Clarifying the role of higher-level cortices in resolving perceptual ambiguity using Ultra High Field fMRI

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

1. Center for Magnetic Resonance Research, University of Minnesota 2021 6th St SE, Minneapolis, MN 55455

* Luca Vizioli and Logan Dowdle

Email: luca.vizioli1@gmail.com; logan.dowdle@gmail.com

Subject Area

Neuroscience

Keywords

Top-down; fMRI; Ultra High Field; Cognitive Neuroscience; Vision; Perception; Cortical Network; Task modulations;

Author Contributions

L.V. designed and conducted the experiment; L.V. and L.D. performed the analyses, wrote the manuscript and generated the figures. E.Y. and K.U. provided guidance on choice and optimization of MR acquisition sequences; G.G. provided guidance on conceptual framework; E.Y., K.U. and G.G. contributed to manuscript formulation and writing.

Abstract

The brain is organized into distinct, flexible networks. Within these networks, cognitive variables such as attention can modulate sensory representations in accordance with moment-to-moment behavioral requirements. These modulations can be studied by varying task demands; however, the tasks employed are often incongruent with the postulated functions of a sensory system, limiting the characterization of the system in relation to natural behaviors. Here we combine domain-specific task manipulations and ultra-high field fMRI to study the nature of top-down modulations. We exploited faces, a visual category underpinned by a complex cortical network, and instructed participants to perform either a stimulus-relevant/domain-specific or a stimulus-irrelevant task in the scanner. We found that 1. perceptual ambiguity (i.e. difficulty of achieving a stable percept) is encoded in top-down modulations from higher-level cortices; 2. the right inferior-temporal lobe is active under challenging conditions and uniquely encodes trial-by-trial variability in face perception.

Introduction

33

34 Decades of human and animal research has demonstrated that cortical areas are functionally and

35 anatomically linked to form distinct brain networks. Understanding the distinct contributions of

36 individual areas within these networks has been challenging in part because of the rich and reciprocal

37 pattern of connections between brain areas (Alexander et al., 1986; Felleman and Van Essen, 1991;

38 Moeller et al., 2008). Discrimination is further confounded by the low latencies with which signals travel

39 between areas (Laughlin and Sejnowski, 2003; Wang et al., 2008), substantial overlaps in the functional

40 sensitivities of neurons of different areas (Arcaro and Livingstone, 2017; Haxby et al., 2001; Vogels and

41 Orban, 1994), and the flexibility of connections necessary to maintain function across a range of tasks

42 and environments (Bassett et al., 2011; Gonzalez-Castillo and Bandettini, 2018; Kabbara et al., 2017). A

43 critical component of this flexibility in sensory systems is top-down control, in which cognitive variables

44 such as attention, value, and memory can alter sensory representations in accordance with explicit or

45 specific behavioral demands. Yet this flexibility also is a major challenge to experimental studies, because

46 it implies that the nature of signal processing both within and between brain areas of a network is not

47 fixed and very much depends on the behavioral context and goals of the subject.

48 While many humans and animal studies have demonstrated the ability of task manipulations to

49 modulate sensory evoked responses, in most cases these studies have employed tasks that are not

50 specific to the domain of the stimulus that was presented. For example, numerous electrophysiological

51 and imaging studies suggest that attending to a particular location in visual space increases the

52 responses of neurons whose receptive fields lie within the attended location irrespective of the

53 selectivities of those receptive fields (Cohen and Maunsell, 2011; Liu et al., 2015; Martinez-Trujillo and

54 Treue, 2004). Similarly, perceptual tasks involving working memory often result in a broad distribution

55 of enhanced signals in sensory areas (Druzgal and D'Esposito, 2003, 2001; Kay and Yeatman, 2017; Leung

3

56 and Alain, 2011; Pessoa et al., 2002). By contrast, top-down modulation from domain-specific tasks can

57 be far more targeted, selectively enhancing only those neurons whose receptive field responses are

58 consistent with behavioral demands (De Martino et al., 2015; Zhang and Kay, 2020). Accordingly, the use

59 of appropriate domain specific tasks is likely to highlight those neurons with relevant selectivities and

60 reveal how they contribute to perception. For example, contrasting the activity elicited by identical face

61 stimuli in an N-back and a fixation task may reveal working memory related modulations that are

62 unspecific to face processing and therefore fail to elucidate the changes related to the specialized

63 processing of faces. Indeed, a N-back face task could be performed without explicit face recognition or

64 perception; additionally, one could imagine that explicit vs. implicit face detection may be underpinned

65 by different neural computations, and therefore, implementing a task with no explicit requirement may

66 represent a confound or lead to a different set of results.

67 Establishing how different areas within a network contribute to perception requires sampling the

68 entire network and then identifying the particular areas that can best explain task performance.

69 However, technical limitations have precluded this approach. Invasive electrophysiological studies are

70 necessarily highly targeted and unable to sample across an entire brain network. In addition, these

71 techniques are frequently performed in non-human primates, with tasks that fail to capture the cognitive

72 sophistication available to humans. Moreover, even whole brain studies, such as those relying upon

73 functional magnetic resonance imaging, are often highly biased in their sampling because of the difficulty

74 of imaging particular areas. For example, a number of high-level areas – crucial for human cognition –

75 are located in cortical loci that are traditionally difficult to image (e.g. due to low SNR and their proximity

76 to air cavities (Chen et al., 2003; Devlin et al., 2000)). These limitations also constrain our ability to

77 establish appropriate animal models for invasive measurements and manipulations because of the

78 possibility that human areas homologous to those in animal studies are difficult to sample with

79 conventional gradient echo fMRI methods (Rajimehr et al., 2009). Signal to noise issues have also limited

80 the ability of fMRI to establish trial-to-trial behavioral covariance, which is a critical requirement for

81 understanding the neural basis of perception (Parker and Newsome, 1998) and has been extensively

82 employed by electrophysiology studies (Desimone et al., 1984; Hubel and Wiesel, 1959; Martinez-Trujillo

83 and Treue, 2004; Perrett et al., 1982). This is a particularly important consideration because, by

84 definition, all areas within a network are co-activated, but, given the richness of connectivity within the

85 network, that co-activation does not necessarily imply perceptual relevance.

86   While the difficulties surrounding magnetic field inhomogeneity and transmit uniformity can be

87 more detrimental at higher fields, there are substantial SNR gains available with ultra-high field magnets

88 (UHF i.e. 7T and above). In common practice, these SNR gains are generally traded for higher resolutions

89 (e.g. submillimeter). In this paper we instead select a relatively lower resolution (i.e. 1.6 mm^3) and

90 leverage the improved SNR available in UHF fMRI to study the entire ensemble of brain areas associated

91 with a behaviorally important high-level perceptual task, namely the detection of faces in poor visibility.

92 We choose faces as these are a well-studied, cross species visual category that is highly meaningful and

93 therefore ideal for investigating high-level, top-down neural modulations. The human face processing

94 network has however proved challenging to study (Kanwisher et al., 1997; Kanwisher and Yovel, 2006;

95 McCarthy et al., 1997) (see Grill-Spector & Weiner 2014 for review), because many of its areas, due to

96 their proximity to air cavities, are heavily affected by the aforementioned fMRI signal difficulties. In order

97 to combat susceptibility related dropout we leverage the improved acceleration available at high fields

98 to increase our imaging resolution to 1.6mm relative to the 2mm gold-standard for 3T studies. Combined

99 with manual shimming, this yields substantial gains in SNR, while reducing drop out (Farzaneh et al.,

100 1990; Olman et al., 2009; Young et al., 1988) in key human face processing areas. In the current work,

101 we varied the visibility of face stimuli by modulating their phase coherence (Figure 1) and instructed the

102 subjects to perform 2 tasks in the scanner: a domain-specific, stimulus relevant task involving perceptual

103 judgment of the visual stimuli (i.e. face detection); and, stimulus irrelevant fixation task. By leveraging

5

104   signal gains derived from our imaging approach, we then contrast these responses to isolate top-down

105   modulations specific to the face domain and link subjective perception to brain activity.

106

Materials and Methods

108   *Participants*

109   10 (5 females) healthy right-handed subjects (age range: 18-31) participated in the study. Of

110   these, 1 participant was re-scanned due to excessive motion during scanning. All subjects had normal,

111   or corrected vision and provided written informed consent. The local IRB at the University of Minnesota

112   approved the experiments.

113

114   *Stimuli and procedure*

115   The experimental procedure consisted of a standard block design face localizer and a fast event-

116   related face paradigm. For both experiments, the stimuli were centered on background of average

117   luminance (25.4 cd/m2, 23.5-30.1). Stimuli were presented on a Cambridge Research Systems

118   BOLDscreen 32 LCD monitor positioned at the head of the 7T scanner bed (resolution 1920, 1080 at 120

119   Hz; viewing distance ~89.5 cm) using a Mac Pro computer. Stimulus presentation was controlled using

120   Psychophysics Toolbox (3.0.15) based code. Participants viewed the images though a mirror placed in

121   the head coil. Behavioral responses were recorded using Cambridge Research Systems button box and

122   Psychophysics Toolbox.

123

124   *Face localizer*

125   *Stimuli*

126   All visual stimuli used for the face localizer consisted of grayscale photographs depicting 20

127   different faces (10 identities × 2 genders) and objects (both taken from Stigliani et al., 2015) and textures

128   of noise.

7

129   The experiment consisted of the presentation of grayscale stimuli drawn from different stimulus

130   categories. There were 11 categories, grouped into 3 stimulus domains: faces (adult, child), no face

131   objects (including characters (word, number), body parts (body, limb), places (corridor, house), and

132   objects (car, instrument)) and phase scrambled faces, computed by randomly scrambling the phase

133   coherence of the face stimuli. Each stimulus was presented on a scrambled background (different

134   backgrounds for different stimuli) and occupied a square region with dimensions 10° × 10°. Noise texture

135   stimuli were created by randomly scrambling the phase of the face images (i.e. 0% phase coherence).

136   The stimuli were equated in terms of luminance, contrast and spatial frequency content by taking

137   the Fourier spectra across stimuli and ensuring that the rotational average amplitudes for a given spatial

138   frequency were equated across images while preserving the amplitude distribution across orientations

139   (Willenbockel et al., 2010). The root mean square contrast (i.e. the standard deviation of pixel intensity)

140   was also kept constant across stimuli.

141

142   *Visual presentation paradigm*

143   Face localizer runs involved presentation of blocks of faces, objects and noise textures. Each run

144   began with presentation of a black fixation cross displayed on a grey background for 12 sec and consisted

145   of 9 randomly presented blocks of images. Each block (3 blocks/category; separated by a 12 sec fixation)

146   involved presentation of 10 different stimuli randomly presented for 800 ms, separated by a 400 ms

147   interstimulus interval (ISI). To ensure that participants' attention was maintained throughout the

148   localizer we implemented a 1-back task, where subjects were instructed to respond to the repetition of

149   2 identical stimuli (representing about 10% of the trials), by pressing a button on a response pad held in

150   their right hand. All participants completed two runs of the block design face localizer, where each block

151   occurred 3 times within a run, for a total run duration of 228 seconds.

152

153   *Event related experiment.*

154   *Stimuli*

155   We used grayscale images of faces (20 male and 20 female), presenting neutral expressions. We

156   manipulated the phase coherence of each face, from 0% to 40% in steps of 10%, resulting in 200 images

157   (5 visual conditions x 20 identities x 2 genders). We equated the amplitude spectrum across all images.

158   Stimuli approximately subtended 9° of visual angle. Faces were cropped to remove external features by

159   centering an elliptical window with uniform gray background to the original images. The y diameter of

160   the ellipse spanned the full vertical extent of the face stimuli and the x diameter spanned 80% of the

161   horizontal extent. Before applying the elliptical window to all face images, we smoothed the edge of the

162   ellipse by convolving with an average filter (constructed using the "fspecial" function with "average"

163   option in MATLAB – see Figure 1A). This procedure was implemented to prevent participants from

164   performing edge detection, rather than the face detection task, by reacting to the easily identifiable

165   presence of hard edges in the face images.

166   Like for the localizer experiment, here too we equated amplitude spectrum across the whole

167   stimulus set. We controlled the Fourier spectra across stimuli, ensuring that the rotational average

168   amplitudes for a given spatial frequency were equated across images while preserving the amplitude

169   distribution across orientations (Willenbockel et al., 2010). The root mean square contrast (i.e. the

170   standard deviation of pixel intensity) was also kept constant across stimuli.

171

172   *Tasks*

173   In the scanner, participants were instructed to maintain fixation on a central cross throughout

174   the run and to perform one of 2 tasks: one domain-specific face detection task, involving perceptual

175   judgment of the visual stimuli; and a second, difficult, non-specific attention fixation task that required

176   responding to a specific color change (i.e. red) of the fixation cross.

177   In the former, participants were instructed to respond as quickly as possible by pressing one of 2

178   buttons on their button box to indicate whether they perceived a face. Subjects' instructions were

179   carefully delivered to indicate that there were no correct answers and that we were instead interested

180   in the subjects' perception only (Figure 1B).

181   The latter was designed to isolate bottom-up stimulus-driven responses. In order to ensure this,

182   we piloted the fixation task to maximize task difficulty and direct the attention away from the face stimuli

183   on 4 participants that were not included in the experimental subject pool. Based on participants'

184   feedback, the fixation tasks entailed pressing a button every time the fixation turned red. Every 500 ms,

185   the fixation changed to one of five colors — specifically red, blue, green, yellow and cyan — in a

186   pseudorandom fashion, avoiding consecutive presentations of the same color (Figure 1C). The color

187   change occurred out of sync with the stimulus presentation onset and frequency of button presses was

188   kept constant across tasks. Visual stimuli were identical across tasks to ensure that any differences

189   observed were related solely to top-down processes. Tasks were blocked by run and counterbalanced

190   across participants.

191

192   *Visual presentation paradigm.*

193   We acquired 3 runs per task. Each run lasted approximately 3 mins and 22 secs and began and

194   ended with a 12-second fixation period. Within each run, we showed 40 images (5 phase coherence

195   levels x 4 identities x 2 genders) presented for 2000 ms, with a 2000 ms interstimulus interval (ISI).

196   Importantly, we introduced 10% blank trials (i.e. 4000 ms of fixation period) randomly interspersed

197   amongst the 40 images, effectively jittering the ISI. Stimulus presentation was pseudorandomized across

198   runs, with the only constraint being the non-occurrence of 2 consecutive presentations of the same

199   phase coherence level. Behavioral metrics, including reaction time and responses to face stimuli

200   indicating participants' perceptual judgments (i.e. face or no face) were generated per subject and then

201   averaged.

202

203   *Ambiguity Calculation*

204   For the face detection task, we calculated the average face response percentage to each

205   phase condition, averaged across runs. These values vary from 0 to 100%, representing consistent non-

206   face to consistent face response respectively. Within each subject, this produces a sigmoidal shaped

207   curve. We mathematically define the ambiguity score, on a subject by subject basis, as the inverse of the

208   absolute distance from the inflection point on the sigmoid that fit this curve This is shown in Equation 1:

209   Eq 1. $(|S_i - S_k| \times -1) + \min(A)$

210   where S_i is the *i*th point of the faceness sigmoid, S_k is the theoretical midpoint of the faceness

211   sigmoid and min(A) is the minimum value of the ambiguity function. While conceptually similar to

212   defining a perceptual threshold, the goal of this analytical approach is to characterize the instability of

213   the percept in a threshold-free, data-driven manner for all visual conditions, rather than determining the

214   precise level of phase coherence information required to achieve a reliable face percept.

215   *MR Imaging Acquisition and Processing*

216   All functional MRI data were collected with a 7T Siemens Magnetom System with a 1 by 32-

217   channel NOVA head coil. T2*-weighted images were collected using sequence parameters (TR 1s,

218   Multiband 5, GRAPPA 2 , 7/8ths Partial Fourier, 1.6mm isotropic, TE 22ms, Flip Angle 52°, Bandwidth

219   1923Hz, 85 slices, FOV 208 x 208 mm) matched to the Human Connectome 7T protocol (Thanh Vu et al.,

220   2017).

221   A number of steps were taken to maximize signal quality of the present study. First, we manually

222   shimmed the B0 field to maximize homogeneity over ventral and anterior-temporal regions. Moreover,

223   we selected our flip angle based on the offset between the selected flip angle and the subject specific

224   flip angle, using the results from a 3dAFI (Actual Flip Angle Image) sequence. In this way we optimized

225   flip angles across the brain to maximize SNR in the ventral and anterior-temporal regions. Finally, by

226   taking advantage of the greater ability to accelerate image acquisition at 7T, we were able to obtain

227   higher resolution images and relative reductions in magnetic susceptibility dropout in these regions

228   (Farzaneh et al., 1990; Olman et al., 2009; Young et al., 1988).

229   T1-weighted anatomical images were obtained using an MPRAGE sequence (192 slices; TR, 1900

230   ms; FOV, 256 x 256 mm; flip angle, 9°; TE, 2.52 ms; spatial resolution, .8 mm isotropic voxels) which were

231   collected with a 3T Siemens Magnetom Prisma[fit] system. Anatomical images were used for visualization

232   purposes only.

233

234   *Functional Image Processing*

235   Dicom files were converted using dcm2niix (Li et al., 2016). Subsequence functional image

236   processing was performed in AFNI version 19.2.10 (Cox, 1996). Conventional processing steps were used,

237   including despiking, slice timing correction, distortion and motion correction, and alignment to each

238   participant's anatomical image. With each task type, the target for time series and anatomical alignment

239   was the Single Band Reference (SBRef) image which is acquired to calibrate coil sensitively profiles prior

240   to the multiband acquisition and has no slice acceleration or T_1-saturation, yielding high contrast (Smith

241   et al., 2013). In order to improve localizer and task alignment, an additional nonlinear transform was

242   computed between the task SBRef and localizer SBRef. All spatial transforms were concatenated and

243   applied in a single step to reduce image blurring, and functional images were inspected to confirm

244   successful registration to anatomical targets.

245

246   *Localizer Task Analyses*

247   For the functional localizer, the data was then smoothed with a Gaussian kernel with FWHM of

248   2 voxels (3.2mm), and each run scaled to have mean 100. For regressors of no interest, the 6 estimated

249   rigid-body motion parameters and polynomials up to order 3 (including linear trend) were added to the

250   design matrix.

251

252   *Functional ROI Definition*

253   Using 3dDeconvolve, the data were passed through a GLM in order to determine the response

254   to faces, objects and noise. Each event was modeled as a 12s box car, convolved with AFNI's SPMG1 HRF

255   estimation. Rather than select face patches by comparing face activation to the average of objects and

256   noise, we instead constrained the statistical map in the following ways: 1) Betas to faces were positive,

257   2) the T-stat used for thresholding was the minimum positive T-stat from faces > objects or faces > noise.

258   In other words, we selected voxels that were significant for faces greater than objects, in conjunction

259   with faces greater than scrambled. Regions of interest (ROIs) were derived in volume space, and all

260   consisted of contiguous clusters made up of 19 or more voxels. Statistical thresholding was adjusted to

261   obtain consistency between subjects, no voxels above $p<0.05$ were considered. To aid in ROI definition,

262   we simultaneously viewed the surface representation of the statistical map using SUMA (Saad and

263   Reynolds, 2012) with FreeSurfer (Dale et al., 1999; Fischl et al., 2004, 2002, 2001) defined cortical

264   surfaces and atlas labels. In addition, to supplement the statistical parametric maps, we used finite

265   impulse response (FIR) deconvolution, with 20 bins, to estimate HRFs in response to each localizer

266   condition (face, scrambled, object). These HRFs were also used to inspect data quality for each ROI.

267

Face Detection and Fixation Color Tasks Analyses

268

269   For the Face/Fix detection tasks, the data were only scaled following initial processing, no

270   smoothing was used. As we were interested in BOLD responses across the whole brain, and we know

271   that hemodynamic response functions differ across cortical regions (Handwerker et al., 2004; Lewis et

272   al., 2018; Taylor et al., 2018), we performed a GLM based finite impulse response (FIR) deconvolution

273   analyses, which does not make assumptions regarding the shape of the HRF. This was done using AFNI's

274   TENTZero method, estimating responses with 1s bins out to 15s post stimulus. Deconvolution was

275   performed separately for each task (face detection and fixation) and phase coherence condition,

276   generating 5 FIR curves for the fixation task and 5 for the face detection task. For each task and

277   independently per subject, ROI, run and condition, we averaged all FIR curves across all voxels. To avoid

278   the contribution of extreme voxels, we trimmed 5% of the values falling at each extreme of the

279   distribution tail to compute the 10% trimmed mean. We then extracted the 4 timepoints corresponding

280   to 4, 5, 6 and 7 seconds after stimulus onset, i.e. those with the largest amplitude within a time window

281   spanning from 2 to 10 TRs after stimulus onset. The 12 extracted amplitudes (4 timepoints per run) were

282   then averaged within experimental tasks to obtain one percent signal change value per subject, ROI, task

283   and condition.

284

Statistical Analyses

285

286   To test for statistically significant differences between conditions (i.e. phase coherence levels)

287   and tasks, we carried out the following statistical tests.

288

289   *Task and Condition Analyses*

290   Independently per ROI, we performed a 2 (tasks) by 5 (phase coherence levels) repeated

291   measures ANOVA with the mean BOLD response as a dependent variable. To control for family wise error

292   rate (FWER), we implemented the following multiple comparison correction procedure: we began by

293   centering the data on the group mean for each condition and task. This procedure effectively put the

294   data under the ideal H0 hypothesis of no difference between the means. We then sampled with

295   replacement the subjects and performed the same 2 x 3 repeated measures ANOVA and stored all F

296   values for the relevant main effects and interactions. We repeated this procedure 10,000 times and

297   selected the 95% largest F values. We used these F values as our new thresholds and considered

298   statistical significance only when p values for the original ANOVA were smaller than .05 *and* the

299   connected F values were larger than their centered counterpart (e.g. (Wilcox, 2005))

300   When appropriate (i.e. for comparisons entailing more than 2 factors), we further performed

301   post-hoc paired sample t-tests on significant (as defined above) main effects and interactions. The same

302   FWER control procedure described above for the ANOVA test was implemented to account for multiple

303   comparisons.

304   Additionally, we performed power analyses to determine effect size and the sample size required

305   to achieve adequate power (see results). Specifically, we computed Hedges' g (Freeman et al., 1986;

306   Hedges, 1981). This choice was motivated by the fact that, unlike Cohen's d, which, especially for small

307   samples (i.e. n < 20), tends to provide positively biased estimators of population effect sizes (Freeman

308   et al., 1986; Lakens, 2013), Hedges' g tends to be unbiased (Cumming, 2012).

15

309

310   *Functional connectivity analyses.*

311   To determine the extent to which task demands modulate functional connectivity amongst face

312   ROIs, we carried out the following analysis. Independently per subject, task and ROI, we averaged (mean)

313   all FIR response curves amongst voxels and concatenated the time courses elicited by each condition to

314   produce a single time course. The concatenation was done to maximize statistical power, as the resulting

315   time course contained 80 timepoints (i.e. 16 FIR timepoints x 5 conditions). For each subject and task,

316   we then computed Pearson's correlation coefficient between the concatenated time courses of all

317   possible pairs of ROIs. This procedure lead to the formulation of 2 connectivity matrices (12 ROIs x 12

318   ROIs) per subject (1 per task) summarizing the similarity of BOLD responses between all pairs of ROIs. To

319   convert the skewed sampling distribution of Pearson's r into a normal distribution we computed Fisher's

320   z transformation (Fisher, 1915). We therefore proceeded to carry out paired sample t-tests between the

321   connectivity estimated obtained for the 2 tasks. To control for FWER, we implemented the same

322   bootstrap procedure on centered data described in the previous paragraphs (see paragraph 4.6.1). For

323   display purposes only, after computing the mean between the Fisher z-normalized connectivity matrices,

324   we computed the inverse of such transformation on the group average connectivity matrices to convert

325   these scores back into meaningful and interpretable Pearson's r.

326   To better visualize the results of our functional connectivity analysis, we further performed

327   classic multidimensional scaling (MDS - using the function "cmdscale" in MATLAB) on the participants

328   average dissimilarity matrix (i.e. 1- Pearson's r). MDS was performed independently per task. For ease

329   of visual comparison, the MDS arrangements of the 2 tasks were aligned by means of linear

330   transformations (including translation, reflection, orthogonal rotation, and scaling) using Procrustes

16

331   rotations (this was implemented with the "procrustes" function in MATLAB) with the sum of squared

332   errors as stress metric.

333   MDS is a useful data driven approach to visualize the data projected into a new space whose

334   dimensions are the first (in this case 2) eigenvectors (i.e. those explaining most of the variance in the

335   data) without any prior hypotheses. MDS is therefore a dimension reduction technique that highlights

336   the dominant features in the data. Results are shown in Figure 4; distance between the points indicate

337   dissimilarity of responses.

338

339   *Brain-behavior correlation.*

340   Next for each ROI, we wanted to assess the relationship between behavioral responses and top-

341   down BOLD modulations. To this end, for each subject and ROI, we computed Pearson's correlation

342   coefficients amongst the ambiguity scores at each phase coherence level and the task difference in

343   average BOLD amplitudes elicited by each condition. We then performed Fisher z transform (see

344   paragraph above) on Pearson's r and carried out one sample t-test for each ROI to determine whether

345   the average group correlation was significantly larger than 0. To control the family wise error rate

346   (FWER), we implemented the same bootstrap procedure on centered data described in the previous

347   paragraphs (see paragraph 4.6.1). For display purposes only, after computing the mean between the

348   Fisher z-normalized correlation scores, we computed the inverse of such transformation on the group

349   average to convert these scores back into meaningful and interpretable Pearson's r.

350

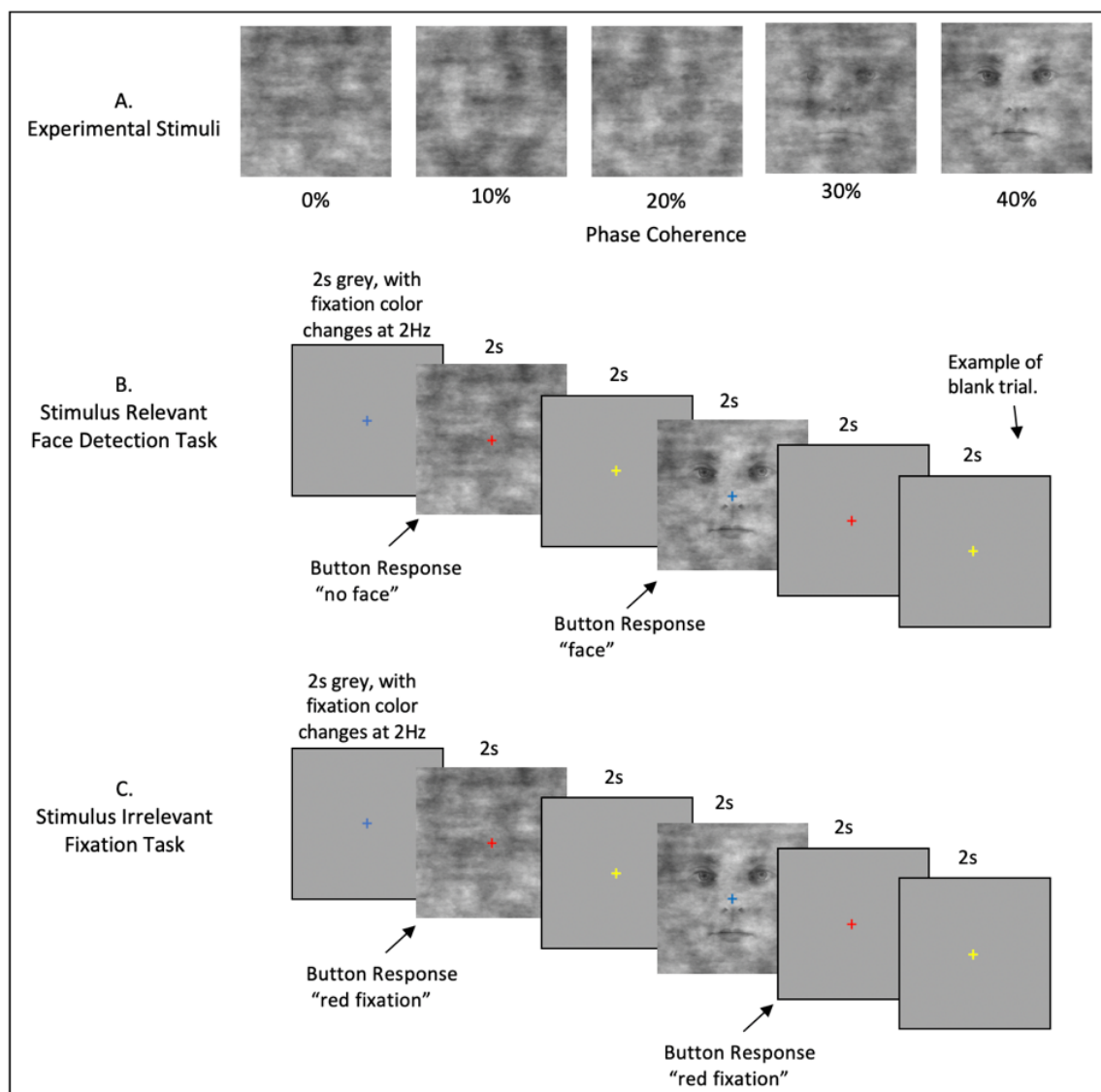351   *Perception-Based Analyses*

17

352   We performed additional analyses to assess BOLD amplitude modulations as a function of

353   percept. Due to the grouping of responses, we considered only the most ambiguous condition (i.e. 20%

354   phase coherence, see results). We allocated each trial to one of 2 new conditions: "*face percept*" and

355   "*no face percept*" on the basis of each participants reported percept. We then repeated the analysis

356   described in section 2.5.4 to estimate BOLD amplitudes of these 2 conditions.

357

358   **Results**

359   Through this work, the prefixes "l" and "r" preceding the name of a ROI will refer to its

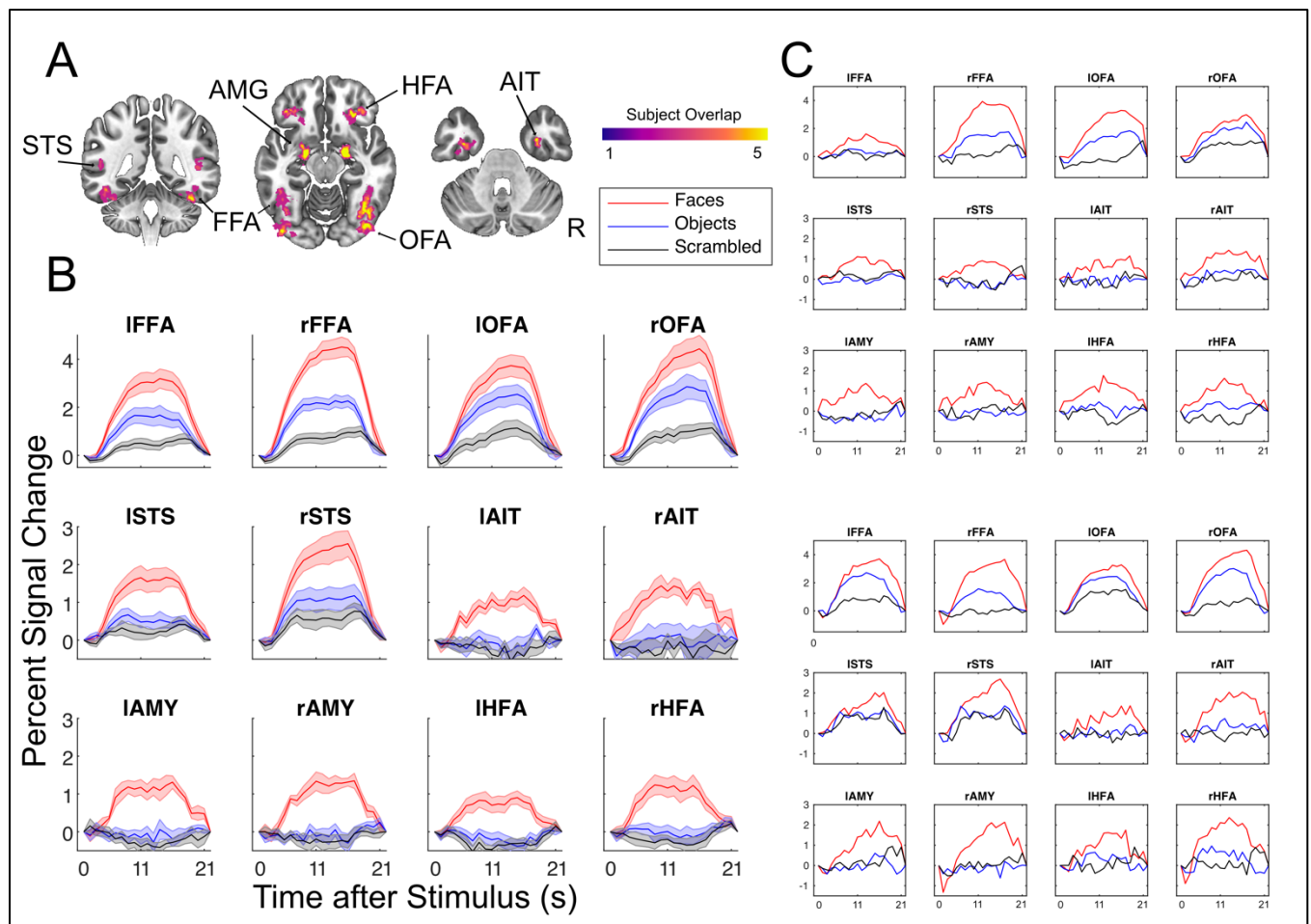360   hemispheric lateralization.

361

362   _Localizer Task._

363   Using the separate face-localizer task, we identified a total of twelve ROIs for all participants. This

364   included typical regions, such as the fusiform face area (FFA, MNI Centers of Mass [L: -41 -52 -19, R: 40 -

365   54 -17]), occipital face area (OFA [L: -44 -83 -13, R:43 -76 -11]), posterior superior temporal sulcus (STS

366   [L: -48 -68 10, R: 50 -57 18]) as well as the amygdalae [L: -21 -7 -15, R: 20 -6 -15]. In addition, we identified

367   ROIs for more difficult regions, including an ROI proximal to the perirhinal cortex, which we are referring

368   to more generally as the anterior inferior temporal (AIT [L: -34 -9 -34, R:32 -6 -39]) cortex as well as an

369   ROI in H-shaped sulcus, which we refer to as the HFA (H-shaped sulcus Face Area, [L: -32 33 -15, R: 30 33

370   -14]). ROIs identification and all subsequent analyses were carried out in native subject space, but to

371   compare their locations across participants and to relate these locations to previous studies, we

372   converted and report their coordinates in MNI space. All cortical regions and their overlap (in the specific

373   slices selected) across participants are shown in Figure 2A.

374

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 inter-trial 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.

384

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. Labels indicate ROI names, right is right. **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 of the Mean shown with shading. **C)** Example results from two single subjects showcasing reliability even within a single subject's ROIs.

Figure 2B shows ROIs average BOLD time courses, constructed from FIR models. All regions, including areas associated with low SNR (AIT, HFA), yield plausible hemodynamic responses. Furthermore, these responses remain HRF-like even at the single subject level (Figure 2C), indicating high data quality.
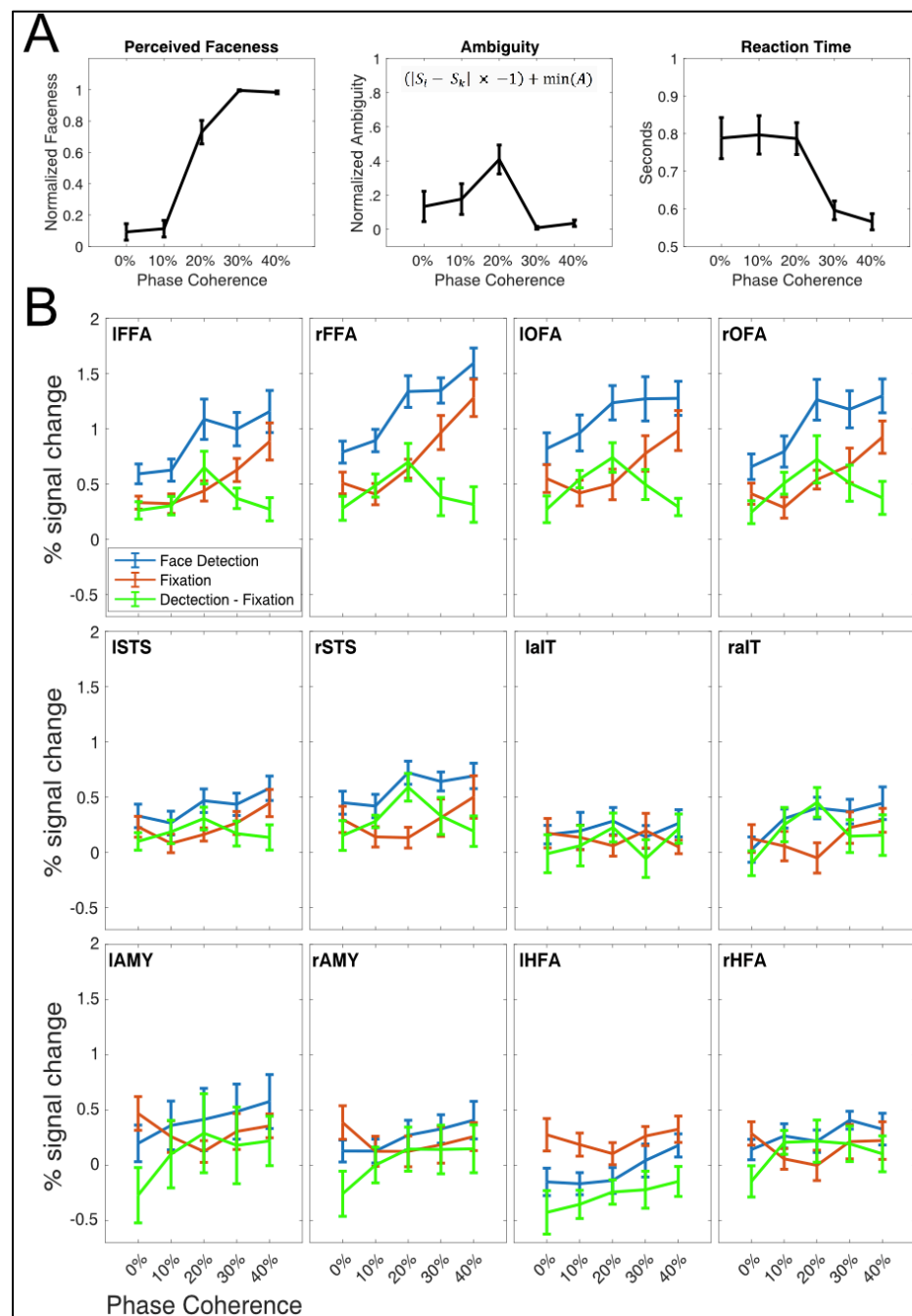
Behavioral Responses and Ambiguity Scores

21

398   The right panel in Figure 3A portrays the face detection group average reaction times. Subjects

399   responded with an average RT of 760 +/- 120 ms. Reaction times indicate that, in reporting their

400   percepts, participants were slowest at the lower end of the phase coherence spectrum (i.e. 0% and 10%),

401   becoming increasingly faster as a function of phase coherence.

402   Figure 3A, left panel, shows the average perceived "faceness", that is the proportion of the time

403   that participants reported seeing a face. By calculating the inverse of the absolute distance from the

404   inflection point (see section 4.4.4.) of the faceness sigmoid, we derive the ambiguity scores (Figure 3A,

405   middle panel), which peaked at 20% phase coherence. Reaction times were shortest for the highest

406   phase coherence (Figure 3A, right panel).

407

Figure 3. A) 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. Ambiguity was computed according to the equation shown in the middle panel, where S_i is the ith 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. **Right Panel:** Group averaged reaction times during the face detection task. For all panels, error bars represent standard errors across subjects. **B)** 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.

23

421

422   *Face and Fixation fMRI analyses*

423   *ANOVA Results*

424   The subjects mean percent signal change in response to each condition, for each ROI, is shown

425   in Figure 3B.

426   The main effect of the task was significant ($p < 0.05$) in the lFFA ($F_{1,9} = 20.559$), rFFA ($F_{1,9} =$

427   13.131), lOFA ($F_{1,9} = 48.595$), rOFA ($F_{1,9} = 14.304$) and the rSTS ($F_{1,9} = 9.182$), with face detection driving

428   larger BOLD responses relative to the fixation task.

429   There was a significant ($p < 0.05$) main effect of the condition (phase coherence level) in the lFFA

430   ($F_{4,36} = 13.136$), rFFA ($F_{4,36} = 35.567$), lOFA ($F_{4,36} = 7.56$), rOFA ($F_{4,36} = 22.866$), lSTS ($F_{4,36} = 7.98$), rSTS ($F_{4,36}$

431    $_= 5.637$), the rAIT ($F_{4,36} = 4.174$). For these ROIs, paired sample post-hoc t-test ($p<.05$, corrected) showed

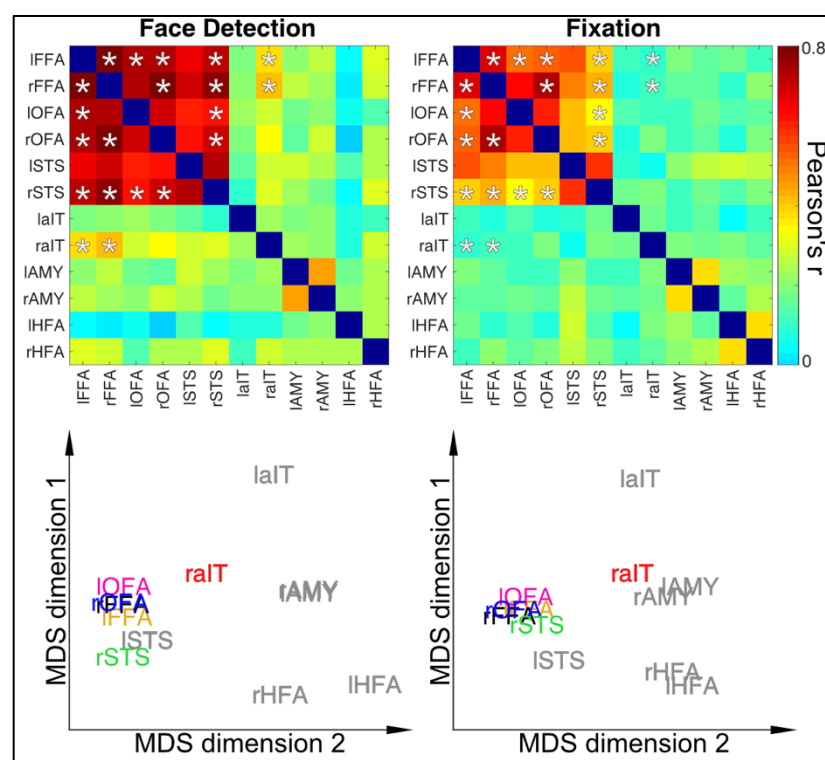432   that the 40% phase coherence is always significantly larger than 0%.

433   There was a significant ($p < 0.05$) task by condition interaction term in the lFFA ($F_{4,36} = 4.96$), rFFA

434   ($F_{4,36} = 3.311$), the rSTS ($F_{4,36} = 3.620$) and the rAIT ($F_{4,36} = 3.84$), indicating that amplitude increase during

435   detection relatively to fixation was different for different phase coherence levels. Post-hoc t-tests carried

436   out across tasks, within each condition revealed that for these ROIs only the 20% phase coherence

437   conditions were always significantly ($p<.05$ corrected) larger than all other conditions. The t-values (lFFA

438   ($t(9) = 4.407$), rFFA ($t(9) = 4.132$), rSTS ($t(9) = 4.684$) and the rAIT ($t(9)=3.376$)) and related effect sizes

439   (Hedges g*: lFFA: 1.361; rFFA: 1.795; rSTS: 1.791; and rAIT: 1.151) further indicate reliable and replicable

440   effects with the current N = 10.

441

442   *Functional Connectivity Measures*

24

443   Functional connectivity significantly increased (p <0.05, corrected) between multiple areas

444   during the face detection task relative to fixation in multiple core face processing areas. Specifically, we

445   found significant increases in functional connectivity between the 1) lFFA and the rFFA, right and lOFA,

446   rSTS and rAIT, 2) rFFA and rOFA, rSTS and rAIT, 3) the lOFA and the rSTS 4) the rOFA and the rSTS

447   (symmetrical connectivity between regions not repeated).
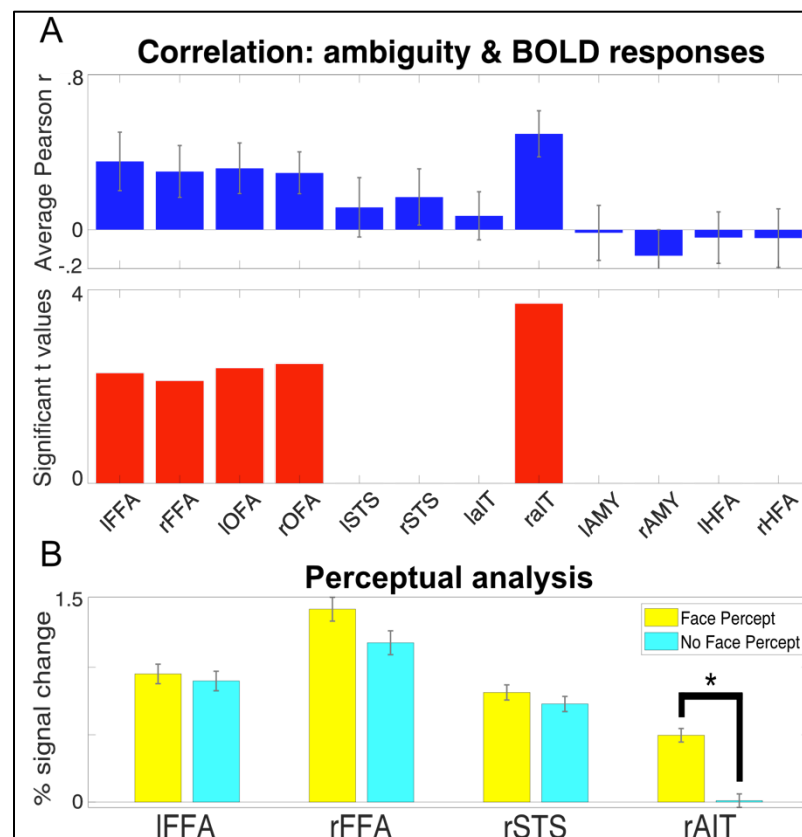

448

Figure 4. 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.

456

457   Multidimensional scaling (MDS), which is a dimensionality reduction technique that allows

458   visualizing the level of similarity amongst data points, was used to summarize task demands related

459   modulations in connectivity. MDS spatial arrangement, where proximity amongst point indicates

460   similarity of responses, highlights how the rAIT is more closely located to the FFAs (i.e. significantly more

461   correlated) during face detection relative to fixation (see Figure 4 Bottom; multidimensional space of

462   fixation rotated onto that of detection using Procrustes transformation).

463

464   *Brain-Behavior Correlations*

465   Figure 5A shows the correlation between the ambiguity score and the difference in the measured

466   brain activity between the stimulus-relevant (SR) and stimulus-irrelevant (SI) tasks. Correlations where

467   significant (p<0.05, corrected) in the bilateral FFA and OFAs and the rAIT.

468

Figure 5. A) Brain-Behavior Correlations in FFAs, OFAs and rAIT. **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 rAIT, the t values are plotted. **B)** 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.

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

485   corrected; see figure 5B). Moreover, paired sample t-tests contrasting the activation of each condition

486   for all ROIs against 0 further indicated significant (p<.05 corrected) above baseline activation for all ROIs

487   and conditions, except for the "no face percept" response in the rAIT, which did not significantly differ

488   from baseline.

489

490

Discussion

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, see Bokde et al., 2005; Druzgal and D'Esposito, 2001; Egner and Hirsch, 2005; Kay and 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"*.

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

513 Our finding that the specificity of task-related modulations is dependent on the nature of the

514 task is consistent with a large body of literature highlighting the conceptual and anatomical differences

515 between different types of attention. Posner and Peterson (1990) delineated 3 attention subsystems

516 devoted to orienting, detection, and alertness. *Our results suggest that a design in which domain-*

517 *specificity difficulty is varied is capable of revealing the distinct modulation associated with different*

518 *attentional subsystems.* The significant (p<.05 corrected) main effect of task, indicates, in accordance

519 with previous reports (Druzgal and D'Esposito, 2001; Kay and Yeatman, 2017; Vuilleumier et al., 2001;

520 Wojciulik et al., 1998) that BOLD amplitude is on average larger during the stimulus relevant (SR) relative

521 to stimulus irrelevant (SI) task in canonical face processing areas: the bilateral FFA, OFA and the rSTS. In

522 line with previous work, (Druzgal and D'Esposito, 2001; Kay and Yeatman, 2017; Ress et al., 2000), we

523 suggest that these increases reflect broad top-down contributions relating to increased vigilance and

524 awareness of the stimuli, which have previously been shown to improve decoding performance (Dobs

525 et al., 2018).

526 By contrast, there are a specific subset of areas that show an interaction between task and

527 stimulus: the bilateral FFA as well as the rAIT and rSTS. Of these only the rAIT showed a significant

528 interaction, driven by the activity elicited by the 20% phase coherence images being larger during the SR

529 compared to the SI. This result indicates that, at least within the context of this work, this region's

530 responses are *exclusively domain-specific top-down modulations*. In this region in fact, other than for the

531 20% phase coherence, task demands do not alter BOLD amplitude to any other condition (Figure 3B).

532 Finally, a large number of regions, specifically the FFAs, OFAs, STS, and rAIT showed a main effect

533 of condition, driven by larger amplitude for the 40% phase relative to the 0% phase. As this comparison

534 is equivalent to a classical 'face vs scrambled face' linear contrast (Chen et al., 2007; Rossion et al., 2012)

535   used to define stimulus selectivity, and consistent with the preferentiality of these regions for this

536   stimulus category, this result is unsurprising and won't be discussed further.

537

538   *Ambiguity*

539   Post-hoc t-tests carried out on the significant task x condition interactions revealed that in all

540   cortical regions exhibiting specific top-down modulations, the stimulus relevant task amplitude increases

541   were always significant *only* for 20% phase coherence. Behaviorally, this condition was also found to be

542   the most perceptually ambiguous, i.e. the condition with the largest reported number of contrasting

543   percepts (i.e. "face" and "no face"). In the context of this study we mathematically defined ambiguity as

544   the inverse of the absolute difference from the inflection point of the sigmoid that describes the face

545   percept/detection behavioral responses (see Equation 1). Both extremes of the phase coherence

546   continuum therefore represent non-ambiguous percepts. That is, at low phase coherence (e.g. 0%)

547   participants consistently reported no face percept and at high coherence (40%) participants consistently

548   reported the presence of a face (Figure 3A). At the maximally ambiguous, 20% condition, the percept

549   was at its most unstable and thus most difficult to categorize, leading to a larger amplitude response.

550   This relative increase is consistent with prior work showing that task difficulty can modulate the BOLD

551   response not only in frontal or parietal regions (Culham et al., 2001; Gould et al., 2003) but also in these

552   category sensitive visual areas (Druzgal and D'Esposito, 2001).

553   Notably, the reaction times measured in this study do not appear to capture this difficulty

554   increase (Figure 3A). This disconnect between reaction time and experimental performance has been

555   noted before in relation to accuracy and attentional cueing (van Ede et al., 2012) or to task difficulty

556   across a wide range of N-back conditions (Lamichhane et al., 2020). In our data, this likely reflects a

31

557   disconnect between the idea of "task difficulty" and "difficulty in categorization" which may provide

558   separate contributions to total reaction time.

559

560   *Top-down driven network reconfiguration*

561   To understand how top-down/context modulated the inter-regional neural communication and

562   thus the connectivity within specialized cortical networks (here, the face network), we compared

563   functional connectivity between tasks (see paragraph 4.6.2). By computing functional connectivity on

564   concatenated FIR curves, we: 1) discard the contribution of ongoing activity (which is not the focus of

565   this specific work); and 2) increase statistical power and thus the reliability of our correlational metric.

566   Broadly speaking, connectivity was significantly greater in core and extended regions (i.e. between FFAs,

567   OFAs and rSTS) during the stimulus relevant task, indicating greater communication among these areas,

568   presumably to fulfill task demands. In particular, we observed significantly (p<.05 corrected) greater

569   connectivity between the rAIT and the FFAs during the face task relative to the fixation (Figure 4). A

570   number of studies have suggested a functional differentiation between the right and lFFA (Meng et al.,

571   2012; Rossion et al., 2000), with the former being more involced in holistic processing and the latter in

572   featural processing. It is therefore likely that the degraded stimuli used here (Figure 1) drive both

573   individual feature detection as well as holistic face processing; or that subjects flexibly adapt their

574   strategy depending on the available information, switching from featural (e.g. looking for eyes) to holistic

575   detection, thus engaging both FFAs.

576   Within the current experimental settings, our data suggest that, during face processing, the rAIT

577   can be recruited by the cortical face network to add an additional resource to resolve ambiguous

578   percepts. These results are consistent with the functional differentiation between core and extended

579   cortical face networks, according to which the core system mediates the representation of more basic

580   aspects of face processing, while the extended system is involved in higher level cognitive functions and

581   can be selectively recruited to act in concert with the regions in the core system (Haxby et al., 2000).

582   Moreover, our results suggest that network connectivity is not static, but flexibly adapts to the

583   contextual demands. This is in line with an ever growing body of literature advocating the plastic nature

584   of functional connectivity during task and at rest (Allen et al., 2014; Cribben et al., 2013; Debener et al.,

585   2006; Doucet et al., 2012; Hutchison et al., 2013; Sadaghiani et al., 2009) in response to cognitive and

586   behavioral demands (Gratton et al., 2018; Hutchison and Morton, 2016). Our results further expand

587   these views, indicating that even subtle changes in task demands, as those implemented here, can have

588   a dramatic impact over local network reconfigurations.

589

590   *Using Stimulus Ambiguity to Differentiate Functional Architecture*

591   Though it is challenging to establish causality, taken together, our results suggest that the rAIT is

592   the source of top-down modulation. First, the correlation between the difference of the BOLD signal

593   between tasks, and each participant's behavior (ambiguity score) was only significant in the bilateral

594   FFAs and OFAs and the rAIT. By using task differences in BOLD responses to identical stimuli, we sought

595   to highlight task specific top-down effects. This finding, indicating correspondence between the

596   ambiguity function and the difference in BOLD amplitude across tasks within these ROIs, suggests that

597   this ambiguity signal originates from one of these regions. As the rAIT is further along the information

598   hierarchy, it is the plausible source of this signal. Moreover, functional connectivity analysis shows

599   increased connectivity between the rAIT and both FFAs, between OFAs and FFAs, but not between the

600   rAIT and either of the OFAs. In addition, the FFAs show both non-specific (significant main effect of task)

601   and domain-specific (significant task x condition interaction) top-down effects, while the rAIT shows only

602   the task-specific, ambiguity related, top-down effects (significant interaction).

603　　　　　Though prior work has often implicated frontal regions or parietal areas as being a source for

604　　　these top-down signals when detecting objects such as faces or houses (Baldauf and Desimone, 2014;

605　　　Kay and Yeatman, 2017), these studies used indirect methods (i.e. 1-back; house or face) to examine

606　　　face detection. By manipulating difficulty within the context of face detection, we uncover a different,

607　　　within-network source of top-down modulation by directly stressing the face processing system. This

608　　　approach is similar to prior work finding modulation within the ventral temporal cortex when viewing

609　　　degraded facial stimuli (Fan et al., 2020).

610　　　　　Furthermore, there is evidence that the AIT is involved in resolving difficult stimuli in both

611　　　macaques and humans. For the former, prior work has implicated this region in learning ambiguous

612　　　stimulus rules related to concurrent discrimination (Bussey et al., 2002). Increases in ambiguity were

613　　　associated with worse performance in the macaques with perirhinal cortex (i.e. anterior inferior

614　　　temporal lobe) lesions. In humans, there is evidence that the AIT is associated with discriminating

615　　　individual face identities (Nasr and Tootell, 2012; Zhang et al., 2016), however these studies employed

616　　　tasks that did not necessarily tap directly into stimulus-specific dimensions. Though these studies offer

617　　　support for our findings, they are unable to disambiguate the relationship between behavior and

618　　　perception, as they either used a memory task, in which subjects could have used an image matching

619　　　strategy (Nasr and Tootell, 2012), or used only a fixation task (Zhang et al., 2016). Additional evidence

620　　　for the involvement of the AIT in ambiguous or difficult stimuli comes from outside the face perception

621　　　literature: this area is active when integrating conceptual and perceptual information (Martin et al.,

622　　　2018) and it is associated with identifying confusable objects (Clarke and Tyler, 2014; Tyler et al., 2013).

623　　　We therefore argue that the results presented here and those reported in previous work represent

624　　　indication that the ambiguity-related top-down signal during face detection originates in the rAIT, is fed

625　　　back to the FFAs and, from there, to the OFAs. Future work can explicitly assess this using specifically

34

626   tailored research methods that provide better evidence of causality, such as depth dependent fMRI

627   analyses (Huber et al., 2017), intracranial EEG or fMRI with dynamic causal modeling (Friston et al., 2019).

628

629   *Neural Correlate of Perception*

630   While FFAs and rSTS show both a main effect of task and a task by condition interaction, only the

631   latter was significant in the rAIT. To better understand these differences and further characterize the

632   response profile of these regions, we grouped BOLD responses according to each participant's percept

633   during the most ambiguous condition (i.e. 20% phase coherence). This analysis shows that, *for the rAIT*

634   *only*, responses elicited by face percepts are significantly larger than those elicited by no face percepts

635   (Figure 5B), indicating a high correspondence between neural and behavioral responses. Moreover,

636   unlike FFAs and rSTS, showing significant above baseline activation for the no face percept condition,

637   the rAIT shows no significant activation when participants failed to perceive a face (Figure 5B). Taken

638   together, these findings represent evidence for a direct link between the rAIT and subjective perception.

639   These observations are consistent with the idea that anterior temporal regions are involved in subjective

640   awareness (Li et al., 2014) and with reports that the AIT organizes visual objects according to their

641   semantic interpretations (Price et al., 2017). These findings are also consistent with the data from the

642   face localizer (Figure 1) in which the rAIT shows the most preferential face response amongst these 4

643   regions, with objects and scrambled images not being significantly different from zero. These effects

644   were not found for the lAIT, suggesting a possible lateralization of these processes in humans. However,

645   in light of the low and inconsistent responses in lAIT, we advocate caution in interpreting these results.

646   Further work is therefore required to characterize the functional role of this and other areas (i.e. HFAs

647   and amygdalae) that displayed comparable responses.

648　　　　The significant activation for the no face percept condition in the FFAs and rSTS can be related to

649　　　　the bottom-up physical properties of the stimulus: regardless of the perceptual state, the 20% phase

650　　　　coherence images always contained a face stimulus. Alternatively, this result can be due to the

651　　　　contextual top-down induced by task demands (here a face detection task). That is, as tasks were blocked

652　　　　by run, during SR runs, participants expected having to resolve perceptual judgment of an ambiguous

653　　　　stimulus, and therefore looked for a face or a face feature in every trial. This is consistent with the

654　　　　observation that, in the absence of a face, the FFA can be activated from contextual cues alone (Cox et

655　　　　al., 2004).

656　　　　　　　　Past work using ambiguous facial stimuli in the form of Mooney faces (Mooney, 1957)

657　　　　have found amplitude differences in the FFA (Andrews and Schluppeck, 2004) or latency differences in

658　　　　the rOFA (Fan et al., 2020). Alternative approaches using bistable perception, such as the face/vase

659　　　　illusion also report that the FFA shows greater activation when faces are perceived (Andrews et al., 2002;

660　　　　Hasson et al., 2001). With binocular rivalry, in which alternative images are shown to each eye, the FFA

661　　　　also increases in activity when faces are perceived (Tong et al., 1998). In our present work we found that

662　　　　the FFA did show a larger, however, non-significant increase when subjects reported seeing a face. This

663　　　　apparent discrepancy with prior reports can be explained by differences in experimental paradigms.

664　　　　Unlike Mooney faces, or face-related bistable stimuli, the 20% phase coherence stimuli used here always

665　　　　have a physical face present. For binocular rivalry, the perceptual ambiguity is similar to bistable

666　　　　perception, however binocular rivalry is ecologically implausible. The defining characteristic of these

667　　　　probes is a dynamic switching amongst percepts despite identical visual input. Here, we have instead

668　　　　focused on the ambiguity of the initial percept, which is more similar to approaches using degraded or

669　　　　partially occluded static stimuli (Flounders et al., 2019; Frühholz et al., 2011). With our stimuli and

670　　　　experimental manipulations, *we isolate a unique and distinctive signal only in the rAIT that distinguishes*

671　　　　*the subjective perception of a face.*

672   Another important implication of the findings presented is their interpretation in relation to non-

673   human primate research. Similar to humans, macaque monkeys possess a specialized cortical network

674   in inferotemporal cortex dedicated to processing faces (e.g. Tsao et al., 2003), however the exact

675   correspondence between human and macaque face-selective areas is still unclear (Tsao et al., 2008a,

676   2006, 2003). While a degree of structural and functional correspondence has been achieved with regards

677   to the core regions (e.g. macaque middle face patches to human FFA (Rajimehr et al., 2009; Tsao et al.,

678   2008a)), identifying a human homologue of the macaque anterior medial (AM) patch (the anterior most

679   patch) has been challenging. Tsao, Moeller, et al. (2008a) for example, failed to uncover a comparable

680   region in humans, attributing such failure to suceptibility related signal drop out due the putative

681   proximity of this area to the ear canal. Rajimehr and colleagues (Rajimehr et al., 2009) uncovered a face

682   specific region in human AIT in 5 out of 10 participants, but were unable to elucidate the nature of its

683   response properties. Here not only were we able to identify a human face selective region in AIT in all

684   subjects, but, importantly, we were able to define its response profile. Our results, linking the function

685   of the face AIT to subjective perception, are in line with recent electrophysiology reports, showing that

686   activity in the AM face patch is related to the animal's individual ability to detect a face (Moeller et al.,

687   2017).

688

689   *Benefits of High Field Imaging*

690   As briefly mentioned in the above paragraph, much of the difficulty in imaging regions such as

691   the AIT relates to low SNR associated because of susceptibility artifacts and inefficient RF transmit fields

692   in the ventral temporal lobes (Devlin et al., 2000). Moreover, areas such as the AIT are typically difficult

693   to image at more conventional (3T) field strengths due to low signal and are often not located in every

694   individual (e.g. 50% in Rajimehr et al., 2009), and often show large deviation from expected

695   hemodynamic responses (e.g. Ramon et al., 2015). These issues can persist independent of field

696   strength.

697   Here we used UHF fMRI to capitalize on higher SNR, CNR, and acceleration performance

698   maximizing fMRI sensitivity in these regions. It should be noted that moving to higher field alone is not

699   sufficient to guarantee increased image quality in these regions. At UHF, while BOLD signal changes and

700   SNR do increase, B0 and B1+ inhomogeneity also increase and have to be dealt with. Here the

701   combination of our experimental design, rigorous analytical approach, manual B0 shimming tailored to

702   all participants with a specific focus on anterior ventral temporal regions and flip angle optimizations

703   (see methods) yielded fruitful results. _For all our participants_ we report consistent and corresponding

704   regions in the anterior inferior temporal cortex on both the left and right that preferentially respond to

705   faces and that are modulated by task demands. This increased sensitivity that led to large effect size and

706   ROI identification in all subjects, stems from a combination of parameters and sequence optimization.

707   Specifically, in addition to the increased SNR that accompanies UHF strength, unlike 3T acquisitions,

708   where no effort has been put forward to for optimizing B0 and flip angles, we manually adjusted B0

709   inhomogeneity and flip angles to maximize SNR for each subject. Moreover, relative to previous human

710   work, where functional voxels measured > 3 mm iso (e.g. (Rajimehr et al., 2009; Ramon et al., 2015)),

711   here, we used 1.6 mm iso voxels, minimizing partial volume effects and spin dephasing, ultimately

712   reducing signal loss in dropout regions (Thanh Vu et al., 2017). While the increased signal due to higher

713   field strengths can be a benefit, using high field alone is insufficient and can be, in fact, at times

714   detrimental. Appropriate consideration of tradeoffs, such as increased $B0/B1^+$ instability, is necessary.

715

716   _Beyond the ventral temporal lobe: future directions_

38

717   Beyond the ventral temporal lobe, there have been reports of face sensitive regions in lateral

718   orbital regions in macaques (Barat et al., 2018; Hadj-Bouziane et al., 2008; Tsao et al., 2008b). However,

719   evidence for the existence of these areas in humans has been mixed, uncovering such regions in

720   approximately half (Troiani et al., 2016) or one-third of the study population (Troiani et al., 2019; Tsao

721   et al., 2008a). In the current work, we are able to identify these areas in all subjects during the localizer

722   portion of the study. Activation was proximal to the H-Shaped sulci, consistent with both prior reports

723   (Troiani et al., 2019) and across subjects, as visualized after normalization to MNI space (Figure 2). Like

724   the other ROIs in the face preferential cortical network, these regions exhibited positive responses to

725   faces and significantly (p< .05) greater activity to faces relative to scrambled stimuli or objects (See

726   Methods).

727   Although we did not observe meaningful stimulus relevant task modulations we did observe an

728   interesting, although non-significant, separation of 2 areas, with regard to their functional connectivity

729   during face detection: the rHFA exhibited a non-significant increase in activation, while its left

730   counterpart exhibited a non-significant decrease. These results should be interpreted with caution, as

731   the primary tasks produced low response amplitudes relative to the localizer. The absence of strong

732   activation in the HFAs, as well as other regions such as the amygdalae and left AIT, during the primary

733   task despite prominent responses in the localizer task is likely due to a number of differences between

734   the tasks. These include, but are not limited to, difference in presentation timings (i.e. 12 secs vs. 2 secs

735   on/off) and the fact that the localizer used stimulus presentations with non-degraded, full faces with

736   variable expressions and gaze directions.

737   That these task differences drove larger responses during the localizer task in the HFAs is

738   congruent with prior findings, namely that these areas have been suggested to be involved in social and

739   emotional aspects of face processing in non-human primates (Barat et al., 2018), aspects that, within the

740   primary task context, are irrelevant. Our failure to find task modulations is therefore consistent with the

741   more complex evaluative role of the frontal cortices (Noonan et al., 2012) and its engagement in social

742   perception tasks (Barat et al., 2018; Beer et al., 2006; Freeman et al., 2010; Mah et al., 2004).

743   Given the enlarged frontal cortices among primate species and in particular the highly developed

744   frontal areas in humans, it will be essential to perform further research that build on these differences

745   and further manipulate context, value and/or salience in order to elucidate the functional role of regions

746   such as the HFAs during face perception and processing.

747

Conclusion

Using faces, a well-studied stimulus underpinned by a dedicated, complex cortical network, in this study we combined high SNR fMRI with domain-specific attention and observed 2 types of top-down scaling mechanisms: 1) a broad gain effect related to drawing attention to the stimuli, manifesting as an unspecific amplitude increase across conditions; and 2) an additional scaling of specific conditions dictated by domain-specific task requirements. We explain the latter in terms of perceptual ambiguity and suggest that the ambiguity signal originates in the rAIT. Importantly, only in the rAIT is both preferentially active under challenging conditions and predictive of the subjects' perceptual judgments. We further show that subtle changes in task demands can lead to dramatic changes in network reconfigurations. Our results suggest that the combination of an explicit face detection and stimulus matched control task 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

762 **Acknowledgments**

766

767

768 **Competing Interests**

769   The authors have no financial or non-financial competing interests associated with this work.

43

References

Alexander GE, DeLong MR, Strick PL. 1986. Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annu Rev Neurosci* **9**:357–381. doi:10.1146/annurev.ne.09.030186.002041

Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. 2014. Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex* **24**:663–676. doi:10.1093/cercor/bhs352

Andrews TJ, Schluppeck D. 2004. Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *Neuroimage* **21**:91–98. doi:10.1016/j.neuroimage.2003.08.023

Andrews TJ, Schluppeck D, Homfray D, Matthews P, Blakemore C. 2002. Activity in the Fusiform Gyrus Predicts Conscious Perception of Rubin's Vase–Face Illusion. *Neuroimage* **17**:890–901. doi:10.1006/nimg.2002.1243

Arcaro MJ, Livingstone MS. 2017. Retinotopic organization of scene areas in macaque inferior temporal cortex. *J Neurosci* **37**:7373–7389. doi:10.1523/JNEUROSCI.0569-17.2017

Baldauf D, Desimone R. 2014. Neural mechanisms of object-based attention. *Science (80-)* **344**:424–427. doi:10.1126/science.1247003

Barat E, Wirth S, Duhamel JR. 2018. Face cells in orbitofrontal cortex represent social categories. *Proc Natl Acad Sci U S A* **115**:E11158–E11167. doi:10.1073/pnas.1806165115

Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST. 2011. Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci U S A* **108**:7641–7646. doi:10.1073/pnas.1018985108

791 Beer JS, John OP, Scabini D, Knight RT. 2006. Orbitofrontal cortex and social behavior: Integrating self-

792   monitoring and emotion-cognition interactions. *J Cogn Neurosci* **18**:871–879.

793   doi:10.1162/jocn.2006.18.6.871

794 Bokde ALW, Dong W, Born C, Leinsinger G, Meindl T, Teipel SJ, Reiser M, Hampel H. 2005. Task difficulty

795   in a simultaneous face matching task modulates activity in face fusiform area. *Cogn Brain Res*

796   **25**:701–710. doi:10.1016/j.cogbrainres.2005.09.016

797 Bussey TJ, Saksida LM, Murray EA. 2002. Perirhinal cortex resolves feature ambiguity in complex visual

798   discriminations. *Eur J Neurosci* **15**:365–374. doi:10.1046/j.0953-816x.2001.01851.x

799 Chen CC, Kao KLC, Tyler CW. 2007. Face configuration processing in the human brain: The role of

800   symmetry. *Cereb Cortex* **17**:1423–1432. doi:10.1093/cercor/bhl054

801 Chen NK, Dickey CC, Yoo SS, Guttmann CRG, Panych LP. 2003. Selection of voxel size and slice orientation

802   for fMRI in the presence of susceptibility field gradients: Application to imaging of the amygdala.

803   *Neuroimage* **19**:817–825. doi:10.1016/S1053-8119(03)00091-0

804 Clarke A, Tyler LK. 2014. Object-specific semantic coding in human perirhinal cortex. *J Neurosci* **34**:4766–

805   4775. doi:10.1523/JNEUROSCI.2828-13.2014

806 Cohen MR, Maunsell JHR. 2011. Using Neuronal Populations to Study the Mechanisms Underlying Spatial

807   and Feature Attention. *Neuron* **70**:1192–1204. doi:10.1016/j.neuron.2011.04.029

808 Cox D, Meyers E, Sinha P. 2004. Contextually Evoked Object-Specific Responses in Human Visual Cortex.

809   *Science (80-)* **304**:115–117. doi:10.1126/science.1093110

810 Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance

811   neuroimages. *Comput Biomed Res* **29**:162–173. doi:10.1006/cbmr.1996.0014

812   Cribben I, Wager TD, Lindquist MA. 2013. Detecting functional connectivity change points for single-

813   subject fMRI data. *Front Comput Neurosci* **7**. doi:10.3389/fncom.2013.00143

814   Culham JC, Cavanagh P, Kanwisher NG. 2001. Attention response functions: Characterizing brain areas

815   using fMRI activation during parametric variations of attentional load. *Neuron* **32**:737–745.

816   doi:10.1016/S0896-6273(01)00499-8

817   Cumming G. 2012. Understanding the new statistics: Effect sizes, confidence intervals, and meta-

818   analysis, Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.

819   New York, NY, US: Routledge/Taylor & Francis Group.

820   Dale AM, Fischl B, Sereno MI. 1999. Cortical Surface-Based Analysis. *Neuroimage* **9**:179–194.

821   doi:10.1006/nimg.1998.0395

822   De Martino F, Moerel M, Ugurbil K, Goebel R, Yacoub E, Formisano E. 2015. Frequency preference and

823   attention effects across cortical depths in the human primary auditory cortex. *Proc Natl Acad Sci U

824   S A* **112**:16036–16041. doi:10.1073/pnas.1507552112

825   Debener S, Ullsperger M, Siegel M, Engel AK. 2006. Single-trial EEG-fMRI reveals the dynamics of

826   cognitive function. *Trends Cogn Sci* **10**:558–563. doi:10.1016/j.tics.2006.09.010

827   Desimone R, Albright TD, Gross CG, Bruce C. 1984. Stimulus-selective properties of inferior temporal

828   neurons in the macaque. *J Neurosci* **4**:2051–2062. doi:10.1523/jneurosci.04-08-02051.1984

829   Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK. 2000. Susceptibility-

830   induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage* **11**:589–600.

831   doi:10.1006/nimg.2000.0595

832   Dobs K, Schultz J, Bülthoff I, Gardner JL. 2018. Task-dependent enhancement of facial expression and

833   identity   representations   in   human   cortex.   *Neuroimage*   **172**:689–702.

834   doi:10.1016/j.neuroimage.2018.02.013

835   Doucet G, Naveau M, Petit L, Zago L, Crivello F, Jobard G, Delcroix N, Mellet E, Tzourio-Mazoyer N,

836   Mazoyer B, Joliot M. 2012. Patterns of hemodynamic low-frequency oscillations in the brain are

837   modulated   by   the   nature   of   free   thought   during   rest.   *Neuroimage*   **59**:3194–3200.

838   doi:10.1016/j.neuroimage.2011.11.059

839   Druzgal TJ, D'Esposito M. 2003. Dissecting contributions of prefrontal cortex and fusiform face area to

840   face working memory. *J Cogn Neurosci* **15**:771–784. doi:10.1162/089892903322370708

841   Druzgal TJ, D'Esposito M. 2001. Activity in fusiform face area modulated as a function of working memory

842   load. *Cogn Brain Res* **10**:355–364. doi:10.1016/S0926-6410(00)00056-2

843   Egner T, Hirsch J. 2005. Cognitive control mechanisms resolve conflict through cortical amplification of

844   task-relevant information. *Nat Neurosci* **8**:1784–1790. doi:10.1038/nn1594

845   Fan X, Wang F, Shao H, Zhang P, He S. 2020. The bottom-up and top-down processing of faces in the

846   human occipitotemporal cortex. *Elife* **9**. doi:10.7554/eLife.48764

847   Farzaneh F, Riederer SJ, Pelc NJ. 1990. Analysis of T2 limitations and off-resonance effects on spatial

848   resolution   and   artifacts   in   echo-planar   imaging.   *Magn   Reson   Med*   **14**:123–139.

849   doi:10.1002/mrm.1910140112

850   Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex.

851   *Cereb Cortex* **1**:1–47. doi:10.1093/cercor/1.1.1

852   Fischl B, Liu A, Dale AM. 2001. Automated manifold surgery: Constructing geometrically accurate and

853   topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* **20**:70–80.

854   doi:10.1109/42.906426

855   Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D,

856   Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. 2002. Whole brain segmentation: Automated

857   labeling of neuroanatomical structures in the human brain. *Neuron* **33**:341–355.

858   doi:10.1016/S0896-6273(02)00569-X

859   Fischl B, Van Der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J,

860   Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. 2004. Automatically Parcellating the Human

861   Cerebral Cortex. *Cereb Cortex* **14**:11–22. doi:10.1093/cercor/bhg087

862   Fisher RA. 1915. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an

863   Indefinitely Large Population. *Biometrika* **10**:507. doi:10.2307/2331838

864   Flounders MW, González-García C, Hardstone R, He BJ. 2019. Neural dynamics of visual ambiguity

865   resolution by perceptual prior. *Elife* **8**. doi:10.7554/eLife.41861

866   Freeman JB, Rule NO, Adams RB, Ambady N. 2010. The neural basis of categorical face perception:

867   Graded representations of face gender in fusiform and orbitofrontal cortices. *Cereb Cortex*

868   **20**:1314–1322. doi:10.1093/cercor/bhp195

869   Freeman PR, Hedges L V., Olkin I, Freeman PR, Hedges L V., Olkin I, Rosenbaum PR, Hedges L V., Olkin I,

870   Freeman PR, Hedges L V., Olkin I. 1986. Statistical Methods for Meta-Analysis., Biometrics.

871   Academic Press. doi:10.2307/2531069

872   Friston KJ, Preller KH, Mathys C, Cagnan H, Heinzle J, Razi A, Zeidman P. 2019. Dynamic causal modelling

873   revisited. *Neuroimage* **199**:730–744. doi:10.1016/j.neuroimage.2017.02.045

874   Frühholz S, Godde B, Lewicki P, Herzmann C, Herrmann M. 2011. Face recognition under ambiguous

875   visual stimulation: fMRI correlates of "encoding styles." *Hum Brain Mapp* **32**:1750–1761.

876   doi:10.1002/hbm.21144

877   Gonzalez-Castillo J, Bandettini PA. 2018. Task-based dynamic functional connectivity: Recent findings

878   and open questions. *Neuroimage* **180**:526–533. doi:10.1016/j.neuroimage.2017.08.006

879   Gould RL, Brown RG, Owen AM, Ffytche DH, Howard RJ. 2003. fMRI BOLD response to increasing task

880   difficulty during successful paired associates learning. *Neuroimage* **20**:1006–1019.

881   doi:10.1016/S1053-8119(03)00365-3

882   Gratton C, Laumann TO, Nielsen AN, Greene DJ, Gordon EM, Gilmore AW, Nelson SM, Coalson RS, Snyder

883   AZ, Schlaggar BL, Dosenbach NUF, Petersen SE. 2018. Functional Brain Networks Are Dominated by

884   Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**:439-452.e5.

885   doi:10.1016/j.neuron.2018.03.035

886   Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role

887   in categorization. *Nat Rev Neurosci* **15**:536–548. doi:10.1038/nrn3747

888   Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RBH. 2008. Perception of emotional

889   expressions is independent of face selectivity in monkey inferior temporal cortex. *Proc Natl Acad*

890   *Sci U S A* **105**:5591–5596. doi:10.1073/pnas.0800489105

891   Handwerker DA, Ollinger JM, D'Esposito M. 2004. Variation of BOLD hemodynamic responses across

892   subjects and brain regions and their effects on statistical analyses. *Neuroimage* **21**:1639–1651.

893   doi:10.1016/j.neuroimage.2003.11.029

894   Hasson U, Hendler T, Bashat D Ben, Malach R. 2001. Vase or face? A neural correlate of shape-selective

895   grouping processes in the human brain. *J Cogn Neurosci* **13**:744–753.

896   doi:10.1162/08989290152541412

897   Haxby J V., Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping

898   representations of faces and objects in ventral temporal cortex. *Science (80-)* **293**:2425–2430.

899   doi:10.1126/science.1063736

900   Haxby J V., Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception.

901   *Trends Cogn Sci* **4**:223–233. doi:10.1016/S1364-6613(00)01482-0

902   Hedges L V. 1981. Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *J Educ*

903   *Stat* **6**:107. doi:10.2307/1164588

904   Hubel DH, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. *J Physiol*

905   **148**:574–591. doi:10.1113/jphysiol.1959.sp006308

906   Huber L, Handwerker DA, Jangraw DC, Chen G, Hall A, Stüber C, Gonzalez-Castillo J, Ivanov D, Marrett S,

907   Guidi M, Goense J, Poser BA, Bandettini PA. 2017. High-Resolution CBV-fMRI Allows Mapping of

908   Laminar Activity and Connectivity of Cortical Input and Output in Human M1. *Neuron* **96**:1253-

909   1263.e7. doi:10.1016/j.neuron.2017.11.005

910   Hutchison RM, Morton JB. 2016. It's a matter of time: Reframing the development of cognitive control

911   as a modification of the brain's temporal dynamics. *Dev Cogn Neurosci*, Flux Congress 2014 **18**:70–

912   77. doi:10.1016/j.dcn.2015.08.006

913   Hutchison RM, Womelsdorf T, Gati JS, Everling S, Menon RS. 2013. Resting-state networks show dynamic

914   functional connectivity in awake humans and anesthetized macaques. *Hum Brain Mapp* **34**:2154–

915   2177. doi:10.1002/hbm.22058

916   Kabbara A, EL Falou W, Khalil M, Wendling F, Hassan M. 2017. The dynamic functional core network of

917   the human brain at rest. *Sci Rep* **7**:2936. doi:10.1038/s41598-017-03420-6

918 Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate
919   cortex specialized for face perception. *J Neurosci* **17**:4302–4311. doi:10.1523/jneurosci.17-11-
920   04302.1997

921 Kanwisher N, Yovel G. 2006. The fusiform face area: A cortical region specialized for the perception of
922   faces. *Philos Trans R Soc B Biol Sci* **361**:2109–2128. doi:10.1098/rstb.2006.1934

923 Kay KN, Yeatman JD. 2017. Bottom-up and top-down computations in word- and face-selective cortex.
924   *Elife* **6**:e22341. doi:10.7554/eLife.22341

925 Lakens D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer
926   for t-tests and ANOVAs. *Front Psychol* **4**:863. doi:10.3389/fpsyg.2013.00863

927 Lamichhane B, Westbrook A, Cole MW, Braver TS. 2020. Exploring brain-behavior relationships in the N-
928   back task. *Neuroimage* **212**:116683. doi:10.1016/j.neuroimage.2020.116683

929 Laughlin SB, Sejnowski TJ. 2003. Communication in neuronal networks. *Science (80-)* **301**:1870–1874.
930   doi:10.1126/science.1089662

931 Leung AWS, Alain C. 2011. Working memory load modulates the auditory "What" and "Where" neural
932   networks. *Neuroimage* **55**:1260–1269. doi:10.1016/j.neuroimage.2010.12.055

933 Lewis LD, Setsompop K, Rosen BR, Polimeni JR. 2018. Stimulus-dependent hemodynamic response
934   timing across the human subcortical-cortical visual pathway identified through high spatiotemporal
935   resolution 7T fMRI. *Neuroimage* **181**:279–291. doi:10.1016/j.neuroimage.2018.06.056

936 Li Q, Hill Z, He BJ. 2014. Spatiotemporal dissociation of brain activity underlying subjective awareness,
937   objective performance and confidence. *J Neurosci* **34**:4382–4395. doi:10.1523/JNEUROSCI.1820-
938   13.2014

939   Li X, Morgan PS, Ashburner J, Smith J, Rorden C. 2016. The first step for neuroimaging data analysis:

940   DICOM to NIfTI conversion. *J Neurosci Methods* **264**:47–56. doi:10.1016/j.jneumeth.2016.03.001

941   Liu B, Zhu T, Zhong J. 2015. Comparison of quality control software tools for diffusion tensor imaging.

942   *Magn Reson Imaging* **33**:276–285. doi:10.1016/j.mri.2014.10.011

943   Mah L, Arnold MC, Grafman J. 2004. Impairment of social perception associated with lesions of the

944   prefrontal cortex. *Am J Psychiatry* **161**:1247–1255. doi:10.1176/appi.ajp.161.7.1247

945   Martin CB, Douglas D, Newsome RN, Man LLY, Barense MD. 2018. Integrative and distinctive coding of

946   visual and conceptual object features in the ventral visual stream. *Elife* **7**:e31873.

947   doi:10.7554/eLife.31873

948   Martinez-Trujillo JC, Treue S. 2004. Feature-based attention increases the selectivity of population

949   responses in primate visual cortex. *Curr Biol* **14**:744–751. doi:10.1016/j.cub.2004.04.028

950   McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J

951   Cogn Neurosci* **9**:605–610. doi:10.1162/jocn.1997.9.5.605

952   Meng M, Cherian T, Singal G, Sinha P. 2012. Lateralization of face processing in the human brain. *Proc R

953   Soc B Biol Sci* **279**:2052–2061. doi:10.1098/rspb.2011.1784

954   Moeller S, Crapse T, Chang L, Tsao DY. 2017. The effect of face patch microstimulation on perception of

955   faces and objects. *Nat Neurosci* **20**:743–752. doi:10.1038/nn.4527

956   Moeller S, Freiwald WA, Tsao DY. 2008. Patches with links: A unified system for processing faces in the

957   macaque temporal lobe. *Science (80-)* **320**:1355–1359. doi:10.1126/science.1157436

958   Mooney CM. 1957. Age in the development of closure ability in children. *Can J Psychol* **11**:219–226.

959   doi:10.1037/h0083717

960   Narumoto J, Okada T, Sadato N, Fukui K, Yonekura Y. 2001. Attention to emotion modulates fMRI activity

961   in human right superior temporal sulcus. *Cogn Brain Res* **12**:225–231. doi:10.1016/S0926-

962   6410(01)00053-2

963   Nasr S, Tootell RBH. 2012. Role of fusiform and anterior temporal cortical areas in facial recognition.

964   *Neuroimage* **63**:1743–1753. doi:10.1016/j.neuroimage.2012.08.031

965   Noonan MP, Kolling N, Walton ME, Rushworth MFS. 2012. Re-evaluating the role of the orbitofrontal

966   cortex in reward and reinforcement. *Eur J Neurosci* **35**:997–1010. doi:10.1111/j.1460-

967   9568.2012.08023.x

968   Olman CA, Davachi L, Inati S. 2009. Distortion and signal loss in medial temporal lobe. *PLoS One* **4**.

969   doi:10.1371/journal.pone.0008160

970   Parker AJ, Newsome WT. 1998. SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception.

971   *Annu Rev Neurosci* **21**:227–277. doi:10.1146/annurev.neuro.21.1.227

972   Perrett DI, Rolls ET, Caan W. 1982. Visual neurones responsive to faces in the monkey temporal cortex.

973   *Exp Brain Res* **47**:329–342. doi:10.1007/BF00239352

974   Pessoa L, Gutierrez E, Bandettini P, Ungerleider L. 2002. Neural correlates of visual working memory:

975   fMRI amplitude predicts task performance. *Neuron* **35**:975–987. doi:10.1016/S0896-

976   6273(02)00817-6

977   Posner MI, Petersen SE. 1990. The Attention System of the Human Brain. *Annu Rev Neurosci* **13**:25–42.

978   doi:10.1146/annurev.ne.13.030190.000325

979   Price AR, Bonner MF, Peelle JE, Grossman M. 2017. Neural coding of fine-grained object knowledge in

980   perirhinal cortex. *bioRxiv* **194829**:1–18. doi:doi.org/10.1101/194829

981   Rajimehr R, Young JC, Tootell RBH. 2009. An anterior temporal face patch in human cortex, predicted by

982   macaque maps. *Proc Natl Acad Sci U S A* **106**:1995–2000. doi:10.1073/pnas.0807304106

983   Ramon M, Vizioli L, Liu-Shuang J, Rossion B. 2015. Neural microgenesis of personally familiar face

984   recognition. *Proc Natl Acad Sci U S A* **112**:E4835–E4844. doi:10.1073/pnas.1414929112

985   Ress D, Backus BT, Heeger DJ. 2000. Activity in primary visual cortex predicts performance in a visual

986   detection task. *Nat Neurosci* **3**:940–945. doi:10.1038/78856

987   Rossion B, Dricot L, Devolder A, Bodart JM, Crommelinck M, De Gelder B, Zoontjes R, Gelder B De,

988   Zoontjes R. 2000. Hemispheric asymmetries for whole-based and part-based face processing in the

989   human fusiform gyrus. *J Cogn Neurosci* **12**:793–802. doi:10.1162/089892900562606

990   Rossion B, Hanseeuw B, Dricot L. 2012. Defining face perception areas in the human brain: A large-scale

991   factorial fMRI face localizer analysis. *Brain Cogn* **79**:138–157. doi:10.1016/j.bandc.2012.01.001

992   Saad ZS, Reynolds RC. 2012. SUMA. *Neuroimage* **62**:768–773. doi:10.1016/j.neuroimage.2011.09.016

993   Sadaghiani S, Hesselmann G, Kleinschmidt A. 2009. Distributed and antagonistic contributions of ongoing

994   activity fluctuations to auditory stimulus detection. *J Neurosci* **29**:13410–13417.

995   doi:10.1523/JNEUROSCI.2592-09.2009

996   Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA,

997   Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen S, Power J, Salimi-

998   Khorshidi G, Snyder AZ, Vu AT, Woolrich MW, Xu J, Yacoub E, Uğurbil K, Van Essen DC, Glasser MF.

999   2013. Resting-state fMRI in the Human Connectome Project. *Neuroimage*, Mapping the

1000   Connectome **80**:144–168. doi:10.1016/j.neuroimage.2013.05.039

1001   Stigliani A, Weiner KS, Grill-Spector K. 2015. Temporal processing capacity in high-level visual cortex is

1002   domain specific. *J Neurosci* **35**:12412–12424. doi:10.1523/JNEUROSCI.4822-14.2015

1003   Szczepanski SM, Pinsk MA, Douglas MM, Kastner S, Saalmann YB. 2013. Functional and structural

1004   architecture of the human dorsal frontoparietal attention network. *Proc Natl Acad Sci U S A*

1005   **110**:15806–15811. doi:10.1073/pnas.1313903110

1006   Taylor AJ, Kim JH, Ress D. 2018. Characterization of the hemodynamic response function across the

1007   majority   of   human   cerebral   cortex.   *Neuroimage*   **173**:322–331.

1008   doi:10.1016/j.neuroimage.2018.02.061

1009   Thanh Vu A, Jamison K, Glasser MF, Smith SM, Coalson T, Moeller S, Auerbach EJ, Uğurbil K, Yacoub E.

1010   2017. Tradeoffs in pushing the spatial resolution of fMRI for the 7T Human Connectome Project.

1011   *Neuroimage*, Cleaning up the fMRI time series: Mitigating noise with advanced acquisition and

1012   correction strategies **154**:23–32. doi:10.1016/j.neuroimage.2016.11.049

1013   Tong F, Nakayama K, Vaughan JT, Kanwisher N. 1998. Binocular rivalry and visual awareness in human

1014   extrastriate cortex. *Neuron* **21**:753–759. doi:10.1016/S0896-6273(00)80592-9

1015   Troiani V, Dougherty CC, Michael AM, Olson IR. 2016. Characterization of face-selective patches in

1016   orbitofrontal cortex. *Front Hum Neurosci* **10**:1–14. doi:10.3389/fnhum.2016.00279

1017   Troiani V, Patti MA, Adamson K. 2019. The use of the orbitofrontal H-sulcus as a reference frame for

1018   value signals. *Eur J Neurosci* **n/a**:ejn.14590. doi:10.1111/ejn.14590

1019   Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH. 2003. Faces and objects in macaque

1020   cerebral cortex. *Nat Neurosci* **6**:989–995. doi:10.1038/nn1111

1021   Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. 2006. A cortical region consisting entirely of face-

1022   selective cells. *Science (80-)* **311**:670–674. doi:10.1126/science.1119983

1023   Tsao DY, Moeller S, Freiwald WA. 2008a. Comparing face patch systems in macaques and humans. *Proc*

1024   *Natl Acad Sci U S A* **105**:19514–19519. doi:10.1073/pnas.0809662105

1025   Tsao DY, Schweers N, Moeller S, Freiwald WA. 2008b. Patches of face-selective cortex in the macaque

1026   frontal lobe. *Nat Neurosci* **11**:877–879. doi:10.1038/nn.2158

1027   Tyler LK, Chiu S, Zhuang J, Randall B, Devereux BJ, Wright P, Clarke A, Taylor KI. 2013. Objects and

1028   categories: Feature statistics and object processing in the ventral stream. *J Cogn Neurosci* **25**:1723–

1029   1735. doi:10.1162/jocn_a_00419

1030   van Ede F, de Lange FP, Maris E. 2012. Attentional cues affect accuracy and reaction time via different

1031   cognitive and neural processes. *J Neurosci* **32**:10408–10412. doi:10.1523/JNEUROSCI.1337-12.2012

1032   Vogels R, Orban GA. 1994. Activity of inferior temporal neurons during orientation discrimination with

1033   successively presented gratings. *J Neurophysiol* **71**:1428–1451. doi:10.1152/jn.1994.71.4.1428

1034   Vuilleumier P, Armony JL, Driver J, Dolan RJ. 2001. Effects of attention and emotion on face processing

1035   in the human brain: An event-related fMRI study. *Neuron* **30**:829–841. doi:10.1016/S0896-

1036   6273(01)00328-2

1037   Wang SSH, Shultz JR, Burish MJ, Harrison KH, Hof PR, Towns LC, Wagers MW, Wyatt KD. 2008. Functional

1038   trade-offs in white matter axonal scaling. *J Neurosci* **28**:4047–4056. doi:10.1523/JNEUROSCI.5559-

1039   05.2008

1040   Wilcox R. 2005. Introduction to Robust Estimation and Hypothesis Testing, Technometrics. Academic

1041   Press. doi:10.1198/tech.2005.s334

1042   Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW. 2010. Controlling low-level image

1043   properties: The SHINE toolbox. *Behav Res Methods* **42**:671–684. doi:10.3758/BRM.42.3.671

1044   Wojciulik E, Kanwisher N, Driver J. 1998. Covert visual attention modulates face-specific activity in the

1045   human fusiform gyrus: fMRI study. *J Neurophysiol* **79**:1574–1578. doi:10.1152/jn.1998.79.3.1574

1046   Young IR, Cox IJ, Bryant DJ, Bydder GM. 1988. The benefits of increasing spatial resolution as a means of

1047   reducing artifacts due to field inhomogeneities. *Magn Reson Imaging* **6**:585–590.

1048   doi:10.1016/0730-725X(88)90133-6

1049   Zhang H, Japee S, Nolan R, Chu C, Liu N, Ungerleider LG. 2016. Face-selective regions differ in their ability

1050   to classify facial expressions. *Neuroimage* **130**:77–90. doi:10.1016/j.neuroimage.2016.01.045

1051   Zhang RY, Kay K. 2020. Flexible top-down modulation in human ventral temporal cortex. *Neuroimage*

1052   **218**:116964. doi:10.1016/j.neuroimage.2020.116964

1053