants' feedback, the fixation tasks entailed pressing a button every time the fixation turned red. Every 500 ms, the fixation changed to one of five colors — specifically red, blue, green, yellow and cyan — in a pseudorandom fashion, avoiding consecutive presentations of the same color (Figure 1C). The color change occurred out of sync with the stimulus presentation onset and frequency of button presses was kept constant across tasks. Visual stimuli were identical across tasks to ensure that any differences observed were related solely to top-down processes. Tasks were blocked by run and counterbalanced across participants.

2.4.3 Visual presentation paradigm.

We acquired 3 runs per task. Each run lasted approximately 3 mins and 22 secs and began and ended with a 12-second fixation period. Within each run, we showed 40 images (5 phase coherence levels x 4 identities x 2 genders) presented for 2000 ms, with a 2000 ms interstimulus interval (ISI). Importantly, we introduced 10% blank trials (i.e. 4000 ms of fixation period) randomly interspersed amongst the 40 images, effectively jittering the ISI. Stimulus presentation was pseudorandomized across runs, with the only constraint being the non-occurrence of 2 consecutive presentations of the same phase coherence level. Behavioral metrics, including reaction time and responses to face stimuli indicating participants' perceptual judgments (i.e. face or no face) were generated per subject and then averaged.

2.4.4 Ambiguity Calculation

For the face detection task, we calculated the average face response percentage to each phase condition, averaged across runs. These values vary from 0 to 100%, representing consistent non-face to consistent face response respectively. Within each subject, this produces a sigmoidal shaped curve. We mathematically define the ambiguity score, on a subject by subject basis, as the inverse of the absolute distance from the inflection point on the sigmoid that fit this curve This is shown in Equation 1:

$$\text{Eq 1. } (|S_i - S_k| \times -1) + \min(A)$$

where S_i is the *i*th point of the faceness sigmoid, S_k is the theoretical midpoint of the faceness sigmoid and min(A) is the minimum value of the ambiguity function.

2.5 MR Imaging Acquisition and Processing

All functional MRI data were collected with a 7T Siemens Magnetom System with a 1 by 32-channel NOVA head coil. T2*-weighted images were collected using sequence parameters (TR 1.6s, Multiband 5, GRAPPA 2 , 7/8ths Partial Fourier, 1.6mm isotropic, TE 22ms, Flip Angle 60°, Bandwidth 1923Hz) matched to the Human Connectome 7T protocol (Thanh Vu et al., 2017).

For each participant we manually shimmed the B0 field to maximize homogeneity over ventral and anterior-temporal regions. Moreover, we optimized flip angles across the brain to maximize SNR.

T1-weighted anatomical images were obtained using an MPRAGE sequence (192 slices; TR, 1900 ms; FOV, 256 x 256 mm; flip angle, 9°; TE, 2.52 ms; spatial resolution, .8 mm isotropic voxels) which were collected with a 3T Siemens Magnetom Prisma^{fit} system. Anatomical images were used for visualization purposes only.

2.5.1 Functional Image Processing

Dicom files were converted using dcm2niix (Li et al., 2016). Subsequence functional image processing was performed in AFNI version 19.2.10 (Cox, 1996). Conventional processing steps were used, including despiking, slice timing correction, distortion and motion correction, and alignment to each participant's anatomical image. With each task type, the target for time series and anatomical alignment was the Single Band Reference (SBRef) image which is acquired to calibrate coil sensitively profiles prior to the multiband acquisition and has no slice acceleration or T_1-saturation, yielding high contrast (Smith et al., 2013). In order to improve

localizer and task alignment, an additional nonlinear transform was computed between the task SBRef and localizer SBRef. All spatial transforms were concatenated and applied in a single step to reduce image blurring, and functional images were inspected to confirm successful registration to anatomical targets.

2.5.2 Localizer Task Analyses

For the functional localizer, the data then smoothed with a Gaussian kernel with FWHM of 2 voxels (3.2mm), and each run scaled to have mean 100. Nuisance regressors included high-pass filtered using a Fourier basis set of three cycles per run (including linear trend) and the 6 estimated rigid-body motion parameters.

2.5.3 Functional ROI Definition

Using 3dDeconvolve, the data were passed through a GLM in order to determine the response to faces, objects and noise. Each event was modeled as a 12s box car, convolved with AFNI's SPMG1 HRF estimation. Rather than select face patches by comparing face activation to the average of objects and noise, we instead constrained the statistical map in the following ways: 1) Betas to faces were positive, 2) the T-stat used for thresholding was the minimum positive T-stat from faces > objects or faces > noise. In other words, we selected voxels that were significant for faces greater than objects, in conjunction with faces greater than scrambled. Regions of interest (ROIs) were derived in volume space, and all consisted of contiguous clusters made up of 20 or more voxels. Statistical thresholding was adjusted to obtain consistency between subjects, no voxels above $p<0.05$ were considered. In addition, to supplement the statistical parametric maps, we used finite impulse response (FIR) deconvolution, with 20 bins, to estimate HRFs in response to each localizer condition (face, scrambled, object). These HRFs were also used to inspect data quality for each ROI.

2.5.4 Face Detection and Fixation Color Tasks Analyses

For the Face/Fix detection tasks, the data were only scaled following initial processing, no smoothing was used. As we were interested in BOLD responses across the whole brain, and we know that hemodynamic response functions differ across cortical regions (Handwerker et al., 2004; Taylor et al., 2018), we performed a GLM based finite impulse response (FIR) deconvolution analyses, which does not make assumptions regarding the shape of the HRF. This was done using AFNI's TENTZero method, estimating responses with 1s bins out to 15s post stimulus. Deconvolution was performed separately for each task (face detection and fixation) and phase coherence condition, generating 5 FIR curves for the fixation task and 5 for the face detection task. For each

task and independently per subject, ROI, run and condition, we averaged all FIR curves across all voxels. To avoid the contribution of extreme voxels, we trimmed 5% of the values falling at each extreme of the distribution tail to compute the 10% trimmed mean. We then extracted the 4 timepoints corresponding to 4, 5, 6 and 7 seconds after stimulus onset, i.e. those with the largest amplitude within a time window spanning from 2 to 10 TRs after stimulus onset. The 12 extracted amplitudes (4 timepoints per run) were then averaged within experimental tasks to obtain one percent signal change value per subject, ROI, task and condition.

2.6 Statistical Analyses

To test for statistically significant differences between conditions (i.e. phase coherence levels) and tasks, we carried out the following statistical tests.

2.6.1 Task and Condition Analyses

Independently per ROI, we performed a 2 (tasks) by 5 (phase coherence levels) repeated measures ANOVA with the mean BOLD response as a dependent variable. To control for family wise error rate (FWER), we implemented the following multiple comparison correction procedure: we began by centering the data on the group mean for each condition and task. This procedure effectively put the data under the ideal H0 hypothesis of no difference between the means. We then sampled with replacement the subjects and performed the same 2 x 3 repeated measures ANOVA and stored all F values for the relevant main effects and interactions. We repeated this procedure 10,000 times and selected the 95% largest F values. We used these F values as our new thresholds and considered statistical significance only when p values for the original ANOVA were smaller than .05 *and* the connected F values were larger than their centered counterpart (e.g. (Wilcox, 2005))

When appropriate (i.e. for comparisons entailing more than 2 factors), we further performed post-hoc paired sample t-tests on significant (as defined above) main effects and interactions. The same FWER control procedure described above for the ANOVA test was implemented to account for multiple comparisons.

Additionally, we performed power analyses to determine effect size and the sample size required to achieve adequate power (see results). Specifically, we computed Hedges' g (Hedges, 1981; Hedges et al., 1985). This choice was motivated by the fact that, unlike Cohen's d, which, especially for small samples (i.e. $n < 20$), tends to provide positively biased estimators of population effect sizes (Hedges et al., 1985; Lakens, 2013), Hedges' g tends to be unbiased (Cumming, 2012).

2.6.2 Functional connectivity analyses.

To determine the extent to which task demands modulate functional connectivity amongst face ROIs, we carried out the following analysis. Independently per subject, task and ROI, we averaged (mean) all FIR response curves amongst voxels and concatenated the time courses elicited by each condition to produce a single time course. The concatenation was done to maximize statistical power, as the resulting time course contained 80 timepoints (i.e. 16 FIR timepoints x 5 conditions). For each subject and task, we then computed Pearson's correlation coefficient between the concatenated time courses of all possible pairs of ROIs. This procedure lead to the formulation of 2 connectivity matrices (12 ROIs x 12 ROIs) per subject (1 per task) summarizing the similarity of BOLD responses between all pairs of ROIs. To convert the skewed sampling distribution of Pearson's r into a normal distribution we computed Fisher's z transformation (Fisher, 1915). We therefore proceeded to carry out paired sample t-tests between the connectivity estimated obtained for the 2 tasks. To control for FWER, we implemented the same bootstrap procedure on centered data described in the previous paragraphs (see paragraph 2.6.1). For display purposes only, after computing the mean between the Fisher z-normalized connectivity matrices, we computed the inverse of such transformation on the group average connectivity matrices to convert these scores back into meaningful and interpretable Pearson's r.

To better visualize the results of our functional connectivity analysis, we further performed classic multidimensional scaling (MDS - using the function "cmdscale" in MATLAB) on the participants average dissimilarity matrix (i.e. 1- Pearson's r). MDS was performed independently per task. For ease of visual comparison, the MDS arrangements of the 2 tasks were aligned by means of linear transformations (including translation, reflection, orthogonal rotation, and scaling) using Procrustes rotations (this was implemented with the "procrustes" function in MATLAB) with the sum of squared errors as stress metric.

MDS is a useful data driven approach to visualize the data projected into a new space whose dimensions are the first (in this case 2) eigenvectors (i.e. those explaining most of the variance in the data) without any prior hypotheses. MDS is therefore a dimension reduction technique that highlights the dominant features in the data. Results are shown in Figure 5; distance between the points indicate dissimilarity of responses.

2.6.3 Brain-behavior correlation.

Next for each ROI, we wanted to assess the relationship between behavioral responses and top-down BOLD modulations. To this end, for each subject and ROI, we computed Pearson's correlation coefficients amongst the ambiguity scores at each phase coherence level and the task difference in average BOLD amplitudes elicited by each condition. We then performed Fisher z transform (see paragraph above) on Pearson's r and carried out one sample t-test for each ROI to determine whether the average group correlation was significantly larger than 0. To control the family wise error rate (FWER), we implemented the same bootstrap procedure on centered data described in the previous paragraphs (see paragraph 2.6.1). For display purposes only, after computing the mean between the Fisher z-normalized correlation scores, we computed the inverse of such transformation on the group average to convert these scores back into meaningful and interpretable Pearson's r.

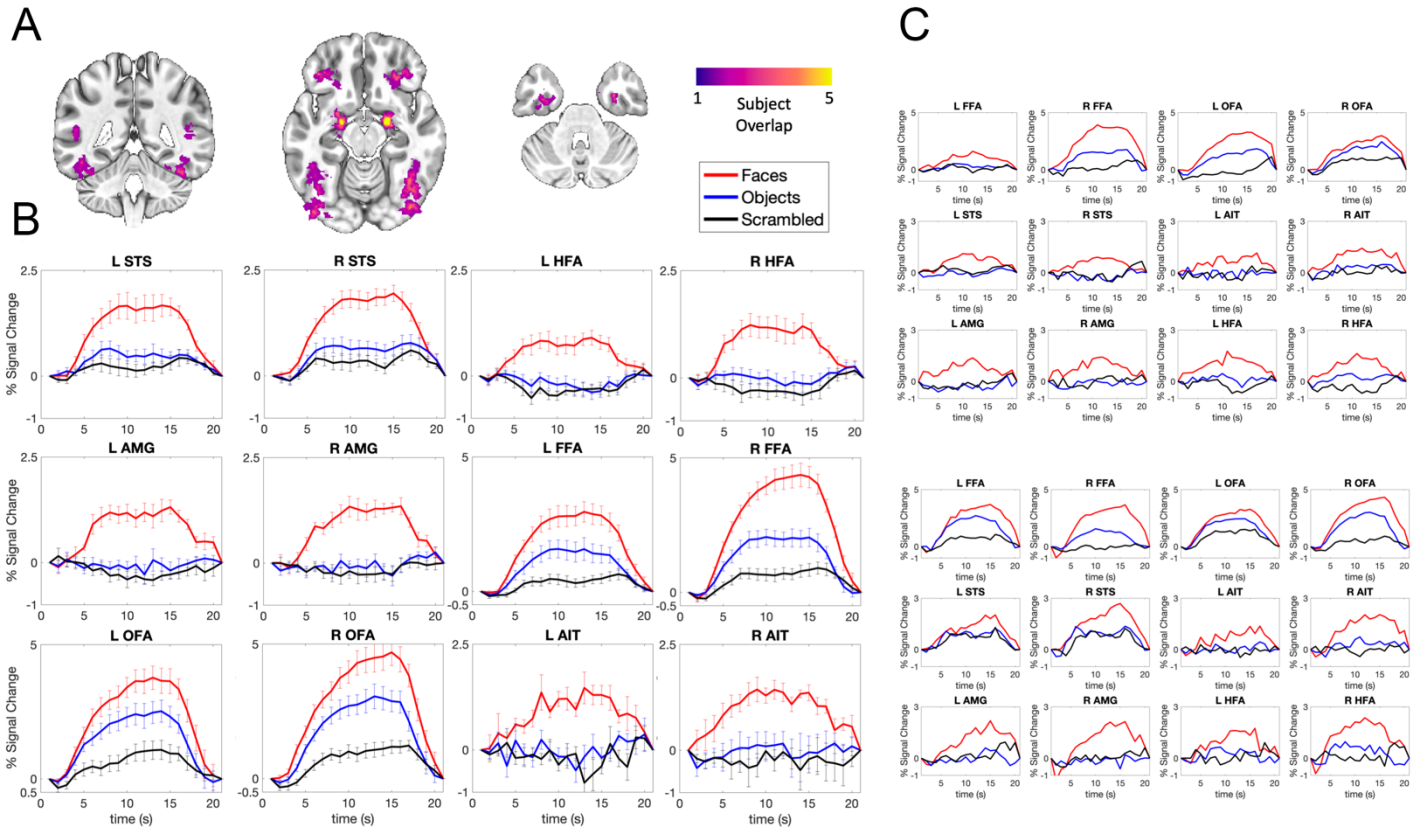
2.6.4 Perception-Based Analyses

We performed additional analyses to assess BOLD amplitude modulations as a function of percept. For the most ambiguous condition (i.e. 20% phase coherence, see results), stimulus grouping was determined on the basis of each participants reported percept. We allocated each trial to one of 2 new conditions: "*face percept*" and "*no face percept*". We then repeated the analysis described in section 2.5.4 to estimate BOLD amplitudes of these 2 conditions.

3. Results

3.1 Localizer Task.

Using the separate face-localizer task, we identified a total of twelve ROIs. This included typical regions, such as the fusiform face area (FFA), occipital face area (OFA), posterior superior temporal sulcus (STS) as well as the amygdalae. In addition, we identified ROIs for more difficult regions, including an ROI proximal to the perirhinal cortex, which we are referring to more generally as the anterior inferior temporal (AIT) cortex as well as an ROI in H-shaped sulcus, which we refer to as the HFA (H-shaped sulcus Face Area). ROIs were derived in subject space, but are combined in MNI space in Figure 2A. Due to subject specific anatomy and cortical folding, the relatively high resolution of the data and the small size of the ROIs, there is not perfect overlap, particularly for cortical ROIs. Despite this, all ROIs are in consistent locations (i.e. HFAs all fall within the H-shaped sulcus area).



Caption Figure 2. Results from the face preferential localizer. **A)** All subject's ROIs combined in MNI template space, highlighting the consistent overlap between subjects and between hemispheres. **B)** Average BOLD responses across all subjects, in units of percent signal change, for each ROI in response to faces, objects and

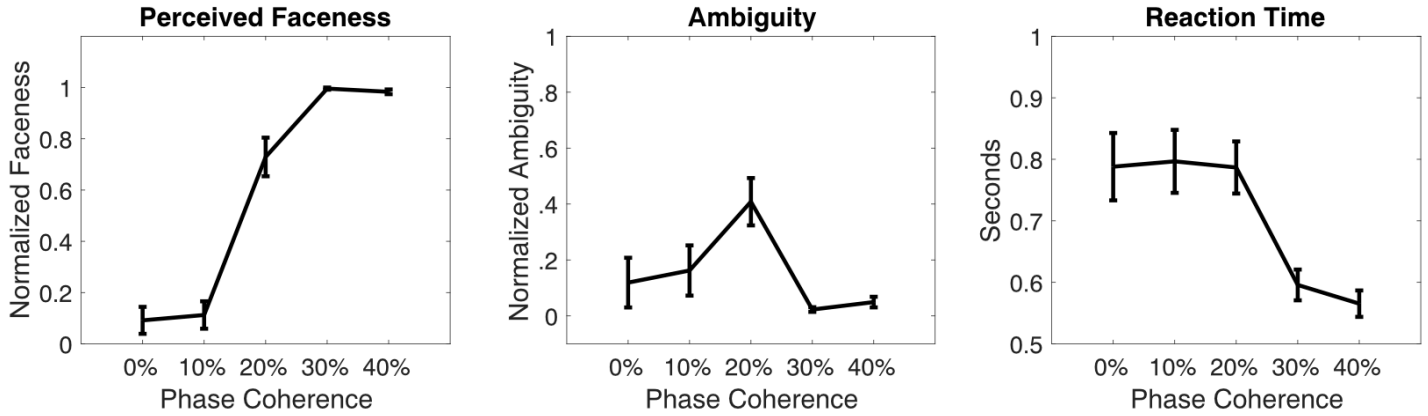
scrambled images. Standard error shown. **C)** Example results from two single subjects showcasing reliability even within a single subject's ROIs.

Figure 2B shows the average BOLD time courses, constructed from FIR models, for these ROIs. For all regions, including areas associated with low SNR (AIT, HFA), this averaging yields plausible hemodynamic responses. Furthermore, these responses remain HRF-like even at the single subject level (Figure 2C).

3.2 Behavioral Responses and Ambiguity Scores

The right panel in Figure 3 portrays the face detection group average reaction times. Subjects responded with an average RT of 760±120 ms. Reaction times indicate that, in reporting their percepts, participants were slowest at the lower end of the phase coherence spectrum (i.e. 0% and 10%), becoming increasingly faster as a function of phase coherence.

Figure 3, left panel, shows the average perceived "faceness", that is the proportion of the time that participants reported seeing a face, which produces a sigmoid shaped curve. This sigmoid shape was present for all subjects. By calculating the inverse of the absolute distance from the inflection point (see section 2.4.4.) we derive the ambiguity score (Figure 3, middle panel), which shows that the stimuli reached their maximum level of ambiguity at 20% phase coherence. Reaction times were shortest for the highest phase coherence (Figure 3, right panel).



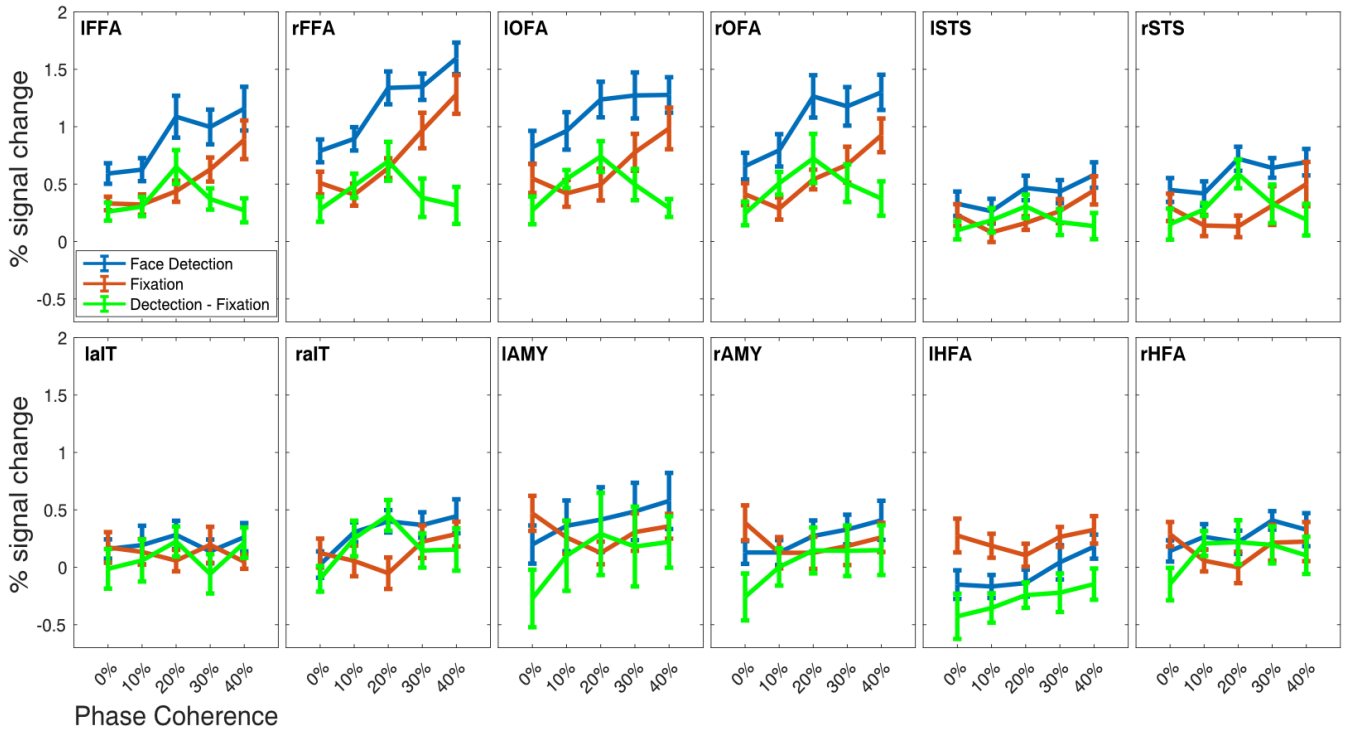
Caption Figure 3. Behavioral Responses to Stimulus Relevant Task. **Left panel:** The group average of perceived "faceness", where 0 represents consistently reporting no face, and 1 represents consistently reporting

the presence of a face. **Central Panel:** The group averaged ambiguity curve, showing that stimuli at 20% were most ambiguous, i.e., were most inconsistently categorized. **Right Panel:** Group averaged reaction times during the face detection task. For all panels, error bars represent standard errors across subjects.

3.3 Face and Fixation fMRI analyses

3.3.1 ANOVA Results

The percent signal change in response to each condition, for each ROI, is shown in Figure 4.



Caption Figure 4. Percent Signal change during event related tasks in all ROIs. In blue, the BOLD responses to the stimulus relevant, domain specific task, in red to the stimulus irrelevant task. For the majority of ROIs, the stimulus relevant BOLD responses are larger across all phase levels relative to the stimulus irrelevant task. The green curve represents the differences between tasks.

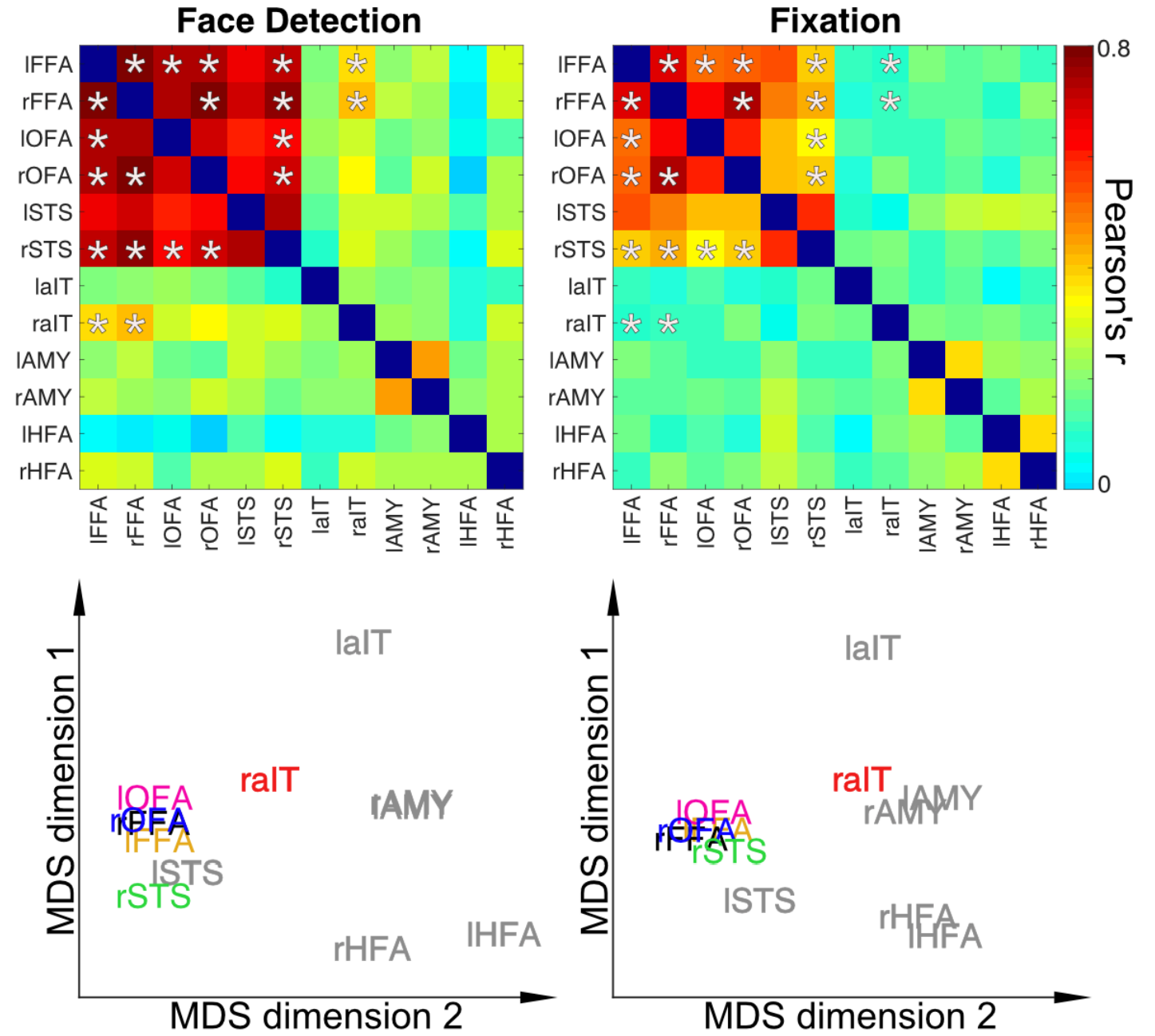
The main effect of the task was significant ($p < 0.05$) in the left FFA ($F_{1,9} = 20.559$), right FFA ($F_{1,9} = 13.131$), left OFA ($F_{1,9} = 48.595$), right OFA ($F_{1,9} = 14.304$) and the right STS ($F_{1,9} = 9.182$), with the face detection task driving larger BOLD responses relative to the fixation task.

There was a significant ($p < 0.05$) main effect of the condition (phase coherence level) in the left FFA ($F_{4,36} = 13.136$), right FFA ($F_{4,36} = 35.567$), left OFA ($F_{4,36} = 7.56$), right OFA ($F_{4,36} = 22.866$), left STS ($F_{4,36} = 7.98$), right STS ($F_{4,36} = 5.637$), the right AIT ($F_{4,36} = 4.174$). For these ROIs, paired sample post-hoc t-test ($p<.05$, corrected) showed that the 40% phase coherence is always significantly larger than 0%.

There was a significant ($p < 0.05$) task by condition interaction term in the left FFA ($F_{4,36} = 4.96$), right FFA ($F_{4,36} = 3.311$), the right STS ($F_{4,36} = 3.620$) and the right AIT ($F_{4,36} = 3.84$), indicating that amplitude increase during detection relatively to fixation was different for different phase coherence levels. Post-hoc t-tests carried out across tasks, within each condition revealed that for these ROIs only the 20% phase coherence conditions was always significantly ($p<.05$ corrected) larger than for all other conditions. The t-values (left FFA ($t(9) = 4.407$), right FFA ($t(9) = 4.132$), right STS ($t(9) = 4.684$) and the right AIT ($t(9)=3.376$)) and related effect sizes (Hedges g*: left FFA: 1.361; right FFA: 1.795; right STS: 1.791; and right AIT: 1.151) further indicate reliable and replicable effects with the current N = 10.

3.3.2 Functional Connectivity Measures

Functional connectivity significantly increased ($p < 0.05$, corrected) between multiple areas during the face detection task relative to fixation in multiple core face processing areas. Specifically, we found significant increases in functional connectivity between the 1) left FFA and the right FFA, right and left OFA, right STS and right AIT, 2) right FFA and right OFA, right STS and right AIT, 3) the left OFA and the right STS 4) the right OFA and the right STS (symmetrical connectivity between regions not repeated).

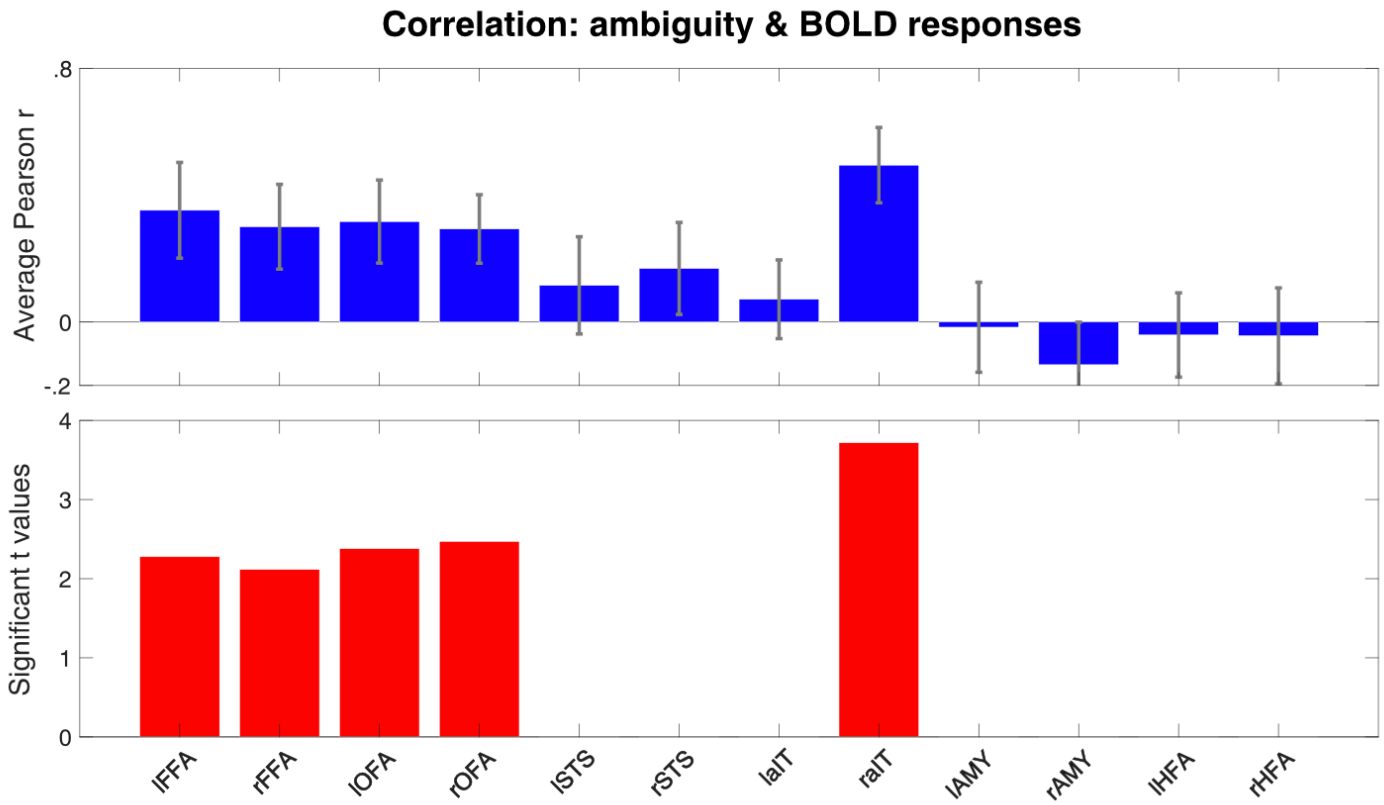


Caption Figure 5. Dynamic Reconfiguration of Face Network as a function of Task **Top Row**: Connectivity Matrices for Stimulus Relevant Face Detection Task (Left) and Stimulus Irrelevant Fixation Task (Right). Asterisks indicate correlation coefficients that significantly ($p<0.05$) difference between tasks and are identical between matrices for visualization purposes. **Bottom Row.** Classic Multidimensional Scaling for connectivity matrices highlights the higher proximity of the rAIT to the core face areas as a function of increased connectivity during face detection relative to the fixation task. ROIs in grey text indicate those regions with no significant connectivity modulations across tasks.

Multidimensional scaling was used to visualize the transformation of connectivity due to task demands. While the core face areas are significantly correlated during both tasks, the MDS makes it apparent that task demands shift the right AIT closer (i.e. significantly more correlated) during face detection relative to fixation. (See Figure 5 Bottom, multidimensional space of fixation rotated onto that of detection using Procrustes transformation)

3.3.3 Brain-Behavior Correlations

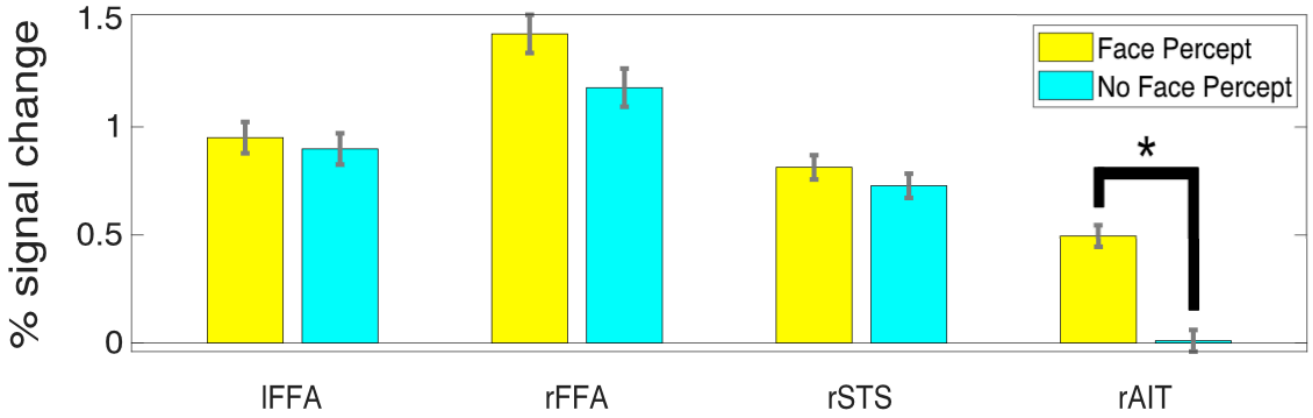
Figure 6 shows the correlation between the ambiguity score and the difference in the measured brain activity between the SR and SI tasks. Correlations where significant ($p<0.05$, corrected) in the bilateral FFA and OFAs and the right AIT.



Caption Figure 6. Brain-Behavior Correlations in FFAs, OFAs and right AIT. **Top.** The Pearson correlation coefficient between the ambiguity score and the difference in task BOLD responses. **Bottom.** For the regions reaching significance ($p<0.05$), the left and right FFAs and OFAs, and right AIT, the t values are plotted.

3.3.4 Perceptual analysis

For the 20% phase coherence condition only (i.e. the most ambiguous percept) we further separated the activity elicited by each trial according to participants perceptual response, creating 2 new conditions: “face percept” and “no face percept”. We investigated amplitude differences between these 2 conditions for those ROIs that showed a significant task x condition interaction. While all areas showed slightly larger BOLD amplitude for the face compared to the non-face percept, paired sample t-tests indicated significant amplitude differences as a function of percept in the rAIT only ($t(9) = 6.365$ $p<.05$ corrected; see figure 7). Moreover, paired sample t-tests contrasting the activation of each condition for all ROIs against 0 further indicated significant ($p<.05$ corrected) above baseline activation for all ROIs and conditions, except for the “no face percept” response in the right AIT, which did not significantly differ from baseline.



Caption Figure 7. Average BOLD response to 20% phase coherence (i.e. the most ambiguous stimulus) for the trials in which participants reported a face percept (yellow) and no face percept (magenta). Responses are reported for the regions that showed a significant ($p<.05$ corrected) task by condition interaction. Star symbol indicates significantly ($p<.05$ corrected) difference in amplitude.

4. Discussion

In this study, we used fMRI to measure top-down modulations of visual processing by varying task demands to identical visual stimuli. We used faces, a well-studied and highly meaningful visual category, and manipulated their visibility by modulating their phase coherence. In the scanner, participants were instructed to perform either 1) a domain-specific, stimulus relevant (SR) task or 2) a stimulus irrelevant (SI) task, while BOLD and behavioral data were collected.

We observed 2 types of top-down related attentional scaling mechanisms: 1) as indicated by the significant ($p<.05$ corrected) main effect of task, and in agreement with previous studies (e.g. Kay & Yeatman, 2017), we report that drawing attention to the stimuli leads to a broad gain effect, with an average increase in BOLD amplitude due to the stimulus relevant task in the FFAs, OFAs and rSTS; 2) the significant ($p<.05$ corrected) task by condition interaction in FFAs, rSTS and right AIT further indicates that specific task demands can additionally scale the BOLD responses to specific visual conditions. Post-hoc t-tests ($p<.05$ corrected) show that, for these areas, responses during face detection are *always* larger than those during fixation task *only for 20% phase coherence*. These differences across tasks can be explained by perceptual ambiguity (defined below). As supported by the correlations between behavioral and neural responses (Figure 6), and by functional connectivity analyses (Figure 5), our data is consistent with the idea that the ambiguity signal originates in the right AIT and is fed back to FFAs and from there to OFAs. Importantly, only in the right AIT we further observed BOLD modulation as a function of perceptual differences (Figure 7), with significantly larger BOLD amplitudes when participants reported seeing a face vs not. This finding represents a first indication of a pure perceptual module in humans and is in line with animal electrophysiology reports.

These results and their implications are discussed in detail below.

4.1 Task Modulations & Ambiguity

Straightforward manipulation of task demands can produce changes in the amplitude of BOLD response in higher level areas, such as the FFA (Druzgal and D'Esposito, 2001; Kay and Yeatman, 2017; Vuilleumier et al., 2001; Wojciulik et al., 1998) or STS (Narumoto et al., 2001). Traditionally, however, these tasks manipulations are often indirectly related to the neural processing of the stimuli at hand, as they serve a more general purpose

of directing attention towards (e.g. N-back) or away from the stimuli (e.g. fixation tasks, Bokde et al., 2005; Druzgal & D'Esposito, 2001; Egner & Hirsch, 2005; Kay & Yeatman, 2017; Wojciulik et al., 1998). Changes in neural responses to identical visual inputs related to these unspecific changes in task demands often reveal broad contributions from attentional networks, including frontal and parietal regions (Szczepanski et al., 2013). While these manipulations can shed light upon the neural basis of general, non-specific top-down mechanisms, such as awareness, working memory demands or vigilance, they may fail to reveal fine-grained top-down modulations pertaining to the processing of a specific stimulus category. The approach used here instead builds on task manipulations that tap into a relevant stimulus dimension. This disambiguates the contributions of various regions *within* dedicated networks (here, the face processing network), by modulating the difficulty within a stimulus relevant task. In the context of this work, we will be referring to the modulatory forces that direct attention towards (e.g. N-Back) or away from (e.g. observe fixation) the stimuli as *"non-specific top-down";* and to the task-specific modulations, as they pertain to task difficulty or cognitive load, as *"domain-specific top-down"*.

4.1.1 Task modulations: differentiating domain-specific vs non-specific top-down

Our finding that the specificity of task-related modulations is dependent on the nature of the task is consistent with a large body of literature highlighting the conceptual and anatomical differences between different types of attention. Posner and Peterson (1990) delineated 3 attention subsystems devoted to orienting, detection, and alertness. *Our results suggest that a design in which domain-specificity difficulty is varied is capable of revealing the distinct modulation associated with different attentional subsystems.* The significant ($p<.05$ corrected) main effect of task, indicates, in accordance with previous reports (Druzgal and D'Esposito, 2001; Kay and Yeatman, 2017; Vuilleumier et al., 2001; Wojciulik et al., 1998) that BOLD amplitude is on average larger during the SR relative to SI task in canonical face processing areas: the bilateral FFA, OFA and the right STS. In line with previous work, (Druzgal and D'Esposito, 2001; Kay and Yeatman, 2017; Ress et al., 2000), we suggest that these increases reflect broad top-down contributions relating to increased vigilance and awareness of the stimuli.

By contrast, there are a specific subset of areas that show an interaction between task and stimulus: the bilateral FFA as well as the right AIT and right STS. Of these only the right AIT showed only a significant interaction, driven by the activity elicited by the 20% phase coherence images being larger during the SR compared to the SI. This result indicates that, at least within the context of this work, this region's responses are

exclusively domain-specific top-down modulations. In this region in fact, other than for the 20% phase coherence, task demands do not alter BOLD amplitude to any other condition (Figure 4).

Finally, a large number of regions, specifically the FFAs, OFAs, STS, and right AIT showed a main effect of condition, driven by larger amplitude for the 40% phase relative to the 0% phase. As this comparison is equivalent to a classical 'face vs scrambled face' linear contrast (Chen et al., 2007; Rossion et al., 2012) used to define stimulus selectivity, and consistent with the preferentiality of these regions for this stimulus category, this result is unsurprising and won't be discussed further.

4.1.2 Ambiguity

Post-hoc t-tests carried out on the significant task x condition interactions revealed that in all cortical regions exhibiting specific top-down modulations, the stimulus relevant task amplitude increases were always significant only for 20% phase coherence. Behaviorally, this condition was also found to be the most perceptually ambiguous, i.e. the condition with the largest reported number of contrasting percepts (i.e. "face" and "no face"). In the context of this study we mathematically defined ambiguity as the inverse of the absolute difference from the inflection point of the sigmoid that describes the face percept/detection behavioral responses (see Equation 1). Both extremes of the phase coherence continuum therefore represent non-ambiguous percepts. That is, at low phase coherence (e.g. 0%) participants consistently reported no face percept and at high coherence (40%) participants consistently reported the presence of a face (Figure 3). At the maximally ambiguous, 20% condition, the percept was at its most unstable and thus most difficult to categorize, leading to a larger amplitude response. This relative increase is consistent with prior work showing that task difficulty can modulate the BOLD response not only in frontal or parietal regions (Culham et al., 2001; Gould et al., 2003) but also in these category sensitive visual areas (Druzgal and D'Esposito, 2001).

Notably, the reaction times measured in this study do not appear to capture this difficulty increase (Figure 3). This disconnect between reaction time and experimental performance has been noted before in relation to accuracy and attentional cueing (van Ede et al., 2012) or to task difficulty across a wide range of N-back conditions (Lamichhane et al., 2020). In our data, this likely reflects a disconnect between the idea of "task difficulty" and "difficulty in categorization" which may provide separate contributions to total reaction time.

4.1.3 Top-down driven network reconfiguration.

To understand how top-down/context modulated the inter-regional neural communication and thus the connectivity within specialized cortical networks (here, the face network), we compared functional connectivity between tasks. By computing functional connectivity on concatenated FIR curves, we: 1) discard the contribution of ongoing activity (which is not the focus of this specific work); and 2) increase statistical power and thus the reliability of our correlational metric. Broadly speaking, connectivity was significantly greater in core and extended regions (i.e. between FFAs, OFAs and right STS) during the stimulus relevant task, indicating greater communication among these areas to fulfill task demands. In particular, we observed significantly ($p<.05$ corrected) greater connectivity between the right AIT and the FFAs during the detection task relative to the fixation (Figure 5). A number of studies have suggested a functional differentiation between the right and left FFA (Meng et al., 2012; Rossion et al., 2000*)*, with the fomer being more involced in holistic processing and the latter in featural processing. It is therefore likely that the degraded stimuli used here (Figure 1) drive both individual feature detection as well as holistic face processing, thus requiring integration and feedback from high-level regions. Alternatively, it is possible that subjects flexibly adapt their strategy depending on the available information, switching from featural (e.g. looking for eyes) to holistic detection, thus engaging both FFAs.

Within the current experimental settings, our data suggest that, during face processing, the right AIT can be recruited by the cortical face network to add an additional resource to resolve ambiguous percepts. These results are consistent with the functional differentiation between core and extended cortical face networks, according to which the core system mediates the representation of more basic aspects of face processing, while the extended system is involved in higher level cognitive functions and can be selectively recruited to act in concert with the regions in the core system (Haxby et al., 2000).

Moreover, our results suggest that network connectivity is not static, but flexibly adapts to the contextual demands. This is in line with an ever growing body of literature advocating the plastic nature of functional connectivity during task and at rest (Allen et al., 2014; Cribben et al., 2013; Debener et al., 2006; Doucet et al., 2012; Hutchison et al., 2013; Sadaghiani et al., 2009) in response to cognitive and behavioral demands (Hutchison and Morton, 2016). Our results further expand these views, indicating that even subtle changes in task demands, as those implemented here, have a dramatic impact over network reconfiguration.

4.2 Using Stimulus Ambiguity to Differentiate Functional Architecture.

Though it is challenging at best to establish causality, our results suggest that the right AIT is the source of top-down modulation. First, the correlation between the difference of the BOLD signal between tasks, and each participant's behavior (ambiguity score) was only significant in the bilateral FFAs and OFAs and the right AIT. By using the difference between tasks, we sought to highlight task specific top-down effects. This finding, indicating correspondence between the ambiguity function and the difference in BOLD amplitude across tasks within these fir ROIs, suggests that this ambiguity signal originates from one of these regions. As the right AIT is further along the information hierarchy, it is plausible that it is the source of this signal. Moreover, functional connectivity analysis shows increased connectivity between the right AIT and both FFAs, between OFAs and FFAs, but not between right AIT and either of the OFAs. In addition, the FFAs show both non-specific (significant main effect of task) and domain-specific (significant task x condition interaction) top-down effects, while the right AIT shows only the task-specific, ambiguity related, top-down effects (significant interaction).

Though prior work has often implicated frontal regions or parietal areas as being a source for these top down signals when detecting objects such as faces or houses (Baldauf and Desimone, 2014; Kay and Yeatman, 2017), these studies used indirect methods (i.e. 1-back; house or face) to examine face detection. By manipulating difficulty within the context of face detection, we uncover a different, within-network source of top-down modulation by directly stressing the face detection system. This approach is similar to prior work finding modulation within the ventral temporal cortex when viewing degraded facial stimuli (Fan et al., 2020).

Furthermore, there is evidence that the AIT is involved in resolving difficult stimuli in both macaques and humans. For the former, prior work has implicated this region in learning ambiguous stimulus rules related to concurrent discrimination (Bussey et al., 2002). Increases in ambiguity were associated with worse performance in the macaques with perirhinal cortex (i.e. anterior inferior temporal lobe) lesions. In humans, this region appears to be active when integrating conceptual and perceptual information (Martin et al., 2018), determining identity (Zhang et al., 2016) or identifying confusable objects (Clarke and Tyler, 2014; Tyler et al., 2013). We therefore argue that the results presented here and those reported in previous work represent a strong indication that the ambiguity-related top-down signal during face detection originates the right AIT, is fed back to the FFAs and, from there, to the OFAs. Future work can more explicitly analyze this using specifically tailored research methods that provide better evidence for causal effects (i.e. fMRI with dynamic causal modeling or intracranial EEG).

4.2.1 Neural Correlate of Perception

Evidence for the unique role of the right AIT in perception comes from the observation that, while FFAs and right STS show both a main effect of task and a task by condition interaction, for the right AIT only the latter was found to be significant. To better understand these differences and further characterize the response profile of these regions, we grouped BOLD responses according to each participant's percept during the most ambiguous 20% condition. This analysis shows that, for the right AIT only, responses elicited by face percepts are significantly larger than those elicited by no face percepts for this area only (Figure 7), indicating a high correspondence between neural and behavioral responses. Moreover, unlike FFAs and right STS, showing significant above baseline activation for the no face percept condition, the right AIT shows no significant activation when participants failed to perceive a face (Figure 7). Taken together, these findings represent evidence for a direct link between the right AIT and subjective perception. These observations are consistent with the idea that anterior temporal regions are involved in subjective awareness (Li et al., 2014) and with reports that the AIT organizes visual objects according to their semantic interpretations (Price et al., 2017). These findings are also consistent with the data form the face localizer (Figure 1) in which the right AIT shows the most preferential face response amongst these 4 regions, with objects and scrambled images not being significantly different from zero.

The significant activation for the no face percept condition in the FFAs and right STS can be related to the bottom up physical properties of the stimulus. Regardless of the perceptual state, the 20% phase coherence images always contained a more or less visible face stimulus. Alternatively, this result can be due to the contextual top-down induced by task demands (here a face detection task). That is, as tasks were blocked by run, during SR runs, participants expected having to resolve perceptual judgment of an ambiguous stimulus, and therefore looked for a face or a face feature in every trial. This is consistent with the observation that, in the absence of a face, the FFA can be activated from contextual cues alone (Cox et al., 2004).

By contrast, the rAIT shows no response when there is no face present but only appears to activate once the stimulus is at its most difficult and ambiguous – consistent with engagement of more regions due to task difficulty (Wang et al., 2013).

Past work using ambiguous facial stimuli in the form of Mooney faces (Mooney, 1957) have found amplitude differences in the FFA (Andrews and Schluppeck, 2004) or latency differences in the rOFA (Fan et al., 2020). Alternative approaches using bistable perception, such as the face/vase illusion also report that the

FFA shows greater activation when faces are perceived (Andrews et al., 2002; Hasson et al., 2001). With binocular rivalry, in which alternative images are shown to each eye, the FFA also increases in activity when faces are perceived (Tong et al., 1998). In our present work we found that the FFA did show a larger, however, non-significant increase when subjects reported seeing a face. This apparent discrepancy with prior reports can be explained by differences in experimental paradigms. Unlike Mooney faces, or face-related bistable stimuli, the 20% phase coherence stimuli used here always have a physical face present. For binocular rivalry, the perceptual ambiguity is similar to bistable perception, however binocular rivalry is ecologically implausible. The defining characteristic of these probes is a dynamic switching amongst percepts despite identical visual input. Here, we have instead focused on the ambiguity of the initial percept, which is more similar to approaches using degraded or partially occluded static stimuli (Flounders et al., 2019; Frühholz et al., 2011). With our stimuli and experimental manipulations, *we isolate a unique and distinctive signal only in the right AIT that distinguishes the subjective perception of a face.*

Another important implication of the findings presented is their interpretation in relation to non-human primate research. Similar to humans, macaque monkeys possess a specialized cortical network in inferotemporal cortex dedicated to processing faces (e.g. Tsao et al., 2003), however the exact correspondence between human and macaque face-selective areas is still unclear (Tsao et al., 2003, 2006, 2008a). While a degree of structural and functional correspondence has been achieved with regards to the core regions (e.g. macaque middle face patches to human FFA – Tsao, Moeller, et al., 2008; Rajimehr et al., 2009), identifying a human homologues of the macaque anterior medial (AM) patch (the anterior most patch) has been challenging. Tsao, Moeller, et al. (2008) for example, failed to uncover a comparable region in humans, attributing such failure to suceptibility related signal drop out due the putative proximity of this area to the ear canal. Rajimehr et al. (2009) uncovered a face specific region in human AIT in 5 out of 10 participants, but were unable to elucidate the nature of its response properties. Here, not only were we able to identify a human face selective region in AIT in all subjects, but, importantly, we were able to define its response profile. Our results, linking the function of the face AIT to subjective perception, are in line with recent electrophysiology reports, showing that activity in the AM face patch is related to the animal's individual ability to detect a face (Moeller et al., 2017).

4.3 Benefits of High Field Imaging

As briefly mentioned in the above paragraph, much of the difficulty in imaging regions such as the AIT relates to low SNR associated because of susceptibility artifacts and inefficient RF transmit fields in the ventral temporal lobes (Devlin et al., 2000). Moreover, areas such as the AIT are typically difficult to image at more conventional (3T) field strengths and are often not located in every individual (e.g. 50% in Rajimehr et al., 2009), nor do they show large deviation from expected hemodynamic responses (e.g. Ramon et al., 2015). These issues can persist independent of field strength.

Here we used high field MRI to capitalize on higher SNR, CNR, and acceleration performance to maximize fMRI sensitivity in these regions. *For all our participants* we report consistent and corresponding regions in the anterior inferior temporal cortex on both the left and right that preferentially respond to faces and that are modulated by task demands. This increased sensitivity that led to large effect size and ROI identification in all subjects, stems from a combination of parameters optimization. Specifically, in addition to the increased SNR that accompanies UHF strength, unlike 3T acquisitions, where no effort has been put forward to for optimizing B0 and flip angles, we manually adjusted B0 inhomogeneity and flip angles to maximize SNR for each subject. Moreover, relative to previous human work, where functional voxels measured > 3 mm iso (e.g. Rajimehr et al., 2009; Ramon et al., 2015), here, we used 1.6 mm iso voxels, minimizing partial volume effects and spin dephasing and ultimately reducing signal loss in dropout regions (Thanh Vu et al., 2017).

4.4. Beyond the ventral temporal lobe: future directions

Beyond the ventral temporal lobe, there have been reports of face sensitive regions in lateral orbital regions in macaques (Barat et al., 2018; Hadj-Bouziane et al., 2008; Tsao et al., 2008b). However, evidence for the existence of these areas in humans has been mixed, uncovering such regions in approximately half (Troiani et al., 2016) or one-third of the study population (Troiani et al., 2019; Tsao et al., 2008a). In the current work, we are able to identify these areas in all subjects during the localizer portion of the study. Activation was proximal to the H-Shaped sulci, consistent with both prior reports (Troiani et al., 2019) and across subjects, as visualized after normalization to MNI space (Figure 2). Like the other ROIs in the face preferential cortical network, these regions exhibited positive responses to faces and significantly ($p < .05$) greater activity to faces relative to scrambled stimuli or objects (See Methods).

Although we did not observe meaningful stimulus relevant task modulations we did observe an interesting, although non-significant, separation of 2 areas, with regard to their functional connectivity during face detection: the right HFA exhibited a non-significant increase in activation, while its left counterpart exhibited a non-significant decrease. These areas have been suggested to be involved in social and emotional aspects of face processing in non-human primates (Barat et al., 2018), aspects that, within the current task context, are irrelevant. Our failure to find task modulations is therefore consistent with the more complex evaluative role of the frontal cortices (Noonan et al., 2012) and its engagement in social perception tasks (Barat et al., 2018; Beer et al., 2006; Freeman et al., 2010; Mah et al., 2004). Given the enlarged frontal cortices among primate species and in particular the highly developed frontal areas in humans, it will be essential to perform further research that manipulates context, value and/or salience in order to elucidate the functional role of the HFAs during face processing. Our results suggest that the combination of ecologically relevant tasks with low noise fMRI capable of resolving previously inaccessible regions of the human brain may be the only way to understand the changes underlying human cognitive flexibility.

5. Acknowledgements

Funding for this study was provided by National Institutes of Health Grants RF1 MH117015 (Ghose), RF1 MH116978 (Yacoub), and P30 NS076408 (Ugurbil). The authors would also like to thank Dr. Kendrick Kay for conceptual discussions regarding stimulus ambiguity and feedback.

6. References

Alexander, G.E., DeLong, M.R., and Strick, P.L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. Annu. Rev. Neurosci. *9*, 357–381.

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., and Calhoun, V.D. (2014). Tracking whole-brain connectivity dynamics in the resting state. Cereb. Cortex *24*, 663–676.

Andrews, T.J., and Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. Neuroimage *21*, 91–98.

Andrews, T.J., Schluppeck, D., Homfray, D., Matthews, P., and Blakemore, C. (2002). Activity in the Fusiform Gyrus Predicts Conscious Perception of Rubin's Vase–Face Illusion. Neuroimage *17*, 890–901.

Arcaro, M.J., and Livingstone, M.S. (2017). Retinotopic organization of scene areas in macaque inferior temporal cortex. J. Neurosci. *37*, 7373–7389.

Baldauf, D., and Desimone, R. (2014). Neural mechanisms of object-based attention. Science (80-.). *344*, 424–427.

Barat, E., Wirth, S., and Duhamel, J.R. (2018). Face cells in orbitofrontal cortex represent social categories. Proc. Natl. Acad. Sci. U. S. A. *115*, E11158–E11167.

Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., and Grafton, S.T. (2011). Dynamic reconfiguration of human brain networks during learning. Proc. Natl. Acad. Sci. U. S. A. *108*, 7641–7646.

Beer, J.S., John, O.P., Scabini, D., and Knight, R.T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. J. Cogn. Neurosci. *18*, 871–879.

Bokde, A.L.W., Dong, W., Born, C., Leinsinger, G., Meindl, T., Teipel, S.J., Reiser, M., and Hampel, H. (2005). Task difficulty in a simultaneous face matching task modulates activity in face fusiform area. Cogn. Brain Res. *25*, 701–710.

Bussey, T.J., Saksida, L.M., and Murray, E.A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. Eur. J. Neurosci. *15*, 365–374.

Chen, C.C., Kao, K.L.C., and Tyler, C.W. (2007). Face configuration processing in the human brain: The role of symmetry. Cereb. Cortex *17*, 1423–1432.

Chen, N.K., Dickey, C.C., Yoo, S.S., Guttmann, C.R.G., and Panych, L.P. (2003). Selection of voxel size and slice orientation for fMRI in the presence of susceptibility field gradients: Application to imaging of the amygdala. Neuroimage *19*, 817–825.

Clarke, A., and Tyler, L.K. (2014). Object-specific semantic coding in human perirhinal cortex. J. Neurosci. *34*, 4766–4775.

Cohen, M.R., and Maunsell, J.H.R. (2011). Using Neuronal Populations to Study the Mechanisms Underlying Spatial and Feature Attention. Neuron *70*, 1192–1204.

Cox, R.W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. *29*, 162–173.

Cox, D., Meyers, E., and Sinha, P. (2004). Contextually Evoked Object-Specific Responses in Human Visual Cortex. Science (80-.). *304*, 115–117.

Cribben, I., Wager, T.D., and Lindquist, M.A. (2013). Detecting functional connectivity change points for single-subject fMRI data. Front. Comput. Neurosci. *7*.

Culham, J.C., Cavanagh, P., and Kanwisher, N.G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron *32*, 737–745.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis (New York, NY, US: Routledge/Taylor & Francis Group).

Debener, S., Ullsperger, M., Siegel, M., and Engel, A.K. (2006). Single-trial EEG-fMRI reveals the dynamics of cognitive function. Trends Cogn. Sci. *10*, 558–563.

Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. *4*, 2051–2062.

Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., and Tyler, L.K. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. Neuroimage *11*, 589–600.

Doucet, G., Naveau, M., Petit, L., Zago, L., Crivello, F., Jobard, G., Delcroix, N., Mellet, E., Tzourio-Mazoyer, N., Mazoyer, B., et al. (2012). Patterns of hemodynamic low-frequency oscillations in the brain are modulated by the nature of free thought during rest. Neuroimage *59*, 3194–3200.

Druzgal, T.J., and D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. Cogn. Brain Res. *10*, 355–364.

Druzgal, T.J., and D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. J. Cogn. Neurosci. *15*, 771–784.

van Ede, F., de Lange, F.P., and Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. J. Neurosci. *32*, 10408–10412.

Egner, T., and Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat. Neurosci. *8*, 1784–1790.

Fan, X., Wang, F., Shao, H., Zhang, P., and He, S. (2020). The bottom-up and top-down processing of faces in the human occipitotemporal cortex. Elife *9*.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex.

Cereb. Cortex *1*, 1–47.

Fisher, R.A. (1915). Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. Biometrika *10*, 507.

Flounders, M.W., González-García, C., Hardstone, R., and He, B.J. (2019). Neural dynamics of visual ambiguity resolution by perceptual prior. Elife *8*.

Freeman, J.B., Rule, N.O., Adams, R.B., and Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. Cereb. Cortex *20*, 1314–1322.

Frühholz, S., Godde, B., Lewicki, P., Herzmann, C., and Herrmann, M. (2011). Face recognition under ambiguous visual stimulation: fMRI correlates of “encoding styles.” Hum. Brain Mapp. *32*, 1750–1761.

Gonzalez-Castillo, J., and Bandettini, P.A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. Neuroimage *180*, 526–533.

Gould, R.L., Brown, R.G., Owen, A.M., Ffytche, D.H., and Howard, R.J. (2003). fMRI BOLD response to increasing task difficulty during successful paired associates learning. Neuroimage *20*, 1006–1019.

Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. *15*, 536–548.

Hadj-Bouziane, F., Bell, A.H., Knusten, T.A., Ungerleider, L.G., and Tootell, R.B.H. (2008). Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. Proc. Natl. Acad. Sci. U. S. A. *105*, 5591–5596.

Handwerker, D. a, Ollinger, J.M., and D’Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage *21*, 1639–1651.

Hasson, U., Hendler, T., Bashat, D. Ben, and Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J. Cogn. Neurosci. *13*, 744–753.

Haxby, J. V., Hoffman, E.A., and Gobbini, M.I. (2000). The distributed human neural system for face perception. Trends Cogn. Sci. *4*, 223–233.

Haxby, J. V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science (80-.). *293*, 2425–2430.

Hedges, L. V. (1981). Distribution Theory for Glass’s Estimator of Effect Size and Related Estimators. J. Educ. Stat. *6*, 107.

Hedges, L. V., Olkin, I., Freeman, P.R., Hedges, L. V., Olkin, I., Rosenbaum, P.R., Hedges, L. V., Olkin, I., Freeman, P.R., Hedges, L. V., et al. (1985). Statistical Methods for Meta-Analysis (Elsevier).

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat’s striate cortex. J. Physiol.

148, 574–591.

Hutchison, R.M., and Morton, J.B. (2016). It's a matter of time: Reframing the development of cognitive control as a modification of the brain's temporal dynamics. Dev. Cogn. Neurosci. *18*, 70–77.

Hutchison, R.M., Womelsdorf, T., Gati, J.S., Everling, S., and Menon, R.S. (2013). Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. Hum. Brain Mapp. *34*, 2154–2177.

Kabbara, A., EL Falou, W., Khalil, M., Wendling, F., and Hassan, M. (2017). The dynamic functional core network of the human brain at rest. Sci. Rep. *7*, 2936.

Kanwisher, N., and Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. Philos. Trans. R. Soc. B Biol. Sci. *361*, 2109–2128.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.

Kay, K.N., and Yeatman, J.D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. Elife *6*, e22341.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front. Psychol. *4*, 863.

Lamichhane, B., Westbrook, A., Cole, M.W., and Braver, T.S. (2020). Exploring brain-behavior relationships in the N-back task. Neuroimage *212*, 116683.

Laughlin, S.B., and Sejnowski, T.J. (2003). Communication in neuronal networks. Science (80-.). *301*, 1870–1874.

Leung, A.W.S., and Alain, C. (2011). Working memory load modulates the auditory "What" and "Where" neural networks. Neuroimage *55*, 1260–1269.

Li, Q., Hill, Z., and He, B.J. (2014). Spatiotemporal dissociation of brain activity underlying subjective awareness, objective performance and confidence. J. Neurosci. *34*, 4382–4395.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., and Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods *264*, 47–56.

Liu, B., Zhu, T., and Zhong, J. (2015). Comparison of quality control software tools for diffusion tensor imaging. Magn. Reson. Imaging *33*, 276–285.

Mah, L., Arnold, M.C., and Grafman, J. (2004). Impairment of social perception associated with lesions of the prefrontal cortex. Am. J. Psychiatry *161*, 1247–1255.

Martin, C.B., Douglas, D., Newsome, R.N., Man, L.L.Y., and Barense, M.D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. Elife *7*, e31873.

Martinez-Trujillo, J.C., and Treue, S. (2004). Feature-based attention increases the selectivity of population

responses in primate visual cortex. Curr. Biol. *14*, 744–751.

De Martino, F., Moerel, M., Ugurbil, K., Goebel, R., Yacoub, E., and Formisano, E. (2015). Frequency preference and attention effects across cortical depths in the human primary auditory cortex. Proc. Natl. Acad. Sci. U. S. A. *112*, 16036–16041.

McCarthy, G., Puce, A., Gore, J.C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. *9*, 605–610.

Meng, M., Cherian, T., Singal, G., and Sinha, P. (2012). Lateralization of face processing in the human brain. Proc. R. Soc. B Biol. Sci. *279*, 2052–2061.

Moeller, S., Freiwald, W.A., and Tsao, D.Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe. Science (80-.). *320*, 1355–1359.

Moeller, S., Crapse, T., Chang, L., and Tsao, D.Y. (2017). The effect of face patch microstimulation on perception of faces and objects. Nat. Neurosci. *20*, 743–752.

Mooney, C.M. (1957). Age in the development of closure ability in children. Can. J. Psychol. *11*, 219–226.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., and Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Cogn. Brain Res. *12*, 225–231.

Noonan, M.P., Kolling, N., Walton, M.E., and Rushworth, M.F.S. (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. Eur. J. Neurosci. *35*, 997–1010.

Parker, A.J., and Newsome, W.T. (1998). SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception. Annu. Rev. Neurosci. *21*, 227–277.

Perrett, D.I., Rolls, E.T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp. Brain Res. *47*, 329–342.

Pessoa, L., Gutierrez, E., Bandettini, P., and Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. Neuron *35*, 975–987.

Posner, M.I., and Petersen, S.E. (1990). The Attention System of the Human Brain. Annu. Rev. Neurosci. *13*, 25–42.

Price, A.R., Bonner, M.F., Peelle, J.E., and Grossman, M. (2017). Neural coding of fine-grained object knowledge in perirhinal cortex. BioRxiv *194829*, 1–18.

Rajimehr, R., Young, J.C., and Tootell, R.B.H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. Proc. Natl. Acad. Sci. U. S. A. *106*, 1995–2000.

Ramon, M., Vizioli, L., Liu-Shuang, J., and Rossion, B. (2015). Neural microgenesis of personally familiar face recognition. Proc. Natl. Acad. Sci. U. S. A. *112*, E4835–E4844.

Ress, D., Backus, B.T., and Heeger, D.J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. Nat. Neurosci. *3*, 940–945.

Rossion, B., Dricot, L., Devolder, A., Bodart, J.M., Crommelinck, M., De Gelder, B., and Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. J. Cogn. Neurosci. *12*, 793–802.

Rossion, B., Hanseeuw, B., and Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. Brain Cogn. *79*, 138–157.

Sadaghiani, S., Hesselmann, G., and Kleinschmidt, A. (2009). Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. J. Neurosci. *29*, 13410–13417.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al. (2013). Resting-state fMRI in the Human Connectome Project. Neuroimage *80*, 144–168.

Stigliani, A., Weiner, K.S., and Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. J. Neurosci. *35*, 12412–12424.

Szczepanski, S.M., Pinsk, M.A., Douglas, M.M., Kastner, S., and Saalmann, Y.B. (2013). Functional and structural architecture of the human dorsal frontoparietal attention network. Proc. Natl. Acad. Sci. U. S. A. *110*, 15806–15811.

Taylor, A.J., Kim, J.H., and Ress, D. (2018). Characterization of the hemodynamic response function across the majority of human cerebral cortex. Neuroimage *173*, 322–331.

Thanh Vu, A., Jamison, K., Glasser, M.F., Smith, S.M., Coalson, T., Moeller, S., Auerbach, E.J., Uğurbil, K., and Yacoub, E. (2017). Tradeoffs in pushing the spatial resolution of fMRI for the 7T Human Connectome Project. Neuroimage *154*, 23–32.

Tong, F., Nakayama, K., Vaughan, J.T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. Neuron *21*, 753–759.

Troiani, V., Dougherty, C.C., Michael, A.M., and Olson, I.R. (2016). Characterization of face-selective patches in orbitofrontal cortex. Front. Hum. Neurosci. *10*.

Troiani, V., Patti, M.A., and Adamson, K. (2019). The use of the orbitofrontal H-sulcus as a reference frame for value signals. Eur. J. Neurosci. *n/a*, ejn.14590.

Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., and Tootell, R.B.H. (2003). Faces and objects in macaque cerebral cortex. Nat. Neurosci. *6*, 989–995.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. Science (80-.). *311*, 670–674.

Tsao, D.Y., Moeller, S., and Freiwald, W.A. (2008a). Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. U. S. A. *105*, 19514–19519.

Tsao, D.Y., Schweers, N., Moeller, S., and Freiwald, W.A. (2008b). Patches of face-selective cortex in the

macaque frontal lobe. Nat. Neurosci. *11*, 877–879.

Tyler, L.K., Chiu, S., Zhuang, J., Randall, B., Devereux, B.J., Wright, P., Clarke, A., and Taylor, K.I. (2013). Objects and categories: Feature statistics and object processing in the ventral stream. J. Cogn. Neurosci. *25*, 1723–1735.

Vogels, R., and Orban, G.A. (1994). Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. J. Neurophysiol. *71*, 1428–1451.

Vuilleumier, P., Armony, J.L., Driver, J., and Dolan, R.J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. Neuron *30*, 829–841.

Wang, M., Arteaga, D., and He, B.J. (2013). Brain mechanisms for simple perception and bistable perception. Proc. Natl. Acad. Sci. U. S. A. *110*, E3350–E3359.

Wang, S.S.H., Shultz, J.R., Burish, M.J., Harrison, K.H., Hof, P.R., Towns, L.C., Wagers, M.W., and Wyatt, K.D. (2008). Functional trade-offs in white matter axonal scaling. J. Neurosci. *28*, 4047–4056.

Wilcox, R. (2005). Introduction to Robust Estimation and Hypothesis Testing (Academic Press).

Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., and Tanaka, J.W. (2010). Controlling low-level image properties: The SHINE toolbox. Behav. Res. Methods *42*, 671–684.

Wojciulik, E., Kanwisher, N., and Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. J. Neurophysiol. *79*, 1574–1578.

Zhang, R., and Kay, K. (2018). Flexible top-down modulation in human ventral temporal cortex. Neuroimage 116964.

Zhang, H., Japee, S., Nolan, R., Chu, C., Liu, N., and Ungerleider, L.G. (2016). Face-selective regions differ in their ability to classify facial expressions. Neuroimage *130*, 77–90.