eNeuro

Research Article: New Research | Cognition and Behavior

Following Eye Gaze Activates a Patch in the Posterior Temporal Cortex That Is Not Part of the Human "Face Patch" System

Segregation of gaze following and face vision

Kira Marquardt[a,*], Hamidreza Ramezanpour[a,b,c,*], Peter W. Dicke[a] and Peter Thier[a,d]

[a]Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research
[b]Graduate School of Neural and Behavioural Sciences, University of Tübingen, Tübingen, Germany
[c]International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Tübingen, Germany
[d]Werner Reichardt Centre for Integrative Neuroscience (CIN), University of Tübingen, Tübingen, Germany

Author Contributions: KM, HR, PWD and PT designed research; KM, PWD performed research; KM and HR analyzed data; KM, HR, PT wrote the paper.

Conflict of Interest: Authors report no conflict of interest.

*K.M. and H.R. contributed equally to this work.

Correspondence should be addressed to either Peter Thier, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Email: thier@uni-tuebingen.de or Hamidreza Ramezanpour, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Email: hamidreza.ramezanpour@gmail.com

1 **Following eye gaze activates a patch in the posterior temporal cortex that is not**

2 **part of the human "face patch" system**

3 Kira Marquardt[1,a], Hamidreza Ramezanpour[1,2,a,b,c], Peter W. Dicke[a], Peter Thier[2,a,d]

4 [a] Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research

5 [b] Graduate School of Neural and Behavioural Sciences, University of Tübingen, Tübingen,

6 Germany

7 [c] International Max Planck Research School for Cognitive and Systems Neuroscience, University

8 of Tübingen, Tübingen, Germany

9 [d] Werner Reichardt Centre for Integrative Neuroscience (CIN), University of Tübingen, Tübingen,

10 Germany

11 [1] Equal Contributions

12 [2] Address for correspondence: PT & HR, Department of Cognitive Neurology, Hertie Institute for

13 Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Email: thier@uni-tuebingen.de &

14 hamidreza.ramezanpour@gmail.com

15 Abbreviated Title: **Segregation of gaze following and face vision**

16

17 Author Contributions: KM, HR, PWD and PT designed research; KM, PWD performed research; KM and
18 HR analyzed data; KM, HR, PT wrote the paper.

20 Number of Figures: 8

21 Number of words for Abstract: 208

22 Number of words for Significance Statement: 111

23 Number of words for Introduction: 647

24 Number of words for Discussion: 1738

25 Conflict of Interest: Authors report no conflict of interest.

Abstract

Humans follow another person´s eye gaze to objects of interest to the other, thereby establishing "joint attention", a first step towards developing a theory of the other´s mind. Previous fMRI studies agree that a "gaze following patch" (GFP) of cortex close to the posterior STS is specifically implicated in eye gaze following. The location of the GFP is in the vicinity of the posterior members of the core face processing system that consists of distinct patches in ventral visual cortex, the STS, and frontal cortex, also involved in processing information on the eyes. To test if the GFP might correspond to one of the posterior "face patches", we compared the pattern of BOLD contrasts reflecting the passive vision of static faces with the one evoked by shifts of attention guided by the eye gaze of others. The viewing of static faces revealed the face patch system. On the other hand, eye gaze following activated a cortical patch (the GFP) with its activation maximum separated by more than 24mm in the right and 19mm in the left hemisphere from the nearest face patch, the superior temporal sulcus face area (STS-FA). This segregation supports a distinct function of the GFP, different from the elementary processing of facial information.

48 **Significance Statement**

49 Human observers follow another person´s eye gaze to objects and locations of interest to the

50 other one, thereby establishing "joint attention", a major step towards developing a theory of

51 the other´s mind. Previous fMRI studies agree that a patch of cortex around the posterior

52 superior temporal sulcus is specifically implicated in eye gaze following. This "gaze following

53 patch" is located in the same region as the posterior elements of the face patch system, also

54 extracting information on the eyes. Using fMRI we show that the gaze following patch is distinct

55 from the face patch system, supporting a role beyond the elementary processing of facial

56 information accommodated by the face patch system.

57

58

59

60

61

62

63

64

65

66

67

68

69 **Introduction**

70 Eye gaze, head, shoulder and trunk orientation are important examples of body cues that offer

71 compelling information on the object and location of interest to the other, drawing the

72 observer´s attention to the same object and location, thereby establishing "joint attention", a

73 first and major step toward developing a theory of the other´s mind (Baron-Cohen, 1994; Baron-

74 Cohen, 1995; Emery, 2000; Langton and Bruce, 2000; Shimojo et al., 2003). In humans, eye gaze

75 is arguably the most important social cue guiding the observer´s attention (Emery, 2000). A

76 precise localization of the relevant machinery has recently been provided by a number of

77 functional magnetic resonance imaging (fMRI) studies. This work has identified a circumscribed

78 region in the posterior temporal sulcus of both hemispheres, adjacent to the middle and

79 superior temporal gyri, often referred to as pSTS region or area (pSTS) or, more loosely the gaze

80 following patch (GFP) (Puce et al., 1998; Allison et al., 2000; Hoffman and Haxby, 2000; Pelphrey

81 et al., 2003; Pelphrey et al., 2004; Materna et al., 2008a; Laube et al., 2011). A cortical area

82 involved in macaque monkeys´ head gaze following, the monkey GFP, has recently been

83 described in a comparable cortical region that may eventually turn out to be homologous with

84 the human GFP in the pSTS (see (Marciniak et al., 2014)).

85 The extraction of eye gaze orientation requires knowledge of the orientation of the eyes relative

86 to the face and ultimately also knowledge about the orientation of the other´s face relative to

87 the observer and the world. This need to care about particular aspects of faces might suggest

88 that eye gaze following may build on information provided by the parts of cortex known to be

89 devoted to the processing of faces, including their constitutive elements such as the eyes,

90 Actually, this influence of the eyes is suggested by a number of studies that have demonstrated

91 that for instance information on identity and emotional expression, provided by the eye region,

4

92 not only influences perception but also the activity in distinct face patches (Fox and

93 Damjanovic, 2006; Chan and Downing, 2011).

94 Actually, the human GFP, lighting up in gaze perception tasks is located in close vicinity to face

95 selective areas in the ventral visual cortex. This raises the possibility that the GFP may actually

96 be one of the members of this face processing network that involves distinct elements in the

97 ventral visual cortex and frontal cortex, namely the occipital face area (OFA), the fusiform face

98 area (FFA), the STS face area (STS-FA) and the inferior frontal face area (IFG-FA) (Kanwisher et

99 al., 1997; Haxby et al., 2000; Tsao et al., 2008). These areas are interconnected and seem to be

100 devoted to particular aspects of faces. For instance, the FFA emphasizes the encoding of

101 constant aspects of the face underlying identity decisions (Grill-Spector et al., 2004). On the

102 other hand, the STS-FA, the face selective area closest to the known location of the GFP has

103 been shown to contribute to encoding changeable aspects of faces such as facial expression and

104 face orientation, the latter an aspect obviously important for gaze following (Puce et al., 1998;

105 Wicker et al., 1998). Could it be that the STS-FA is actually part of the machinery for gaze

106 following, rather than being confined to providing information on face orientation? In this case,

107 we would expect at least partial overlap between the GFP and the STS-FA. In view of the

108 interindividual variability in the location of the GFP and also the STS-FA, the question if the two

109 overlap or not requires testing the same subjects in gaze following and face perception tasks.

110 Using well controlled fMRI paradigms in the same set of subjects, we show that the two systems

111 are actually well-segregated, a finding that clearly indicates that the GFP accommodates a

112 functionality not found in the face selective areas, although most probably building on pertinent

113 information contributed by the latter.

114 **Material and Methods**

115   **Subjects and Instrumentation**

116   Eleven adult male and nine adult female subjects from an age range of 21 to 46 years (mean =

117   26 years, SEM = 5.5 years) participated in the current study. All participants were right-handed,

118   healthy and had normal or corrected to normal vision. Subject were provided with transparent

119   and comprehensible information about the study goals and the procedures involved and gave

120   their written consent. Participants ran a training behavioural session before an imaging session

121   to minimize errors inside the MRI scanner caused by potential misunderstanding of tasks rules

122   or a lack of practice.

123   The study was approved by the Ethics Review Board of Tübingen Medical School and was

124   carried out in accordance with the principles of the 1964 Declaration of Helsinki.

125   In the training session subjects' eye movements were recorded deploying a commercial Eye

126   Tracker (Chronos Vision C-ETD). During the imaging session subjects' heads were stabilized by

127   foam rubber to minimize residual head movements. The visual stimuli (32 x 24° visual angle)

128   were presented on a translucent screen using an LCD projector (NEC GT 950, 1024 × 768 pixels)

129   viewed by the subjects via a two-mirror system with 60cm distance between the translucent

130   screen and the subject´s eyes. During the imaging procedure a certified, MRI compatible Eye

131   Tracker (SMI iView X MRI-LR) was used to record the subjects´ eye movements. The recorded

132   eye movements were evaluated offline after the experiments.

133   **Visual Stimuli and Experimental Tasks**

134   The participants had to perform three tasks: the first one required the observer to extract the

135   portrait's eye-gaze direction and to make a saccade towards one out of a set of 5 spatial targets

136   which the portrayed "demonstrator" looked at (gaze following task). The second one also

137   required an indicative saccade to targets singled out by information provided by the same

138 demonstrator portraits. However, in contrast to the first task, a different rule applied. Now,

139 rather than following the demonstrator´s gaze, the observer was required to make a saccade to

140 the target which had the same colour as the portrayed demonstrator´s iris (colour matching

141 task). Note that the visual information provided in the two tasks was the same, i.e. in both tasks

142 the iris colour varied from trial to trial, adopting the distinctive colour of one of the 5 targets,

143 arranged on a horizontal line met by the demonstrator´s gaze axis. Using the same visual stimuli

144 for the gaze following task and the control task and requiring the same behavioural responses,

145 any differences in the associated BOLD responses would have to be caused by differences

146 between the cognitive operations induced by the two sets of cues. Finally, participants were

147 subjected to a third experiment that required fixation of a small dot while passively viewing

148 images of faces and non-face stimuli, centered on the fixation dot, not requiring any

149 behavioural response (passive face perception task).

150 The portraits used in the gaze following/ colour matching tasks (collectively referred to as the

151 "active tasks") were photographs of a female in front of a white background. She was either

152 looking directly straight into the camera (baseline fixation picture) or to one out of five dot

153 targets arranged on a horizontal board, 25° below the straight ahead axis in the fronto-

154 orthogonal plane with a visual angle of 12.5° between targets The digital photographs were

155 processed using Adobe Photoshop CS5 to replace the original background with a black and

156 white random dot pattern and to colour the portrait´s iris and the five targets with five different

157 colours (dark blue, light blue, green, light brown, dark brown).

158 The tasks were run in separate blocks. Each block started with a written task instruction on the

159 projection screen (either "gaze following" or "colour matching") present for 5 seconds. The

160   whole block lasted for 95 seconds and contained 10 trials. Each trial started with a baseline

161   fixation picture with direct gaze (lasted for 5 seconds), immediately followed by one out of five

162   possible portraits ("target portraits"), present for 4 sec, with the demonstrator´s gaze directed

163   at a specific target and exhibiting a distinct iris colour. Within one block these 10 trials were

164   sorted randomly. The whole experiment contained four sessions, each involving two blocks of

165   gaze following and two colour matching one after another.

166   During the presentation of the baseline fixation picture, the subjects were asked to fixate a

167   small dot with 0.3° visual angle radius presented between the demonstrator´s eyes oriented

168   straight ahead. This fixation dot was also present in the target portraits for the first 1 sec and

169   then turned off. The disappearance of the fixation point served as the "go"-signal for the

170   participants to perform their saccade to the target singled out by the prevailing rule (gaze

171   following vs. colour matching). The subjects had to stay with their eye-gaze on the chosen target

172   until the baseline fixation picture, now serving as go signal, appeared again (See Figure 1 and

173   Figure 2). Implementing this go signal seemed to be necessary in order to allow us to reveal

174   differences in BOLD signals between gaze following and colour matching. Otherwise possibly

175   dominating BOLD signals associated with undelayed saccades might have concealed the

176   differential BOLD activity associated with the preceding processes.

177   The stimuli deployed in the "passive face perception" task (in short "passive task") were

178   photographs of human faces (females and males), hands and bodies plus manmade objects of

179   daily life as well as food, each subtending 12° visual angle. Facial stimuli were taken from by the

180   Radboud Face Database (Langner et al., 2010), showing females and males with averted gazes.

181   Adobe Photoshop CS5 was used to create scrambled versions of all photographs and to replace

182   the backgrounds by the same black and white random dot background which was used in the

183 gaze following paradigm. Stimuli were presented in four sessions, each containing 10 blocks of

184 16 photographs. The sequence of the blocks was the same in each session, but photographs

185 were randomly distributed within a block. Each block lasted 38 seconds and started with the

186 presentation of a fixation dot in front of a black and white random dot background for 5 sec,

187 followed 16 photographs (presentation-time = 1 sec each) with black screens present for 0.2 sec

188 in between. During presentation subjects were asked to maintain fixation of a small dot in the

189 middle of the screen while viewing the photographs.

190 **MRI-Imaging and preprocessing**

191 A 3 Tesla MR-Scanner (Siemens Magnetom Trio Tim syngo MR B17) was used to scan subjects'

192 brains. We used a T2*-weighted echo-planar sequence (TE = 35ms, TR = 3000ms, flip angle =

193 90°) covering the whole brain (44 transverse slices, matrix 64 x 64, slice thickness 2.5mm, in-

194 plane resolution 3x3) for image-acquisition during the experiments and a T1-weighted,

195 magnetization-prepared, rapid acquisition gradient-echo sequence (MP-RAGE with TE = 2.92 TR

196 = 2300ms, TI = 1100, flip angle = 8°, 176 x 256 x 256 voxel, voxel size 1.0 x 1.0 x 1.0mm) for the

197 structural, anatomical scans. A total of 945 images were taken from each subject.

198 The preprocessing and analysis of the MR-Images was done with the statistical parametric

199 mapping program package SPM8 (Wellcome Department of Cognitive Neurology, London, UK,

200 http://www.fil.ion.ucl.ac.uk/spm) running on Matlab 2013. Images of each subject were

201 reoriented by setting the origin to the anterior commissure and slice-time-corrected (number of

202 slices =44, TR = 3 sec, TA = 2.93, slice order = interleaved descending, reference slice = 22).

203 Functional scans were spatially realigned (registered to first and mean images resliced). The

204 anatomical scan was coregistered to the mean volume of the functional images and was

205 normalized to the Montreal Neurological Institute (MNI) space (Friston et al., 1995). Functional

206 images were normalized to the anatomical scan and then smoothed using a 7mm full width half

207 maximum gaussian filter. Time series in each voxel were high-pass-filtered with a cut-off-

208 frequency of 1/128Hz.

209 **MRI Data Analysis**

210 To estimate the BOLD activation patterns associated with the experimental tasks we assumed a

211 standard hemodynamic response function, reflecting the task variables according to a general

212 linear model (GLM). In the active task, the onset of the portrait defined time zero of the ensuing

213 event trace. We distinguished three different event types: fixation, gaze following and colour

214 matching. In the passive task, the appearance of the first image in each block determined time

215 zero of an event trace spreading across the whole block.

216 The estimated head-movements of the subjects during the sessions were considered as

217 regressors of no interest in the GLM model in addition to covariates of interests (the

218 experimental conditions: fixation, gaze following, colour matching, faces and non-faces). For the

219 active tasks the following contrasts were calculated for each subject: the response to gaze

220 following and colour matching versus baseline fixation and the response to gaze following

221 versus colour matching and vice versa. For the passive task, contrasts between responses to

222 faces and all-nonface stimuli including the scrambled faces were calculated. T-statistics were

223 deployed to identify significant changes (p<0.0001 for the active task and a more conservative

224 threshold of p<0.001 for the passive task taking into account its lower statistical power) in the

225 BOLD-signal at the level of individual subjects. In order to test whether results obtained for

226 individual subjects are valid at the population level, we carried out a second level analysis,

227 deploying a random effects model, comparing the average activation for a given voxel with the

228 variability of that activation over the examined population (Friston et al., 1999). The average

229   activation for a given voxel was taken as significant if the probability p provided by t-statistics

230   fell below 0.0001 (uncorrected) for that voxel and in at least 6 neighbouring ones. To optimally

231   visualize and measure the cortical representations, statistical t-maps were projected onto

232   inflated and flattened reconstructions of cortical surface gray matter using Caret

233   (http://brainvis.wustl.edu/wiki/index.php/Caret: About).

234   **Results**

235   **Behavioral findings**

236   In the active experiment, participants were instructed to identify the target either by following

237   the portrait's eye-gaze (gaze following) or, alternatively, to identify it based on a colour match

238   with the iris of the portrayed demonstrator and to execute a saccade to the target. In the first

239   case, eye colour and in the second case eye gaze direction had to be discounted. The two

240   variants of the active task did not differ with respect to the visual information available or the

241   oculomotor behaviour prompted but with regard to the cognitive strategy required to solve the

242   task. One might argue that the two different strategies to be pursued might have been

243   associated with different levels of difficulty and, consecutively, also different subjective task

244   loads. This did not seem to be the case as task performance was very similar. Participants

245   performed the task in the scanner with high accuracy well above chance level (20%) in the gaze

246   following condition [correct responses: mean = 87%, SEM = 11%] as well as in the colour

247   matching condition [correct responses: mean = 88%, SEM = 11%]. Kolmogorov-Smirnov test

248   showed that reaction times and correct responses showed a normal distribution. A paired t-test

249   showed no significant difference in the number of correct responses (p=0.61) or in reaction

250   times (p=0.32) between the two conditions [gaze following reaction time: mean = 711ms, SEM =

251   366ms; and colour matching reaction time: mean = 736ms, SEM = 341ms] (Figure 3).

252 **BOLD responses to gaze following and colour matching**

253 To identify brain areas activated during gaze following we looked at the contrast of gaze

254 following versus baseline fixation in a second level analysis of the group data. This comparison

255 delineated several brain areas in both hemispheres which had a significantly higher BOLD signal

256 (p<0.0001, in a cluster of 6 connected voxels each) (see Figure 4), among them dorsolateral

257 prefrontal cortex, premotor cortex, the supplementary motor area, cuneus, precuneus, fusiform

258 gyrus, posterior middle temporal gyrus, inferior temporal gyrus, middle occipital gyrus,

259 clustrom, middle frontal gyrus, inferior parietal lobule, superior parietal lobule, supramarginal

260 gyrus, precentral gyrus, cingulate gyrus, superior frontal gyrus, lingual gyrus, superior occipital

261 gyrus, parahippocampal gyrus and cerebellum. This pattern was very similar to the one obtained

262 when calculating the colour matching versus baseline fixation contrast (Figure 5). The close,

263 qualitative match between the patterns associated with the two tasks is not unexpected, given

264 the fact that both require the extraction of specific cues from faces to localise distinct objects in

265 order to shift one´s attention to them.

266 In order to identify cortical regions specifically or more strongly activated by the need to exploit

267 gaze direction we calculated the BOLD contrast between gaze following and colour matching. A

268 significant contrast (statistical criteria as before) was found in a patch of cortex bilaterally in the

269 posterior part of the middle and inferior temporal gyrus specifically with the peak contrast at

270 Talaraich coordinates right (50,-64,2) and left (-54,-67,6) (See Figure 6). This location of activity

271 is similar to gaze following and gaze processing related activity found in previous fMRI studies

272 (Hoffman and Haxby, 2000; Hooker et al., 2003; Pelphrey et al., 2005b; Materna et al., 2008a).

273 We will refer to the activated patch as the gaze following patch (GFP) and the cortical region in

274 which it is located as the posterior STS (pSTS).

12

275 **BOLD responses to the passive vision of faces**

276 We identified cortex activated by the passive vision of static faces by delineating regions for

277 which the contrast faces vs. non-face objects (biological as well as non-biological objects and

278 scrambled faces were pooled) was significant in the second level analysis (p<0.001, uncorrected,

279 6 connected voxels). In accordance with previous studies (Ishai et al., 2005; Gobbini and Haxby,

280 2006; Fox et al., 2009) we found significant BOLD contrasts in the mid fusiform gyrus bilaterally

281 (these voxels are the FFA), the right inferior occipital gyrus (these voxels form the OFA), the

282 posterior superior temporal sulcus bilaterally (these voxels correspond to the STS-FA) as well as

283 in the right inferior frontal gyrus (these are the voxels which make the IFG-FA). The highest

284 BOLD contrast to faces was identified in the functionally defined STS-FA, located at Talaraich

285 coordinates right (51,-42,12) and left (-57,-48,8). After identifying the face-selective regions in

286 the second level analysis, the BOLD time series underwent spatial smoothing with a 8-mm

287 FWHM gaussian blur and the clusters of face selective regions were extracted as a mask

288 mapped on the cortical surfaces in order assess their spatial relationship to the GFP later on.

289 **The pSTS gaze following patch and the face patch are segregated**

290 The fact that the GFP and the STS-FA, exhibiting the strongest BOLD contrast were found in the

291 same posterior part of the STS suggested that the two might overlap or, eventually, be even

292 fully congruent. To investigate this possibility we projected the two GFP and the face patches,

293 including the one in the pSTS region onto an inflated 3D representation of cortical surfaces using

294 the PALS-B12 atlas of human cerebral cortex (Van Essen, 2005). This rendering did not exhibit

295 any indication of overlap between the gaze following patch and any of the face selective

296 regions. Actually, the boundaries of the GFP and the ones of the nearest STS-FA were separated

297 by a gap of 4 mm (Figure 7). We next defined the GFP and the STS-FA as our region of interests

298 (spheres with the diameter of 5 mm centered on the coordinates of the peak activities in these

299 two areas in each individual subjects in order to compare the response levels as captured by the

300 contrast values for passive perception of static faces and gaze following. As shown in Figure 8,

301 the average contrast values in the GFP for the passive face perception task did not differ

302 significantly from zero (t-test, p=0.20), meaning that there was no selectivity for faces. Likewise,

303 the mean contrast values in the STS-FA during gaze following did not differ significantly from

304 zero (t-test, p=0.49), correspondingly expressing a lack of selectivity to gaze following. Hence,

305 we may conclude that the GFP and the STS-FA are neighbouring, yet non-overlapping areas with

306 different functions. In 6 out of 20 subjects we could not delineate a significantly activated GFP

307 and STS-FA at the level of the individual. Hence, these 6 subjects had to be excluded from a

308 comparison of gaze following related activity with activity in individual delineated STS-FA and

309 vice versa i.e. face selectivity test in the GFP. We also resorted to a conjunction analysis as an

310 alternative to a random effect analysis (Heller et al., 2007). This approach allows the assessment

311 of how many subjects exhibit selective activations in each voxel and therefore shows the extent

312 of overlap between gaze following related activity and activity evoked by static faces within and

313 across subjects. This analysis did not show any overlap in individual subjects, passing the

314 significance threshold of p<0.001 (uncorrected).

315 **Discussion**

316 With two separate fMRI experiments, performed on the same subjects, we tried to map the

317 cortical areas underlying gaze following and the establishment of joint attention and/or the

318 passive perception of static faces. The two experiments were run on the same subjects in order

319 to find out if the cortical structures involved overlap. In the first experiment, consisting of two

320 tasks, subjects were asked to either follow the eye gaze direction of portrayed demonstrators

321   towards distinct spatial targets or, alternatively, to shift attention to the target whose colour

322   corresponded to the one of the demonstrator´s iris. In accordance with previous work (Materna

323   et al., 2008a), we found that a gaze following patch (GFP) lighted up bilaterally in the posterior

324   part of the middle temporal gyrus when the BOLD signal evoked by eye gaze following was

325   contrasted with the BOLD signal evoked in the colour matching condition. Assuming that this

326   contrast is able to eliminate activity due to visual stimulation or the indicative saccades required

327   in both tasks, we may conclude that the neuronal machinery in the gaze following patch in the

328   pSTS might be responsible for the calculations needed to shift the observer´s attention based on

329   eye gaze. Unlike the shifts of attention evoked by our colour matching paradigm, gaze following

330   is reflexive (Friesen and Kingstone, 1998). However, this does not mean that it would not be

331   subject to cognitive control. Indeed, careful psychophysical experiments on monkey head gaze

332   following (Marciniak et al., 2015), probably homologous to human gaze following, clearly show

333   that with the exception of a small early reflex component, a substantial part of the gaze

334   following response can be suppressed. Hence, we can be confident that the BOLD contrast used

335   to identify the GFP reflects differences in gaze following related processing and its cognitive

336   control. Our paradigm vetoed an immediate behavioural response to the gaze cue as subjects

337   had to delay the response until the occurrence of the go signal. Hence, one might be concerned

338   that the GFP activity we observed in this experiment might differ from the normal pattern

339   evoked by spontaneous gaze following. However, the spatial coordinates of the GFP identified

340   here are in accordance with our previous findings on activations evoked by spontaneous gaze

341   following (Materna et al., 2008a).

342   In the second experiment, we used a classical static face localizer to map the face selective

343   regions, potentially involved in extracting information on face and eye gaze orientation in order

15

344   to clarify the anatomical relationship between the GFP and the members of this "face patch

345   system". Actually, we did not observe any overlap between the GFP and any of the face patches,

346   in particular not with a patch in the posterior STS (STS-FA), which in view of its localization as

347   described by previous work (Kanwisher et al., 1997; Haxby et al., 2000), might have been

348   expected to overlap with the GFP. One might argue that a lack of overlap between the two is

349   not surprising, given the fact that the GFP is orchestrating shifts of attention guided by the eyes,

350   i.e. just one out of many elements that make up faces and possibly not that influential in the

351   STS-FA. However, the following consideration speaks against the validity of this criticism. As

352   already shown by Wollaston in the 19th century (Wollaston, 1824), estimates of eye gaze

353   depend on concurrent information on the orientation of the face. And this latter information is

354   available in the GFP. This was shown by Laube et al. (Laube et al., 2011) who could establish that

355   the influence of head or face orientation on perceived eye direction, first described by

356   Wollaston finds its correlate in changes of the BOLD signal in the GFP. On the other hand,

357   previous fMRI work on face perception has suggested that one of the hallmarks of the STS-FA is

358   a stark interest in the changing aspects of faces which - like changes in eye and face orientation

359   – are important for gaze following (Hoffman and Haxby, 2000; Lee et al., 2010). Hence, the fact

360   that the GFP and the STS-FA are distinct, although both handling information on oriented faces

361   and most probably also oriented eyes, clearly indicates different functional roles. On the other

362   hand, the anatomical vicinity may suggest an exchange of pertinent information between the

363   two. However, if the GFP handles information on averted faces, why does it not light up in the

364   passive viewing experiment? The answer is that its activation is most probably contingent on

365   the presence of an object serving as goal for the gaze and observer´s the intention to follow

366   gaze.

367 We found the maximum BOLD response to faces in the STS-FA rather than in the FFA or OFA as

368 many other studies (Engell and Haxby, 2007). The reason is that - in our passive task to elicit

369 maximal responses in the STS-FA - the set of face stimuli used was confined to pictures of

370 emotionally neutral faces with averted eyes with the head straight, known to be less suitable for

371 the FFA or OFA (Hoffman and Haxby, 2000; Narumoto et al., 2001). On the other hand, in most

372 of the studies yielding stronger responses in the FFA or OFA, the emphasis was on faces

373 exhibiting direct eye gaze, stimuli that seem to favor identity-processing.

374 In (Pitcher et al., 2011), a face selective area in the right pSTS was reported which responded

375 three times more strongly to dynamic faces than to static faces. Hence, one may speculate that

376 the current study using static stimuli underestimated the true size of the STS-FA by and

377 therefore failed to reveal an overlap between the GFP and the STS-FA may. We can not exclude

378 the possibility that more powerful face stimuli might have expanded the activated areas with

379 the consequence of some overlap to emerge. However given the fact that the mean Talaraich

380 coordinates of the pSTS patch center as given by (Pitcher et al., 2011), (54, -38,4), and the

381 coordinates of the GFP in our study, (50,-64,2) are separated by 26mm Euclidean distance

382 clearly supports the conclusion of largely non-congruent patches at least when a static face

383 localizer is used to map face selective areas.

384 Non-human primates follow head gaze in order to establish joint attention. This behavior

385 emerges very early during the development of the individual (Tomasello and Carpenter, 2005;

386 Tomasello et al., 2007). According to Marciniak et al. (2015) it is characterized by key features

387 that make human eye gaze following reflex-like, namely swiftness and incomplete cognitive

388 control. As said earlier, monkey head-gaze following activates a patch of cortex (the monkey

389 GFP) whose location bilaterally in the posterior STS is reminiscent of the location of the human

390 GFP. Also the monkey GFP is anatomically distinct, not showing overlap with any of the face

391 patches that can be activated by the passive vision of faces (Tsao et al., 2003; Tsao et al., 2006).

392 As a matter of fact, the spatial relationship of the monkey GFP with respect to the posterior face

393 patch (PL) and the middle face patches (ML & MF) is reminiscent of the spatial relationship of

394 the human GFP to the most posterior face selective area (OFA) and the two more anterior ones

395 (FFA & STS-FA). This lends further support to the notion of a close correspondence of the

396 respective architectures. The major difference seems to be ability of the human architecture to

397 integrate social cues, providing directional information, other than head cues such as eye

398 direction or the direction of fingers (Materna et al., 2008b; Laube et al., 2011). In other words,

399 both species seem to exhibit a common core architecture for gaze following, possibly reflecting

400 homologous ancestry.

401 The notion of separate, yet possibly interdependent cortical structures for the processing of

402 faces and in particular faces showing gaze aversion and gaze following is interesting with regard

403 to observations on subjects with autism spectrum disorder (ASD). At least some ASD persons

404 seem to be able to distinguish between different eye gaze positions when tested in

405 discrimination tasks, suggesting an intact face processing network. However, they fail to use

406 information provided by the other´s face to follow her/his gaze and establish joint attention

407 (Baron-Cohen, 1995; Leekam et al., 1998; Leekam et al., 2000). In accordance with these

408 behavioural observations, Pelphrey et al., reported a lack of differentiation in the STS BOLD

409 responses of ASD subjects when confronted with averted target-directed and averted not-

410 target-directed eye gaze stimuli, a deficit that may reflect an inability to integrate information

411 on the other' s gaze and the object of interest (Pelphrey et al., 2005a). The tentative conclusion

412 suggested by these findings may be one of differential vulnerability of the face processing

413　network and the GFP with the latter selectively compromised in ASD. However, what exactly is

414　the added value of the GFP? At this stage, the lack of knowledge of the neuronal computations

415　inside the GFP, does not allow more than an admittedly rather vague speculation. We think that

416　the GFP may be needed to convert directional information on eye and face/ head orientation as

417　well as directional information offered by other parts of the body into a "vector" describing the

418　necessary shift of the observer´s 'spotlight of attention' to the place of interest. Moreover, in

419　order to ultimately lead to the establishment of joint attention devoted to an object found in a

420　particular place, the GFP may also help to integrate information on the object at stake. Finally,

421　in order to be viable these calculations require the integration of knowledge on the observer´s

422　view point. A final remark pertains a possible role of the most anterior member of the face

423　processing network, the IFG-FA, located at the junction of inferior frontal sulcus and the

424　precentral sulcus in gaze following. There is evidence that the BOLD response of IFG-FA to faces

425　is primarily driven by the eyes i.e., the response to faces with eyes is lower than the

426　presentation of the eyes alone and higher than to faces without eyes (Chan and Downing,

427　2011). In view of these findings and, moreover, the proximity of the IFG-FA to the frontal eye

428　field, the authors speculated that it might contribute to analyse others´ gaze in order to elicit

429　gaze following movements of the observer. Hence, future work will have to address the

430　possibility that not the face patch immediately neighbouring the GFP but a much more remote

431　anterior face patch, the IFG-FA, may serve as the major source of directional information

432　provided by the eyes and the face.

433

eNeuro Accepted Manuscript

References

443   Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. Trends in
444   cognitive sciences 4:267-278.
445   Baron-Cohen S (1994) How to build a baby that can read minds: Cognitive mechanisms in mindreading.
446   Cahiers de Psychologie Cognitive/ Current Psychology of Cognition 13:513-552.
447   Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press.
448   Chan AW, Downing PE (2011) Faces and eyes in human lateral prefrontal cortex. Frontiers in human
449   neuroscience 5:51.
450   Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze.
451   Neuroscience and biobehavioral reviews 24:581-604.
452   Engell AD, Haxby JV (2007) Facial expression and gaze-direction in human superior temporal sulcus.
453   Neuropsychologia 45:3234-3241.
454   Fox CJ, Iaria G, Barton JJ (2009) Defining the face processing network: optimization of the functional
455   localizer in fMRI. Hum Brain Mapp 30:1637-1651.
456   Fox E, Damjanovic L (2006) The eyes are sufficient to produce a threat superiority effect. Emotion
457   (Washington, DC) 6:534-539.
458   Friesen CK, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze.
459   Psychonomic Bulletin & Rev 5:490-495.
460   Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? NeuroImage 10:1-5.
461   Gobbini MI, Haxby JV (2006) Neural response to the visual familiarity of faces. Brain research bulletin
462   71:76-82.
463   Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not
464   generic within-category identification. Nature neuroscience 7:555-562.
465   Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception.
466   Trends in cognitive sciences 4:223-233.
467   Heller R, Golland Y, Malach R, Benjamini Y (2007) Conjunction group analysis: an alternative to
468   mixed/random effect analysis. NeuroImage 37:1178-1185.
469   Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human
470   neural system for face perception. Nature neuroscience 3:80-84.
471   Hooker CI, Paller KA, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2003) Brain networks for
472   analyzing eye gaze. Brain research Cognitive brain research 17:406-418.
473   Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network.
474   Brain research bulletin 67:87-93.

475   Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate
476   cortex specialized for face perception. The Journal of neuroscience : the official journal of the
477   Society for Neuroscience 17:4302-4311.
478   Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, Van Knippenberg A (2010) Presentation and
479   validation of the Radboud faces database. Cognition & Emotion 24:1377—1388.
480   Langton SR, Bruce V (2000) You must see the point: automatic processing of cues to the direction of
481   social attention. Journal of experimental psychology Human perception and performance
482   26:747-757.
483   Laube I, Kamphuis S, Dicke PW, Thier P (2011) Cortical processing of head- and eye-gaze cues guiding
484   joint social attention. NeuroImage 54:1643-1653.
485   Lee LC, Andrews TJ, Johnson SJ, Woods W, Gouws A, Green GG, Young AW (2010) Neural responses to
486   rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG.
487   Neuropsychologia 48:477-490.
488   Leekam SR, Hunnisett E, Moore C (1998) Targets and cues: gaze-following in children with autism.
489   Journal of child psychology and psychiatry, and allied disciplines 39:951-962.
490   Leekam SR, Lopez B, Moore C (2000) Attention and joint attention in preschool children with autism.
491   Developmental psychology 36:261-273.
492   Marciniak K, Dicke PW, Thier P (2015) Monkeys head-gaze following is fast, precise and not fully
493   suppressible. Proceedings Biological sciences / The Royal Society 282.
494   Marciniak K, Atabaki A, Dicke PW, Thier P (2014) Disparate substrates for head gaze following and face
495   perception in the monkey superior temporal sulcus. eLife 3.
496   Materna S, Dicke PW, Thier P (2008a) Dissociable roles of the superior temporal sulcus and the
497   intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. Journal of
498   cognitive neuroscience 20:108-119.
499   Materna S, Dicke PW, Thier P (2008b) The posterior superior temporal sulcus is involved in social
500   communication not specific for the eyes. Neuropsychologia 46:2759-2765.
501   Narumoto J, Okada T, Sadato N, Fukui K, Yonekura Y (2001) Attention to emotion modulates fMRI activity
502   in human right superior temporal sulcus. Brain research Cognitive brain research 12:225-231.
503   Pelphrey KA, Viola RJ, McCarthy G (2004) When strangers pass: processing of mutual and averted social
504   gaze in the superior temporal sulcus. Psychological science 15:598-603.
505   Pelphrey KA, Morris JP, McCarthy G (2005a) Neural basis of eye gaze processing deficits in autism. Brain :
506   a journal of neurology 128:1038-1048.
507   Pelphrey KA, Singerman JD, Allison T, McCarthy G (2003) Brain activation evoked by perception of gaze
508   shifts: the influence of context. Neuropsychologia 41:156-170.
509   Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G (2005b) Functional anatomy of biological
510   motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand
511   movements. Cerebral cortex (New York, NY : 1991) 15:1866-1876.
512   Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic
513   versus static information in face-selective cortical regions. NeuroImage 56:2356-2363.
514   Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing
515   eye and mouth movements. The Journal of neuroscience : the official journal of the Society for
516   Neuroscience 18:2188-2199.
517   Shimojo S, Simion C, Shimojo E, Scheier C (2003) Gaze bias both reflects and influences preference.
518   Nature neuroscience 6:1317-1322.
519   Tomasello M, Carpenter M (2005) The emergence of social cognition in three young chimpanzees.
520   Monographs of the Society for Research in Child Development 70:vii-132.
521   Tomasello M, Hare B, Lehmann H, Call J (2007) Reliance on head versus eyes in the gaze following of
522   great apes and human infants: the cooperative eye hypothesis. Journal of human evolution
523   52:314-320.

524   Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans.
525   Proceedings of the National Academy of Sciences of the United States of America 105:19514-
526   19519.
527   Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-
528   selective cells. Science (New York, NY) 311:670-674.
529   Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque
530   cerebral cortex. Nature neuroscience 6:989-995.
531   Van Essen DC (2005) A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral
532   cortex. NeuroImage 28:635-662.
533   Wicker B, Michel F, Henaff MA, Decety J (1998) Brain regions involved in the perception of gaze: a PET
534   study. NeuroImage 8:221-227.
535   Wollaston WH (1824) On the apparent direction of eye in a portrait. Philosophical Transactions of the
536   Royal Society of London Series B 114:247-256.

537

Figure Legends

539   **Figure 1.** Sequence of visual stimuli in the active task. At the beginning of each block of trials a

540   written instruction (either gaze following or colour matching) was presented on the screen for

541   five seconds. Each trial started with a baseline fixation picture with direct gaze (lasting for 5

542   seconds), immediately followed by one out of five possible portraits ("target portraits"), present

543   for 4 sec, with the demonstrator´s gaze directed at a specific target and exhibiting a distinct iris

544   colour. Subjects were not allowed to make an eye movement until the disappearance of the

545   fixation target. Afterwards alternately 10 fixations (each 5s duration) and 10 trial pictures (each

546   4 s duration) were presented. The demonstrator has agreed her portrait to be published.

547

548   **Figure 2.** Illustration of the first experiment's stimulus. The eyes of the person are directed to

549   the dark blue target (gaze cue), but the person´s iris colour corresponds to the light brown

550   target (colour cue). According to the introduced condition at the beginning of the block, the

551   subject would have to make a saccade towards the dark blue target (gaze following condition)

552   or towards the light brown target (colour matching condition). The demonstrator has agreed

553   her portrait to be published.

554

555

556   **Figure 3.** Behavioural data for gaze following (dark grey) and colour matching (light grey)

557   showing neither a significant difference in the mean accuracy nor mean reaction time (= time

558   between the "go"-signal and the start of the eye-movement (N= 20 sessions, 160 correct trials).

559   Error bars represent standard errors.

560

561   **Figure 4.** MRI group data showing the BOLD response for the contrast gaze following versus

562   baseline fixation.

563

564   **Figure 5.** MRI group data showing the BOLD response for the contrast colour matching versus

565   baseline fixation.

566

567   **Figure 6.** MRI group data showing the BOLD response for the contrast gaze following versus

568   colour matching. Activation maximum in right hemisphere in Talaraich coordinates (50,-64,2)

569

570   **Figure 7.** Spatial organization of face selective areas and the gaze following patch
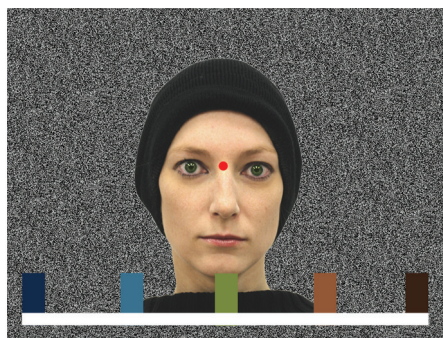
571

572   **Figure 8.** Selectivity of the individually defined STS-FA to gaze following in contrast to the

573   selectivity of the GFP to static face perception. Error bars indicate 90% confidence intervals. In

574   the right STS-FA (Talaraich coordinates of the peak (51,-42,12)), the mean contrast values for

575   gaze following is not significantly different from zero, in accordance with the assumption of a

576   lack of gaze following selectivity (t-test, p=0.49). On the other hand, the contrast value for face

577   perception in the right GFP (Talaraich coordinates of the peak (50, -64,2)) is not significantly

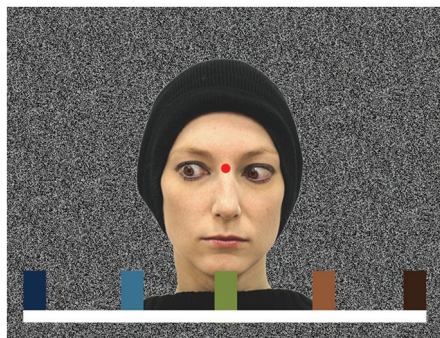578   different from zero meaning no face-selectivity (t-test, p=0.20).

579

Baseline Fixation   Spatial Target   Go Signal

5 seconds   1 seconds   3 seconds

t value

0   12

t value

0   12