What's behind a Face: Person Context Coding in Fusiform Face Area as Revealed by Multivoxel Pattern Analysis

J. van den Hurk1,2, F. Gentile1,2 and B.M. Jansma1,2

1Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, the Netherlands and 2Maastricht Brain Imaging Center (M-BIC), 6200 MD Maastricht, the Netherlands

Address correspondence to Job van den Hurk, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands. Email: job.vanden hurk@maastrichtuniversity.nl

The identification of a face comprises processing of both visual features and conceptual knowledge. Studies showing that the fusiform face area (FFA) is sensitive to face identity generally neglect this dissociation. The present study is the first that isolates conceptual face processing by using words presented in a person context instead of faces. The design consisted of 2 different conditions. In one condition, participants were presented with blocks of words related to each other at the categorical level (e.g., brands of cars, European cities). The second condition consisted of blocks of words linked to the personality features of a specific face. Both conditions were created from the same 8 x 8 word matrix, thereby controlling for visual input across conditions. Univariate statistical contrasts did not yield any significant differences between the 2 conditions in FFA. However, a machine learning classification algorithm was able to successfully learn the functional relationship between the 2 contexts and their underlying response patterns in FFA, suggesting that these activation patterns can code for different semantic contexts. These results suggest that the level of processing in FFA goes beyond facial features. This has strong implications for the debate about the role of FFA in face identification.

Keywords: conceptual face processing, face identification, fMRI, pattern recognition, semantic selection

Introduction

When seeing the face of a friend, we first recognize it as being familiar. Instantly, we recall biographic information about him or her and retrieve his or her name. From a cognitive point of view, the identification of a face follows the matching of a set of facial features against hard-wired templates of known faces (Bruce and Young 1986). Once we recognize the perceived face as being someone familiar, we have instant access to personal-semantic specific knowledge about that individual’s occupation and socioeconomic status (Bruce and Young 1986), followed by access to name (Craigie and Hanley 1993, 1997). The mechanism of face perception and the consequent identification is mediated by a complex network of brain regions (Haxby et al. 2000). While the processing of facial stimuli is understood relatively well up to the level of face identification based on visual features, the stage of person-specific knowledge access remains unclear.

One specific brain region that has been shown to play a central role in face processing is the fusiform face area (FFA; Puce et al. 1995; Kanwisher et al. 1997), a small region located on the ventral bank of the temporal lobe that shows a particular sensitivity toward faces over other stimuli, for example, objects. Next to face recognition, a wealth of evidence showed that FFA is responsible for face identification (Gauthier et al. 2000; Hoffman and Haxby 2000; Barton et al. 2002; Haxby et al. 2002; Andrews and Ewbank 2004; Grill-Spector et al. 2004; Loffler et al. 2005). However, several neuroimaging (Sergent et al. 1992; Kriegeskorte et al. 2007; Brambati et al. 2010) and lesion studies (Evans et al. 1995; Tranel et al. 1997; Gainotti et al. 2003; Tranel 2006) suggested that actual facial identity is maintained in anterior inferior temporal (aIT) cortex and is then back-projected to FFA (Kriegeskorte et al. 2007).

Although the identity of a familiar face comprises visual features and conceptual and biographic knowledge about that face, previous studies either did not dissociate those 2 components experimentally or found no effects in FFA (Mur et al. 2010). Therefore, it remains still unknown whether the face identity representation in FFA is confined solely to visual features or whether it is extended with coding for conceptual knowledge about a person. In order to address this issue, it would be valuable to investigate the role of FFA in the processing of semantic information about a face, in the absence of direct facial information (visual features).

In the present study, we hypothesize that FFA has access to a semantic context when no face is visible during the presentation of that context. Using high-field functional magnetic resonance imaging (fMRI), we measured the functional response of FFA in 2 different conditions: a person-specific (PER) and a categorical (CAT) condition. In the former, we presented blocks of words related to a specific face that was presented at the beginning of the block. The CAT condition consisted of a series of words all belonging to the same category. As the words used for both conditions (PER, CAT) were selected from the same set of stimuli, any item-related confound is to be excluded. We subsequently used a machine learning classifier (linear support vector machine [SVM]) to analyze our data. This method allowed us to look at the neural response of FFA in a multivariate fashion. More specifically, instead of analyzing each and every voxel separately, we considered the pattern of activity of the total amount of voxels in FFA. We investigated whether the 2 different contexts (PER, CAT) elicited different patterns.

The use of pattern recognition as method to analyze fMRI data has proven to be an effective way of decoding brain states from distributed patterns of activity rather than focusing on the average response of an entire region (Cox and Savoy 2003; Kamitani and Tong 2005; Haynes and Rees 2006; Norman et al. 2006; Formisano, De Martino, Bonte, et al. 2008; Mur et al. 2009; Pereira et al. 2009).

Materials and Methods

Participants

Seven right-handed healthy adults with normal or corrected to normal vision participated in this study (5 females, mean age 24.3 years old). Participants were screened for fMRI compatibility, signed informed
consent and were financially compensated for their time. After acquisition, one participant had to be excluded from the analysis as it was not possible to localize FFA. The study was approved by the Ethical Committee of the Faculty.

Stimuli
We created 2 conditions from an 8 × 8 word matrix (Table 1), consisting of 8 categories (professions, capital cities, house types, sports, hobbies, domestic animals, music styles, and car brands). The words used were all Dutch, the native language of all participants. For the category (CAT) condition, 8 blocks each consisting of 8 words were created by using the items from a single category per block (the columns in the word matrix). Each CAT block thus represented 1 of 8 categories (i.e., “professions”). For the person (PER) condition, 8 blocks each consisting of 8 words were created by using one word from each category (the rows in the word matrix): each resulting sequence of 8 words could be interpreted as a keyword description of an individual profile (i.e., a surgeon living in Amsterdam, playing soccer, etc.). By using this approach, we were able to create 2 distinct semantic contexts while controlling for visual and semantic input of individual items across conditions.

For the person and category condition, we used 8 different face images (provided by Mark Steyvers http://psie.exp.ss.uc.edu/research/software.htm) and 8 images of objects (softball glove, house, car, book, dog, construction worker, globe, and saxophone) as instruction for the person and category conditions, respectively.

Procedure
The experiment consisted of 3 runs in a block design. Each run consisted of 8 blocks from each condition (CAT, PER) and started with a fixation cross lasting for 26 s. At the beginning of each block, an instructional image (face or object) was presented for 6 s. The image was followed by a fixation cross that lasted for 10 s. This interval allowed the blood oxygen level-dependent (BOLD) response elicited by the stimulus in the instruction phase to return back to baseline before the onset of the actual block. Next, a block of 8 words was presented, where each word lasted for 1500 ms, followed by 500 ms of fixation (see Fig. 1). Participants were instructed to pay close attention to the block of words and to associate each of them with the image in the instruction (i.e., to associate the category sports with the image of a softball glove, within the category condition). The intention of this approach was to encourage the participants to actively create a semantic context out of the presented words they were facing. They were informed that after each run, all images were presented one by one. The task of the participants was to recall as many words as possible related to that category or person. This additional debriefing phase was included to encourage participant to attend to the block of words inside the scanner.

Data Acquisition
Functional and anatomical images were acquired on a 3-T Siemens Magnetom Allegra head scanner (Siemens Medical System, Erlangen, Germany) using a standard head coil. Thirty-two oblique axial slices (3.5 × 3.5 mm, interslice distance 0 mm) covering the entire cortical volume were collected using a standard echo-planar imaging sequence (repetition time [TR] = 2000 ms, matrix size = 64 × 64, echo time [TE] = 30 ms). For each run, we collected 355 functional volumes, of which the first 2 were excluded due to T1 saturation. In addition to the experimental runs, an independent FFA localizer run was included. Anatomical images covering the whole brain were obtained after the acquisition of this run was temporally aligned using a cortex-based alignment procedure, an algorithm that uses individual curvature information to align corresponding gyri and sulci across participants (Goebel et al. 2006). The analysis was performed using a random effects General Linear Model (GLM), using ‘subject’ as random variable. A design matrix was created, using a predictor for both conditions. The predicted time courses were adjusted for the typical hemodynamic response delay by convolution with a canonical (double gamma) hemodynamic response function. No Statistical contrasts were considered at the multisubject whole-brain level, FDR corrected at q = 0.05. We subsequently performed a region-of-interest (ROI) analysis of the person > category contrast on the individual subject level. We focused on individual left and right FFA as ROI, as identified by the independent localizer run.

The data were also analyzed via a pattern classification approach. We used a SVM, a model that performs binary classification on a data set by placing all cases in a multidimensional space. Each individual case (or example) is expressed as a vector of N features in the N-dimensional space. The examples are labeled as belonging to 1 of 2 experimental conditions. An SVM training algorithm then defines an N-1 dimensional hyperplane that optimally separates the data in 2 categories. The resulting model can be used to predict whether a new example falls into one category or the other, depending on which side of the hyperplane the example falls. Through this means, the accuracy of the trained model can be assessed by considering the ratio of correctly classified examples.

In the present study, the features were represented by the voxels identified by the FFA localizer. In addition, in order to analyze the right and left FFA independently, we considered 2 different feature spaces comprising the voxels from the right and left FFA, respectively. Each example in such a multidimensional space was represented by a multidimensional vector comprising the hemodynamic responses from all voxels in FFA (right or left) related to each of the 48 blocks measured along the 3 experimental runs. The a priori assumption on the ROI decreases the amount of irrelevant features, thereby increasing the generalization performance of the classifier (Cox and Savoy 2003; Kamitani and Tong 2005; Formisano, De Martino, and Valente 2008). The examples from this example vector consisted of the BOLD response averaged across the time window surrounding the stimulus onset, computed for each voxel of the feature space. Subsequently, this value was baseline corrected by considering the mean amplitude of the BOLD response of the volumes immediately preceding the onset of the block (volumes [-1,0]) and subtracting this baseline from the original value.

Analysis
Functional and anatomical data were preprocessed and analyzed using the BrainVoyager QX2.1 package (Brain Innovation, Maastricht, the Netherlands). Functional volumes were first corrected for slice scan-time differences and 3D head motion. In order to enhance the subsequent alignment of the functional images to the anatomical volume, the first and second runs were corrected with the third run as intra- and interrun reference, as the acquisition of this run was temporally adjacent to the anatomical scan. Subsequently, linear trends and low-frequency temporal drifts were removed from the data using a high-pass filter, removing temporal frequencies below 2 cycles per run. No spatial smoothing was performed. After the preprocessing, functional data were coregistered to the high-resolution anatomical volume and normalized to Talairach space.

For each participant, the location of the FFA was determined bilaterally by an independent localizer run, using gray scale images of faces, scrambled faces, and houses. The procedure defines the FFA as the result of a conjunction analysis (FDR [False Discovery Rate] corrected at q = 0.05) between a face responsive contrast (%BOLD signal change faces > scrambled faces) and a face selective contrast (%BOLD signal change faces > houses).

Conventional univariate statistical analysis of functional data was performed on an average cortical surface, obtained by reconstructing both hemispheres for each participant. The surfaces were subsequently aligned using a cortex-based alignment procedure, an algorithm that uses individual curvature information to align corresponding gyri and sulci across participants (Goebel et al. 2006). The analysis was performed using a random effects General Linear Model (GLM), using ‘subject’ as random variable. A design matrix was created, using a predictor for both conditions. The predicted time courses were adjusted for the typical hemodynamic response delay by convolution with a canonical (double gamma) hemodynamic response function. No Statistical contrasts were considered at the multisubject whole-brain level, FDR corrected at q = 0.05. We subsequently performed a region-of-interest (ROI) analysis of the person > category contrast on the individual subject level. We focused on individual left and right FFA as ROI, as identified by the independent localizer run.

examples (20 per condition) and a testing set made of 8 examples (4 per condition). For each participant individually, a linear SVM type-1 (Muller et al. 2001; Mourao-Miranda et al. 2005), as implemented in BrainVoyager QX2.1, was used to differentiate the 2 patterns of activation (person, category). The test subset was used to assess the accuracy of the trained classifier. This process was repeated 40 times for each participant with different train and test subsets, after which the average accuracy was computed and tested against chance level using a one-sample t-test. See Figure 2 for a summary of the pattern classification approach used.

Results
A univariate RFX GLM of the entire cortical volume was performed, but statistical analysis of the person > category contrast showed no active voxels in both the left and right hemisphere. A ROI analysis of both left and right FFA was subsequently carried out on the individual subjects (see Table 2). FFA showed a clear response to the blocks of words, but this response was generally equal in both conditions. The statistical contrasts yielded no significant difference between the person and category contexts in any of the 6 left FFAs nor in 4 of the right.

In contrast, the multivariate pattern analysis revealed a difference between conditions. Results from the pattern classification are depicted in Table 3 and Figure 3. Even though the overall level of activity in both conditions was the same, the pattern of activation across voxels in FFA did reflect the semantic context of the word blocks. The classifier was able to discriminate the response patterns in FFA that underlie the 2 semantic contexts significantly above chance level in all participants (see Table 3). With one exception, all participants showed this distributed coding in FFA bilaterally. The multi-voxel effect found in FFA is without directionality toward 1 of the 2 conditions. A classification accuracy that significantly deviates from chance level tells us that the patterns of responses elicited by the conditions are sufficiently distinct from each other to get noticed by the classifier.

Discussion
We aimed to further delineate the functional role of FFA in the context of semantic association about people. The present study investigated if FFA is able to discriminate between a person-specific and a categorical semantic context, while keeping visual input constant across conditions. Using an IMRI
block design, we measured the response of FFA during presentation of blocks of words in which the words were either related in a categorical or a personal context. Both conditions elicited a relatively strong response in FFA, but univariate statistical contrasts yielded no difference in BOLD signal between them within FFA or elsewhere. Nonetheless, multivoxel pattern analysis revealed that activations in FFA do actually code for the semantic contexts used. This suggests that 1) multivariate pattern analysis was able to reveal neural coding of informational contents, where univariate analysis could show

### Table 2

<table>
<thead>
<tr>
<th>Subject</th>
<th>Left FFA</th>
<th>Right FFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
<td>( P )</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>1.22</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>-1.15</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>0.93</td>
<td>0.35</td>
</tr>
<tr>
<td>5</td>
<td>1.30</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>-0.69</td>
<td>0.49</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Left FFA</th>
<th>Right FFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Correct</td>
<td>( t )</td>
</tr>
<tr>
<td>1</td>
<td>62.2</td>
<td>4.78</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>4.37</td>
</tr>
<tr>
<td>3</td>
<td>60.9</td>
<td>3.90</td>
</tr>
<tr>
<td>4</td>
<td>55.9</td>
<td>2.38</td>
</tr>
<tr>
<td>5</td>
<td>61.3</td>
<td>4.68</td>
</tr>
<tr>
<td>6</td>
<td>56.3</td>
<td>2.21</td>
</tr>
</tbody>
</table>

### Figure 2

Summary of the analysis approach. 1. An event-related average plot shows the univariate contrast between person (red line) and category (blue line) context during the presentation of words (yellow interval, [0,8] volumes) in left FFA (defined by independent localizer), in which no significant difference was observed. As can be seen in the top right inset, the BOLD signal is at baseline prior to the presentation of the block of words, indicating no remaining effect of the instruction images on FFA activity. 2. The matrices represent the responses of the selected voxels (x-axis) for all trials (y-axis). The trials are randomly assigned to a training set containing 20 labeled trials from each condition, and a test set with 4 unlabeled trials from both conditions. 3. A linear SVM is trained with the train set. 4. The accuracy of the classifier after training is assessed with the independent test set. The trials in the test set are not used in the training phase, guaranteeing an unbiased measure of the accuracy of the classifier after training. Steps 2–4 are iterated 40 times, after which the mean accuracy is computed and tested against chance level of 50%. See text for details.
no differences between conditions—which has significant methodological implications—and 2) FFA codes for semantic contexts of word sets, presented free of direct face stimuli. In other words, although the words comprising the individual person profiles could be related to previously presented instructional face stimuli—no actual face stimuli were presented in the trials here analyzed. The fact that activity in FFA nonetheless was found to contain sufficient information about the semantic contexts of the trials, has significant and far-reaching implications for models about the role of FFA in face and identity coding. This we discuss in more detail below.

In light of the role of FFA in face identification (Gauthier et al. 2000; Hoffman and Haxby 2000; Barton et al. 2002; Haxby et al. 2002; Andrews and Ewbank 2004; Grill-Spector et al. 2004; Loffler et al. 2005; Kriegeskorte et al. 2007; Mur et al. 2010), our findings illustrate—for the first time—the possibility that the coding of face identity in FFA goes beyond visual features and might include conceptual and biographical knowledge about a face in the absence of visual facial information.

In the present design, we carefully attempted to equalize both conditions by selecting the words belonging to each condition from the same word matrix. However, the instructional images we presented at the beginning of each block differed between conditions. As the neural responses elicited by the images were irrelevant for our hypothesis, we excluded the corresponding volumes from our analysis. We can also rule out the influence of the instruction over the block of words, and, in turn, on the contextual processing. The long interval between the instruction and the block of words (10,000 ms) ensured that the hemodynamic response related to the instructional images returned to baseline before the block onset. This prevented spilling-over effects from the image (as can be seen in Fig. 2). This approach excluded any direct interference from the instruction with the activation pattern within in FFA during the run. Still, one could argue that FFA plays a role in the working memory of faces. From this perspective, the face that is presented prior to the block of words could potentially bias the activity in FFA in the person-specific condition, offering an alternative explanation for our findings. Several studies, however, have demonstrated that the maintenance of faces in working memory exhibit delay-period activation in FFA (Druzgal and D’Esposito 2003; Postle et al. 2003; Ranganath et al. 2004; Johnson et al. 2007). Our results show that the BOLD signal in FFA returns to baseline in the interval between the presentation and the block of words, indicating no transient effect of working memory. In addition, an effect of face working memory on the activation of FFA would most likely result in a stronger BOLD signal in the person-specific condition compared with the category condition, an effect that was not apparent in our data. These 2 points indicate that it is unlikely that the picture of the face is directly responsible for the difference between the 2 experimental conditions.

The exact source of the observed person versus category effect in FFA, however, remains subject to discussion. It is possible that FFA receives top-down modulation from one or multiple areas from the semantic (person or category identification) network (Martin and Chao 2001; Martin 2007) that were involved in the processing of the contexts during the task. However, since the univariate contrast between the person and category conditions did not reveal any areas of interest, the possible source of any modulatory effects remains subject to speculation. Several studies observed a wide variety of modulatory effects on FFA, ranging from modulation by face working memory load (Druzgal and D’Esposito 2001), likely via feedback from prefrontal cortex (Druzgal and D’Esposito 2003), task specificity (Reddy et al. 2007; Chiu et al. 2010), face/nonface category expectation (Puri et al. 2009), and language describing faces (Aziz-Zadeh et al. 2008). Moreover, a study by von Kriegstein et al. (2005) demonstrated that familiar voices activate FFA during a speaker identification task. Based on what has been established about the role of aIT cortex in retrieval of semantic knowledge and face identification (Sergent and Signoret 1992; Evans et al. 1995; Haxby et al. 2000; Grabowska et al. 2001; Gainotti et al. 2003; Tranel 2006; Kriegeskorte et al. 2007; Patterson et al. 2007; Brambati et al. 2010) one might suspect this brain region to be involved in feedback back conceptual information to FFA. Still, it has to be taken into consideration that participants in the current study were provided with semantic information online during the task, which is in contrast with studies that had the participants study conceptual knowledge ahead of the scanning task, allowing learned information about identity to consolidate. It is therefore possible that aIT did not directly modulate FFA. The actual source might be part of the face identification network and that aIT acts as a semantic hub (Patterson et al. 2007), communicating with other areas involved in perception. Future studies might investigate the functional or effective connectivity between FFA and other areas to resolve this debate, for instance with use of Granger causality mapping (Goebel et al. 2003; Roebroeck et al. 2005), dynamic causal modeling (Friston et al. 2003), or methods based on causal entropy (Hinrichs et al. 2006).

Alternatively, the person versus category effect within FFA can be a local process within FFA and adjacent areas without input from temporal and frontal sites. In general, the early position of FFA in the visual system hierarchy (Haxby et al. 1991) may suggest that FFA likely does not directly process semantics. But there are empirical estimates on the time course of conceptual encoding (Thorpe et al. 1996; Schmitt et al. 2001; Indefrey and Levelt 2004) that estimate conceptual encoding starting around 0–150 ms after stimulus onset. Even-related potential signatures related to facial feature and identity encoding are within this time window (N170 and N250) with estimated sources within occipital face area (OFA) and FFA, respectively (Bentin et al. 1996; Schweinberger et al. 2002; Kaufmann et al. 2009). Therefore, it may well be that facial
encoding and conceptual/semantic access run in parallel. The access might take place in a distributed manner within face selective regions, as shown by Haxby and colleagues. They discussed the distributed representation of faces, cats, and objects in human ventral temporal cortex (Haxby et al. 2001). They show that the distinct pattern of responses associated with each of the categories is not exclusively due to regions that show maximal response to that category. In fact, when these regions are excluded from the analysis, looking at the patterns of responses of the remaining areas was sufficient to identify the category being watched by the participant. In what they termed object form topography, they illustrate that complex categorical and semantic features underlying visual objects might be represented by distributed patterns across category-specific regions. Hypothetically, presenting a semantic context related to one of these categories could elicit a spatially distinct pattern of responses that to some extent corresponds to the category the context refers to.

Conclusion
In the current study, we compared the responses of bilateral FFA with words that were visually identical—but organized in different sets to define 2 conditions: a person context and a category context. Our results showed that multivoxel pattern analysis could reveal differential coding in FFA for the 2 types of semantic associations (semantic contexts). Univariate analysis showed no difference in overall response. Overall, the data showed hypoperfusion in the left middle frontal gyrus. This has strong implications for models of the role of FFA in semantic and identity coding. Future research should aim to reveal how FFA interacts with other regions related to semantic and memory processing, to uncover the mechanism of this differential coding.

Funding
Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) ASPASIA (grant no. 015.001.053).

Notes
The authors would like to thank Valérie Goffaux and Russel Epstein for providing the stimuli for the FFA localizer and Tom de Graaf, Mariele Mur and Sanne Rutten for valuable discussions and comments on earlier versions of this manuscript. Conflict of Interest: None declared.

References


