Selectivity for the Human Body in the Fusiform Gyrus

Marius V. Peelen and Paul E. Downing
School of Psychology, University of Wales, Bangor, United Kingdom

Submitted 17 May 2004; accepted in final form 3 August 2004

Peelen, Marius V. and Paul E. Downing. Selectivity for the human body in the fusiform gyrus. J Neurophysiol 93: 603–608, 2005. First published August 4, 2004; doi:10.1152/jn.00513.2004. Functional neuroimaging studies have revealed human brain regions, notably in the fusiform gyrus, that respond selectively to images of faces as opposed to other kinds of objects. Here we use fMRI to show that the mid-fusiform gyrus responds with nearly the same level of selectivity to images of human bodies without faces, relative to tools and scenes. In a group-average analysis (n = 22), the fusiform activations identified by contrasting faces versus tools and bodies versus tools are very similar. Analyses of within-subjects regions of interest, however, show that the peaks of the two activations occupy close but distinct locations. In a second experiment, we find that the body-selective fusiform region, but not the face-selective region, responds more to stick figure depictions of bodies than to scrambled controls. This result further distinguishes the two foci and confirms that the body-selective response generalizes to abstract image formats. These results challenge accounts of the mid-fusiform gyrus that focus solely on faces and suggest that this region contains multiple distinct category-selective neural representations.

INTRODUCTION

A major current theme in cognitive neuroscience is the effort to understand the brain systems involved in perceiving the identities, emotional states, and intentions of other people. To date, much of this research has focused on the perception of faces, with particular concentration on face-selective activity in the fusiform gyrus. Functional magnetic resonance imaging (fMRI) studies have shown that the “fusiform face area” (FFA) responds strongly and selectively to human faces as compared with a wide variety of controls (Halgren et al. 1999; Kanwisher et al. 1997; Puce et al. 1995). FFA activity closely tracks awareness of the presence of a face (Andrews and Schluppeck 2004; Andrews et al. 2002; Hasson et al. 2001; Tong et al. 1998) as well as trial-by-trial psychophysical performance on face detection and identification tasks (Grill-Spector et al. 2004). One interpretation of these findings is that the FFA represents a cortical “module” for face processing.

Strong selectivity for a stimulus class, however, does not necessarily imply a dedicated system for processing that class. One alternative proposal is that the FFA may instead be better conceived as a mechanism for distinguishing visually similar exemplars of any object class for which the viewer has substantial expertise (Tarr and Gauthier 2000). This is supported, for example, by the finding that bird and car experts show increased FFA activity when viewing birds and cars, respectively (Gauthier et al. 2000). Others have gone further to suggest that the FFA, and highly-selective activation “peaks” in general, have no special functional role, but instead form part of a broad neural network that represents objects by a distributed collection of feature analyzers (Haxby et al. 2001; but see Spiridon and Kanwisher 2002). Thus the FFA has been at the center of a broader debate about the organization of ventral stream object representations (Cohen and Tong 2001; Kanwisher 2000; Levy et al. 2001).

The face, however, is not the only source of socially relevant cues; the rest of the body also conveys such information. Although human bodies and faces are visually dissimilar, they have other features in common. Both provide cues to identity, emotion, intention, age, and gender. For both bodies and faces, the differences between exemplars are metric rather than qualitative, and for both, most adults will have developed substantial perceptual expertise. Moreover, recent studies have shown behavioral effects for bodies that were generally thought to be specific for faces, such as a “body-inversion” effect (Reed et al. 2003) and an advantage for bodies in attentional capture (Downing et al. 2004).

There is some evidence that this conceptual and behavioral similarity between faces and bodies may be reflected in neural activity in the fusiform gyrus. One fMRI study found a significantly higher response in the FFA to bodies than to other object categories, although the response to bodies was significantly lower than to faces (Kanwisher et al. 1999). In this study, however, analysis of body-related activity was limited to a region functionally defined by contrasting faces versus houses. Another fMRI study, in which subjects were scanned while viewing segments of a James Bond movie, showed that bodies activate the fusiform gyrus near the FFA (Bartels and Zeki 2004). The stimuli in this study were not controlled, so it is not clear whether bodies, faces, or both were responsible for this activation. A recent event-related potential (ERP) study showed that the face-selective N170 is also elicited by bodies (with faces blurred), suggestive of shared underlying neural processes (Stekelenburg and de Gelder 2004). Note, however, that the relationship of the N170 to the FFA is unclear. Furthermore, the response to bodies in this study could also reflect contextual enhancement of the blurred faces (Cox et al. 2004). Finally, there is mixed evidence on whether the FFA response to the bodies of non-human animals is higher than the response to inanimate objects (Chao et al. 1999; Kanwisher et al. 1999; Grill-Spector et al. 2004).

Investigations of the macaque visual system also provide relevant findings. Single-unit recordings have revealed neurons distributed throughout the macaque temporal lobe, particularly in the superior temporal sulcus (STS), that respond to views of the face and body and to actions involving them (Jellema and...
Perrett 2003; Perrett et al. 1982; Wachsmuth et al. 1994). More recently, fMRI in the macaque has shown face- and body-selective activations in adjacent regions of the lower bank of the STS (Tsao et al. 2003). A homology between this region and human FFA has been proposed (Gauthier and Logothetis 2000; Tsao et al. 2003) although conclusive evidence is still lacking. Thus in the macaque the representations of faces and bodies appear to be intertwined in a region that may map to the fusiform gyrus in humans.

The preceding evidence leads to the prediction that human bodies, even without faces, may be represented in the human mid-fusiform gyrus. This possibility has not been tested rigorously to date. Therefore the three objectives of the present study were to use fMRI to measure the selectivity of neural responses to headless bodies in the fusiform gyrus, to compare the locations of body- and face-related activity in this region, and to test whether body selectivity in the fusiform gyrus generalizes to more abstract depictions of the human form.

**METHODS**

**Participants**

Twenty-two healthy adult volunteers (12 female) were recruited from the University of Wales, Bangor community for the first experiment. Eight subjects (4 female; 1 of whom participated in the 1st experiment) participated in the second experiment. Participants satisfied all requirements in volunteer screening and gave informed consent. Ethics approval was obtained from the School of Psychology at the University of Wales, Bangor, United Kingdom, and the North-West Wales Health Trust. Participation was compensated at £20 per session.

**Experimental paradigm**

The first experiment consisted of images of faces, human bodies without heads, outdoor scenes, and handheld tools presented in a blocked design (see Fig. 1A). There were 40 exemplars of each category, divided into two sets (A and B).

Subjects performed four runs, each lasting 315 s (105 functional volumes). Runs 1 and 2 contained stimuli from set A and runs 3 and 4 contained stimuli from set B. Each functional run consisted of 21 15-s blocks. Blocks 1, 6, 11, 16, and 21 were a fixation-only baseline condition. Each of the remaining blocks comprised presentation of 20 exemplars from a single category. Within a block, each stimulus was presented for 300 ms, with an ISI of 450 ms, during which a central fixation point appeared on the screen. Two versions of block order were adopted. The first half and second half of one version were swapped to create the second version. The order of blocks was symmetrically counterbalanced within each version so that the first half of each version was the mirror order of the second half. The result is that the mean serial position of each condition was equated, reducing the possible contribution of linear confounds to the results.

The second experiment tested 15 stick figures and 15 scrambled control items (see Fig. 1B). Fifteen-second blocks from each condition were alternated with intervening fixation baseline blocks as in the preceding text. Two orders of each design were tested between subjects to counterbalance for order effects. In each block, 15 images were presented (300 ms on, 700 ms off for each item).

In both experiments, subjects performed a “one-back” repetition-detection task, in which they were asked to press a button whenever...
an image occurred twice in immediate succession. Two image repetitions occurred at randomly selected points in each block.

**Functional imaging and analysis**

**DATA ACQUISITION.** A 1.5 T Philips MRI scanner with a SENSE (Pruessmann et al. 1999) parallel head coil was used. For functional imaging, an EPI sequence was used (TR = 3,000 ms, TE 50 ms, flip angle 90°, FOV = 240, 30 axial slices, 64 × 64 in-plane matrix, 5 mm slice thickness). The scanned area covered the whole cortex and most of the cerebellum.

**DATA ANALYSIS.** Preprocessing and statistical analysis of MRI data were performed using BrainVoyager 4.9 (Brain Innovation, Maastricht, The Netherlands). The first three volumes of each run were discarded to avoid differences in T1 saturation. Functional data were motion-corrected, and low-frequency drifts were removed with a temporal high-pass filter (0.006 Hz). Functional data were manually co-registered with three-dimensional (3D) anatomical T1 scans (1 × 1 × 1.3-mm resolution), and then resampled to isotropic 3 × 3 × 3-mm voxels with trilinear interpolation. The 3D scans were transformed into Talairach space (Talairach and Tournoux 1988), and the parameters for this transformation were subsequently applied to the co-registered functional data.

To generate predictors for the multiple-regression analyses, the event time series for each condition were convolved with a delayed gamma function (delta = 2.5 s; tau = 1.25 s) to model the hemodynamic response (Boynton et al. 1996). Voxel time series were z-normalized for each run, and additional predictors accounting for baseline differences between runs were included in the design matrix.

**ROI DEFINITION AND PEAK-VOXEL COMPARISON.** For each subject, we located the most significantly activated voxel for the contrast faces versus tools and the contrast bodies versus tools, within a restricted part of ventral cortex (Talairach coordinates: 30 < x < 50, -60 < y < -30, -30 < z < -5). Regions of interest (ROIs) for the first experiment were defined as the set of contiguous voxels that were significantly (P < 0.05, uncorrected for multiple comparisons) activated within 3 mm in the anterior/posterior, superior/inferior, and medial/lateral direction of the most significantly activated voxel. For the first experiment, runs 1 and 3 were combined to define the ROIs for runs 2 and 4, and vice versa. Thus the data used for ROI definition were independent from the data reported. For the second experiment, face- and body-selective ROIs were identified based on an independent set of localizer data.

To further examine the spatial relationship between ROIs, the locations of the most significantly activated voxels for faces and bodies (vs. tools) were compared within subjects. We calculated three measures of the linear distance between the most significant voxels. Measure 1 was the distance between peaks within a category and stimulus set. The locations of the most significant voxels of runs 1 + 3 and runs 2 + 4 were compared for faces and bodies separately. Thus in this comparison, each set of data contained one run for each of the two stimulus sets. This measure serves as a baseline estimate of how the peak location for identical stimuli varies between scans. Measure 2 was the distance between peaks within a category but across stimulus sets. The locations of the most significant voxels of runs 1 + 2 and runs 3 + 4 were compared, again for faces and bodies separately. This test shows how spatially consistent responses are to different exemplars of the same category. Measure 3 was the distance between peaks across categories. The locations of the most significant voxels of runs 1 + 3 (e.g., for faces) and runs 2 + 4 (e.g., for bodies) were compared. Likewise, peak voxels from runs 1 + 2 and runs 3 + 4 were compared, again across categories. This measure tests whether the face- and body-selective foci occupied different locations.

Note that in these measures, the mean distance between peak voxels can be smaller than the distance between two adjacent voxels in a single scan. These distance measures are conservative in the sense that they could fail to detect differences in the location of peaks that are smaller than the scanned resolution but could not artifically create such a difference where one did not exist at a finer scale.

**RESULTS**

A whole-brain group-averaged multiple regression analysis (fixed effects) contrasting faces versus tools and bodies versus tools revealed significant activation in the right fusiform gyrus for both contrasts. The fusiform activations were comparable in location and extent (Fig. 2A). The face contrast, but not the body contrast, activated a right ventral occipital region, corresponding to the occipital face area (OFA) (Puce et al. 1996). The body contrast, but not the face contrast, activated an occipito-temporal region bilaterally, corresponding to the “extrastriate body area” (EBA) (Downing et al. 2001).

For each subject, we identified the right hemisphere fusiform face-selective region (FSR; defined by faces vs. tools) and the right hemisphere fusiform body-selective region (BSR; defined by bodies vs. tools) as regions of interest (Fig. 2B and Table 1). The labels we give these regions are intended for explanatory convenience; we do not necessarily assume that there are 2 functionally independent regions). For each ROI, we calculated the mean parameter estimate from the regression analysis for bodies, faces, tools, and scenes (Fig. 3A).

A repeated-measures ANOVA revealed a significant interaction between category and ROI [F(3,19) = 7.4, P < 0.005]. The FSR responded more to faces than bodies (t21 = 4.2, P < 0.001), tools (t21 = 9.7, P < 0.001), and scenes (t21 = 8.1, P < 0.001) and more to bodies than tools (t21 = 7.3, P < 0.001) and scenes (t21 = 6.3, P < 0.001) consistent with previous findings in the FFA (Kanwisher et al. 1999). In contrast, the BSR did not distinguish faces and bodies (t21 = 0.5, P = 0.60) and responded more to faces and bodies than tools (t21 = 6.9, P < 0.001 and t21 = 7.4, P < 0.001) and scenes (t21 = 5.9, P < 0.001 and t21 = 7.8, P < 0.001). No differences were observed in the response to tools and scenes in either region (both P’s > 0.35). The interaction between category and ROI indicates that the two ROIs occupied at least partially different locations.

To compare the location of fusiform BSR and FSR activity within subjects, we calculated the distance between the most significant voxels: within-category, within-stimulus set; within-category, between-stimulus set; and between-categories (see METHODS). No effect of stimulus set (measure 1 vs. measure 2) was observed [F(1,21) = 0.02, P = 0.9] nor did this depend on category [F(1,21) = 0.4, P = 0.54], showing that different exemplars of the same category produce similar activation peaks. In contrast, we found a significantly larger distance across categories [measure 3; 7.5 ± 0.7 (SE) mm] than within categories (measure 2; 5.7 ± 0.6), t21 = 3.0, P < 0.01, indicating distinct peaks of activation for bodies and faces. We also tested whether there was a consistent spatial relationship between the FSR and BSR within subjects. The difference in locations between ROIs did not significantly differ from zero in the x, y, and z directions, all P > 0.20, indicating that there was no consistent relationship between the two ROIs, at least in 3D space.

In a second experiment, we compared headless stick figure depictions of bodies and scrambled control figures (Fig. 1B). Abstract stimuli such as these minimize the differences in
image features between body and nonbody stimuli. A significant interaction, $F(1,7) = 7.6$, $P < 0.05$, showed that the difference between stick figures and controls differed between the BSR and the FSR. The response to stick figures was significantly higher than to controls in the BSR ($t_7 = 2.6$, $P < 0.05$), but not in the FSR ($t_7 = 0.5$, $P = 0.62$). The OFA likewise showed no significant difference, ($t_7 = 1.1$, $P = 0.33$). There was a significant effect in both the right EBA ($t_7 = 3.4$, $P < 0.05$) and the left EBA ($t_7 = 3.5$, $P < 0.05$), replicating previous findings (Downing et al. 2001). Response magnitudes in all of these regions are given in Fig. 3B. The selective response to stick figures as compared with scrambled controls was strong enough to show up in a whole-brain, group-averaged, fixed-effects multiple regression analysis. Significant activation ($P < 0.0005$ uncorrected) was found in the right fusiform gyrus (peak Talairach coordinates: $41, -38, -21$) and bilaterally in the vicinity of the EBA (left: $-49, -74, 9$; right: $47, -62, 6$).

**DISCUSSION**

We found nearly identical activation in the fusiform gyrus for faces and bodies (vs. tools) in a whole-brain group analysis. Analyses of individually defined ROIs, however, showed a different selectivity pattern for the region defined by faces compared with that defined by bodies. There were distinct peaks of activation within the fusiform gyrus for the two different categories but not for different stimulus sets within a category. Finally, we confirmed that the BSR (but not the FSR) distinguished intact and scrambled versions of human stick-figures. This result shows that the selective response to bodies found here is not entirely due to differences in low-level image features between bodies and other objects.

One account of our findings could be that perceiving bodies leads observers to mentally image the missing faces, which in turn activates face-selective neurons. This is unlikely to explain our results for several reasons. The body stimuli were unfamiliar and not associated to a particular face; it may not be possible to mentally image a “generic” face in this situation. Second, a previous study has compared the FFA response to front, side, and back views of heads (Tong et al. 2000). The back of the head, particularly when viewed in context with frontal views of the same individuals, would be expected to induce mental imagery for the missing face at least as strongly as bodies, yet FFA
responses to this condition were comparable to inanimate objects. Finally, even explicit instruction to image specific faces produces relatively weak FFA activation that is not sufficient to explain the strong body selectivity seen here (O’Craven and Kanwisher 2000).

A related proposal is that bodies provide a context that causes ambiguous perceptual input to be perceived as a face. Cox et al. (2004) showed that the mid-fusiform activity produced by a Gaussian blur was increased when it was shown in a face-like contextual relationship to a body (i.e., atop the neck). (The selectivity of this region for faces:scenes, however, was low, raising the possibility that it did not correspond perfectly to the FFAs as defined in previous reports.) Here, however, we showed only headless bodies without any stimulus that could be contextually enhanced.

Another possible account of the present findings relates to retinal eccentricity (Levy et al. 2001). Lateral visual areas, including the fusiform gyrus, are thought to be important for center-biased visual field representations, whereas more medial areas are thought to be important for periphery-biased field representations. Bodies and faces are likely to be represented more centrally than scenes; this could account for the higher response to bodies and faces relative to scenes. Tools, however, often contain fine details that need central processing to discriminate, and the response in this region to tools was relatively low. Moreover, we found no significant difference in either fusiform ROI between the response to tools and scenes, arguing against a central visual field bias in this part of the fusiform gyrus.

How do the present findings improve our understanding of the ventral temporal lobe? The mid-fusiform gyrus is part of a nonretinotopic, high-level object representation system. Our results show that some neurons in this region distinguish bodies, as well as faces, from other object kinds. Other object categories (such as cars, birds, and “greebles”) have also been shown to produce an enhanced, although much less selective, response in this region, particularly in subjects who are experts at distinguishing among exemplars of these categories (Gauthier et al. 2000). One possibility is that the same population of neurons is engaged by all of these categories. Activity in these neurons could thus reflect cognitive processes that are independent of the stimulus categories involved, such as distinguishing among visually similar exemplars. This seems improbable, at least for the case of faces and bodies, however, in light of our finding that selective foci for these categories occupy distinct locations. We suggest instead that different, possibly overlapping, populations of selective neurons co-exist at close quarters. Faces and bodies are represented by larger populations and/or more selective neurons than other categories. Existing data do not speak to the root cause of this bias, which could be present from birth or could develop through experience.

These findings must also be considered in light of the recent proposal that objects are represented in the ventral stream by distributed patterns of neural activity. On this view, highly selective peaks, such as those seen here for faces and bodies, play no unique role in this representation (Haxby et al. 2001).

**TABLE 1.** Talairach coordinates and T values of the most significant voxel for faces and bodies (vs. tools) in the mid-fusiform gyrus

<table>
<thead>
<tr>
<th>Subject</th>
<th>Faces-Tools</th>
<th>Bodies-Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td>-50</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>-47</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>-38</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>-53</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>-53</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>-47</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>-50</td>
</tr>
<tr>
<td>8</td>
<td>44</td>
<td>-47</td>
</tr>
<tr>
<td>9</td>
<td>44</td>
<td>-38</td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>-44</td>
</tr>
<tr>
<td>11</td>
<td>38</td>
<td>-53</td>
</tr>
<tr>
<td>12</td>
<td>35</td>
<td>-50</td>
</tr>
<tr>
<td>13</td>
<td>41</td>
<td>-53</td>
</tr>
<tr>
<td>14</td>
<td>35</td>
<td>-32</td>
</tr>
<tr>
<td>15</td>
<td>38</td>
<td>-44</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>-38</td>
</tr>
<tr>
<td>17</td>
<td>41</td>
<td>-32</td>
</tr>
<tr>
<td>18</td>
<td>41</td>
<td>-56</td>
</tr>
<tr>
<td>19</td>
<td>44</td>
<td>-38</td>
</tr>
<tr>
<td>20</td>
<td>38</td>
<td>-35</td>
</tr>
<tr>
<td>21</td>
<td>35</td>
<td>-47</td>
</tr>
<tr>
<td>22</td>
<td>35</td>
<td>-53</td>
</tr>
</tbody>
</table>

**FIG. 3.** A: Parameter estimates for faces, bodies, tools, and scenes in the face-selective (FSR) and body-selective (BSR) fusiform regions of interest (as defined by faces vs. tools and bodies vs. tools, respectively). Regions of interest (ROIs) were defined with independent data sets. Selectivity in the face-selective region (ratio to tools) was 2.0 for faces and 1.6 for bodies. Selectivity in the body-selective region was 1.7 for both faces and bodies. Error bars indicate SE. B: parameter estimates for body stick figures and scrambled controls in key ROIs. Error bars indicate SE.
One way to test this prediction is to ask for which brain regions activity is systematically related to performance on perceptual tasks. A recent study using this approach (Grill-Specter et al. 2004) showed a positive correlation between successful face detection and identification and FFA activity. This relationship was not observed in other ventral areas, even those (such as the “parahippocampal place area” (Epstein and Kanwisher 1998)) where the response to faces is consistently low, and could in principle provide information relevant to face processing. This finding suggests that regions with high category selectivity may indeed have a privileged role in vision.

These considerations lead to possible avenues for further research. Scanning at higher resolution may help to elucidate the spatial relationship among subregions of the fusiform gyrus that show selectivity for faces, bodies, or other object kinds. Further stimulus or task manipulations will be necessary to determine whether these selective foci can be functionally dissociated. Finally, it will be important to determine for which categories activity in the fusiform region predicts trial-by-trial performance on perceptual tasks. If this relationship is found for the body-selective region identified here, it would provide additional evidence that the fusiform gyrus plays a previously unsuspected role in visual processing of the human body.

Acknowledgments

We thank T. Bedson for technical assistance and A. Chan, R. Epstein, M. Hoerger, N. Kanwisher, D. Linden, and I. Morrison for helpful comments.

References


