Abstract

We used event-related fMRI to study recognition memory of newly learned faces. Caucasian subjects memorized unfamiliar, neutral and happy South Korean faces and 4 days later performed a memory retrieval task in the MR scanner. We predicted that previously seen faces would be recognized faster and more accurately and would elicit stronger neural activation than novel faces. Consistent with our hypothesis, novel faces were recognized more slowly and less accurately than previously seen faces. We found activation in a distributed cortical network that included face-responsive regions in the visual cortex, parietal and prefrontal regions, and the hippocampus. Within all regions, correctly recognized, previously seen faces evoked stronger activation than novel faces. Additionally, in parietal and prefrontal cortices, stronger activation was observed during correct than incorrect trials. Finally, in the hippocampus, false alarms to happy faces elicited stronger responses than false alarms to neutral faces. Our findings suggest that face recognition memory is mediated by stimulus-specific representations stored in extrastriate regions; parietal and prefrontal regions where old and new items are classified; and the hippocampus where veridical memory traces are recovered.

Keywords: Encoding; False memory; fMRI; Retrieval

1. Introduction

Face perception is a highly developed visual skill in humans. Functional brain imaging studies have suggested that face processing is mediated by activation within a distributed neural system that includes visual, limbic and prefrontal regions [12,18]. The cortical network for face perception includes regions in extrastriate cortex that process the identification of individuals [11], the superior temporal sulcus, where gaze direction and speech-related movements are processed [14], the amygdala and insula, where facial expressions are processed [17,28], and regions of the reward circuitry, including the nucleus accumbens and orbitofrontal cortex, where the assessment of facial beauty is processed [1,15,26].

Previous face memory studies have compared encoding with retrieval [13], encoding of faces with encoding of objects [21], retrieval of target faces with retrieval of non-target faces [6], or recognition of famous faces with recognition of newly learned faces [24]. As these studies tested within-session memory effects, it is currently unknown how face representations are stored and retrieved from long-term memory. The aim of this study was to investigate recognition memory of newly learned, unfamiliar faces. To that end, we instructed Caucasian subjects, who had limited exposure to Asian faces in their environment, to memorize unfamiliar, South Korean faces, and tested their recognition memory in the MR scanner. We hypothesized that relative to novel South Korean faces, the previously seen faces would be recognized faster and more accurately and would elicit stronger activation in face-responsive regions. Moreover, based on previous reports of valence enhancement [17,27,41], we predicted facilitated behavioral and neuronal responses to happy faces as compared with neutral faces.

The second aim of the study was to investigate the neural correlates of false memory. Due to the visual similarity between the South Korean faces, we predicted that subjects would mistake novel faces for previously seen faces and would therefore make many incorrect responses. We have recently shown that correctly recognized, novel art paintings, which were visually similar to previously seen paintings and were likely to be mistaken for familiar ones, evoked decreased activation in the hippocampus [44]. We therefore predicted enhanced activation in the hippocampus for incorrectly recognized novel faces, namely for false alarms.
Our results indicate that recognition memory of newly learned faces is mediated by a distributed cortical network that includes visual, parietal and prefrontal regions. Within all regions, previously seen faces evoked stronger activation than novel faces. Moreover, in parietal and prefrontal cortices, activation during correct trials was stronger than activation during incorrect trials. Finally, in the hippocampus, false alarms to happy faces evoked stronger activation than false alarms to neutral faces.

2. Methods

2.1. Subjects

Twelve normal, right-handed subjects (six males, mean age 25 ± 2.5 years) with normal vision participated in the study. All subjects gave informed written consent for the procedure in accordance with protocols approved by the University Hospital. All subjects were Caucasian students from Zurich University who were unfamiliar with Asian faces (according to the 2000 Swiss census only 1.3% of the population are Asians).

2.2. Stimuli and tasks

Stimuli were displayed using Presentation (http://www.neurobs.com, version 9.13) and were projected with a magnetically shielded LCD video projector onto a translucent screen placed at the feet of the subject. During the encoding session, subjects were instructed to memorize unfamiliar, South Korean faces (32 neutral, 32 happy, see http://www.wisdom.weizmann.ac.il/~michel/bc/cauc_kor_ims.html). In a separate behavioral pilot, five naive subjects, who did not participate in the fMRI experiment, rated the visual similarity between selected neutral and happy faces and the reminder of faces from the database, on a scale from 0 (dissimilar) to 3 (very similar). To induce false recognition during retrieval, novel faces were chosen with similarity scores higher than 1.5. During encoding, each picture was presented on the center of the screen for 5 s and was randomly repeated four times. Four days later, subjects performed a memory retrieval task in the MR scanner. In an event-related design, the familiar faces were randomly mixed with novel faces (32 neutral, 32 happy). Pictures were presented every 3 s and subjects pressed a button to indicate whether they had seen each face before. In each run, four face epochs of 24 s each alternated with four fixation epochs of 15 s each (total of 156 s).

2.3. Data acquisition

Data were collected using a 3T Philips Intera whole body MR scanner (Phillips Medical Systems, Best, The Netherlands). Changes in blood-oxygenation level-dependent MRI signal were measured by using sensitivity encoded gradient-echo echoplanar sequence (SENSE) [30], with 35 axial slices, TR = 3 s, TE = 35 ms, flip angle = 52°, field of view = 220 mm, acquisition matrix = 80 × 80, reconstructed voxel size = 1.72 mm × 1.72 mm × 4 mm, and SENSE acceleration factor R = 2.

For each subject, high-resolution spoiled gradient-recalled echo structural images were collected (180 axial slices, TR = 20 ms, TE = 2.3 ms, field of view = 220 mm, acquisition matrix = 224 × 224, reconstructed voxel size = 0.9 mm × 0.9 mm × 0.75 mm). These high-resolution T1 images provided detailed anatomical information for the region of interest (ROI) analysis and were used for 3D normalization to the Talairach and Tournoux brain atlas [40].

2.4. Data analysis

Accuracies and reaction times were calculated for each subject. To compare correct recognition of previously seen and novel faces, repeated-measures ANOVAs were computed for accuracies and reaction times with stimulus type (familiar/novel) and valence (happy/neural) as factors.

Functional MRI data were analyzed in BrainVoyager QX Version 1.3 (Brain Innovation, Maastricht, The Netherlands). All volumes were realigned to the first volume, corrected for motion artifacts and spatially smoothed using a 5 mm FWHM Gaussian filter. The main effect of faces (activation evoked by faces versus activation evoked by the fixation cross) was analyzed using multiple regression with box-car functions that were convolved with a canonical hemodynamic response function [10]. A set of face-responsive ROIs was anatomically defined for each subject with clusters that showed a significant effect (p < 0.0001). These regions included the inferior occipital gyrus (IOG), lateral fusiform gyrus (LFG), superior temporal sulcus (STS), intraparietal sulcus (IPS), amygdala, insula, inferior frontal gyrus (IFG) and anterior cingulate cortex (ACC). Additionally, the hippocampus was anatomically defined in subjects who showed significant activation (p < 0.05, uncorrected) when the response to previously seen faces was contrasted with the response to novel faces. Nine subjects showed significant activation in the right hippocampus and eleven subjects showed significant activation in the left hippocampus.

In each subject and each ROI, the mean parameter estimates were calculated for stimulus (familiar/novel), valence (neutral/happy) and trial (correct/incorrect) type. These values were used for between-subjects random-effects analyses and for repeated-measures ANOVAs with Greenhouse–Geisser correction.

3. Results

3.1. Behavioral data

Previously seen faces were recognized faster and more accurately (1182 ms; 79%) than novel faces (1535 ms, F1,11 = 35.01, p < 0.0005; 59%, F1,11 = 8.5, p < 0.05). Happy faces were recognized more accurately (familiar: 80%; novel: 61%) than neutral faces (familiar: 77%; novel: 58%), however this difference did not reach statistical significance (F1,11 = 1.72, p = 0.22).

The comparison of reaction times during correct and incorrect trials revealed a significant interaction between stimulus (familiar/novel) and trial (correct/incorrect) type (F1,11 = 33.3, p < 0.0005). Thus, reaction times for previously seen faces were longer in incorrect than in correct trials (F1,11 = 37.15, p < 0.0001), whereas reaction times for novel faces were longer in correct than incorrect trials (F1,11 = 11.1, p < 0.01) (Fig. 1).

3.2. Imaging data

The main effect of faces revealed activation in a distributed cortical network that included visual, parietal and prefrontal regions (Table 1). Hereafter we describe the differential activation evoked by previously seen and novel faces within several ROIs.

3.3. Activation in visual and parietal regions

In the visual cortex, bilateral activation was found in the IOG and LFG. As activation in the left and right hemispheres was very similar, the parameter estimates were averaged across both hemispheres (Fig. 2). In both regions, hits evoked stronger activation than misses, correct rejections, and false alarms (IOG: F3,33 = 11.34, p < 0.0005; LFG: F3,33 = 11.3, p < 0.0005). Regardless of stimulus type, activation was stronger in correct than incorrect trials (IOG: F1,11 = 18.4, p < 0.005; LFG: F1,11 = 18.9, p < 0.005). Furthermore, a significant interaction was found between stimulus type and valence: previously seen, neutral faces evoked stronger activation than novel ones, whereas previously seen, happy faces and novel, happy faces
evoked similar activation (IOG: $F_{1,11} = 9.9, p < 0.01$; LFG: $F_{1,11} = 5.3, p < 0.05$).

The main effect of faces also revealed bilateral activation in the intraparietal sulcus (IPS) (Fig. 2). In this region, which is implicated in many attention-related and memory retrieval tasks, hits evoked stronger activation than misses, correct rejections and false alarms ($F_{1,10} = 7.8, p < 0.005$). Additionally, regardless of stimulus type, activation in correct trials was stronger than activation in incorrect trials ($F_{1,10} = 21.1, p < 0.005$). The difference between activation evoked by neutral and happy faces was not significant.

### 3.4. Activation in prefrontal cortex

Faces evoked bilateral activation in the IFG, insula, and ACC. Within these regions, (Fig. 3, top), stronger activation was found during correct trials for both neutral and happy faces (IFG: $F_{1,7} = 9.3, p < 0.05$; insula: $F_{1,11} = 8.5, p < 0.05$; ACC: $F_{1,11} = 10.4, p < 0.01$).

In the left IFG (Fig. 3, bottom), contrasting activation evoked by misses with activation evoked by false alarms revealed an interaction with valence ($F_{1,7} = 10.2, p < 0.05$). Thus, false alarms to happy faces evoked stronger activation than false alarms to neutral faces. The difference between activation elicited by misses to happy faces and activation evoked by misses to neutral faces was not statistically significant.

### 3.5. Activation in the hippocampus

Activation in the hippocampus was localized when responses evoked by previously seen faces were contrasted with responses evoked by novel faces. As shown in Fig. 4, hits elicited stronger activation than misses, correct rejections and false alarms in both hemispheres ($F_{3,21} = 7.3, p < 0.01$). Regardless of stimulus type,

![Behavioral data](image)

#### Table 1

<table>
<thead>
<tr>
<th>Region</th>
<th>$N$</th>
<th>Volume ($\text{cm}^3$)</th>
<th>$x$ (mean ± S.E.M.)</th>
<th>$y$ (mean ± S.E.M.)</th>
<th>$z$ (mean ± S.E.M.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L IOG</td>
<td>12</td>
<td>12.3 ± 0.4</td>
<td>−39 ± 0.8</td>
<td>−75 ± 1.0</td>
<td>−14 ± 1.0</td>
</tr>
<tr>
<td>R IOG</td>
<td>12</td>
<td>13.2 ± 0.1</td>
<td>38 ± 1.0</td>
<td>−75 ± 1.0</td>
<td>−14 ± 1.1</td>
</tr>
<tr>
<td>L LFG</td>
<td>12</td>
<td>12.3 ± 0.3</td>
<td>−34 ± 0.8</td>
<td>−49 ± 1.2</td>
<td>−17 ± 0.9</td>
</tr>
<tr>
<td>R LFG</td>
<td>12</td>
<td>12.6 ± 0.3</td>
<td>33 ± 1.2</td>
<td>−48 ± 1.3</td>
<td>−18 ± 1.1</td>
</tr>
<tr>
<td>L IPS</td>
<td>12</td>
<td>10.8 ± 0.5</td>
<td>−34 ± 1.4</td>
<td>−47 ± 1.3</td>
<td>43 ± 1.7</td>
</tr>
<tr>
<td>R IPS</td>
<td>11</td>
<td>10 ± 1.0</td>
<td>30 ± 1.6</td>
<td>−52 ± 1.7</td>
<td>40 ± 1.3</td>
</tr>
<tr>
<td>L STS</td>
<td>5</td>
<td>4.4 ± 0.4</td>
<td>−48 ± 1.8</td>
<td>−44 ± 2.1</td>
<td>5 ± 1.9</td>
</tr>
<tr>
<td>R STS</td>
<td>8</td>
<td>3.5 ± 1.1</td>
<td>49 ± 2.2</td>
<td>−48 ± 2.3</td>
<td>8 ± 1.4</td>
</tr>
<tr>
<td>L AMG</td>
<td>5</td>
<td>7.6 ± 1.3</td>
<td>−17 ± 1.2</td>
<td>−8 ± 1.4</td>
<td>−11 ± 1.7</td>
</tr>
<tr>
<td>R AMG</td>
<td>7</td>
<td>7.4 ± 1.4</td>
<td>17 ± 0.9</td>
<td>−8 ± 0.8</td>
<td>−10 ± 0.6</td>
</tr>
<tr>
<td>M ACC</td>
<td>12</td>
<td>12 ± 0.3</td>
<td>1 ± 0.7</td>
<td>15 ± 1.1</td>
<td>45 ± 1.1</td>
</tr>
<tr>
<td>L IFG</td>
<td>8</td>
<td>9.2 ± 0.7</td>
<td>−41 ± 2.5</td>
<td>17 ± 2.9</td>
<td>25 ± 2.1</td>
</tr>
<tr>
<td>R IFG</td>
<td>12</td>
<td>11.8 ± 0.4</td>
<td>44 ± 1.8</td>
<td>11 ± 2.0</td>
<td>26 ± 1.8</td>
</tr>
<tr>
<td>L INS</td>
<td>12</td>
<td>11.4 ± 0.6</td>
<td>−31 ± 1.1</td>
<td>20 ± 0.9</td>
<td>6 ± 1.0</td>
</tr>
<tr>
<td>R INS</td>
<td>12</td>
<td>12.2 ± 0.4</td>
<td>33 ± 0.9</td>
<td>20 ± 0.8</td>
<td>6 ± 1.1</td>
</tr>
</tbody>
</table>

Significant clusters ($p < 0.0001$) were selected for each subject based on the contrast between faces and fixation. $N$ indicates the number of subjects who showed significant activation in each region. Volumes were calculated before spatial normalization. Coordinates are in the normalized space of the Talairach and Tournoux atlas [40]. L, left; R, right; M, medial.
Fig. 2. Activation in visual and parietal regions. Top: from left to right, coronal sections, taken from individual subjects, illustrating activation in the inferior occipital gyrus (IOG), lateral fusiform gyrus (LFG), and intraparietal gyrus (IPS). Bottom: mean parameter estimates were averaged across all subjects and both hemispheres. In this and subsequent figures: hits, correctly recognized, previously seen faces; misses, incorrectly recognized, previously seen faces; correct rejections, correctly recognized novel faces; false alarms, incorrectly recognized novel faces.

Fig. 3. Activation in prefrontal regions. Top: a coronal section, taken from one individual, illustrating activation evoked by faces in the inferior frontal gyrus (IFG), insula and anterior cingulated cortex (ACC). Bottom: a coronal section, taken from one individual, showing bilateral activation in the IFG. Mean parameter estimates were averaged across all subjects.
activation in correct trials was stronger than activation in incorrect trials in both hemispheres ($F_{1, 7} = 10.1, p < 0.05$). Additionally, a significant interaction between valence and hemisphere was found ($F_{1, 7} = 12.1, p < 0.05$), indicating that in the right hemisphere false alarms to happy faces evoked stronger activation than false alarms to neutral faces.

4. Discussion

We used a unique set of stimuli, namely South Korean faces, to study recognition memory of newly learned, visually similar pictures. Behaviorally, our subjects recognized previously seen faces faster and more accurately than novel faces. It is of interest that response latencies for misses were significantly longer than the response latencies for hits. Moreover, response latencies for hits were significantly longer than the latencies for false alarms. These behavioral data suggest that it took subjects longer to decide whether a face was novel. Presumably due to the high visual similarity between the South Korean faces, subjects hesitated before pressing the buttons, pondering whether they had previously seen these faces.

The main effect of faces revealed activation within a distributed neural system [12,18]. Within the extrastriate face-specific regions, namely the IOG and LFG, previously seen faces evoked stronger activation than novel faces. We have recently found that recognition memory of familiar, as compared with novel, art paintings was associated with enhanced activation in face- and object-selective regions in the visual cortex [44]. These findings suggest that encoding of visual stimuli results in stimulus-specific representations, consistent with numerous visual imagery [19,25], working memory [9] and associative memory retrieval [31] studies, which showed stimulus-specific memory traces in the human visual cortex.

Contrary to our prediction, our data indicate that recognition memory of happy faces was very similar to the recognition of neutral faces. In both extrastriate regions, regardless of stimulus type (previously seen or novel), happy and neutral faces evoked similar activation. A speculative explanation may be the nature of the stimuli used in our experiment, namely unfamiliar South Korean faces. Future studies will determine the extent to which recognition memory of unfamiliar, other-race faces is associated with valence enhancement.

Activation in the IPS revealed stronger responses to the familiar than to the novel faces. The IPS, a region of the dorsal frontoparietal attention network, has been implicated in many cognitive studies of attention, specifically in target detection tasks [7,22,37]. In our task, previously seen faces, for which the correct response was “Yes, I had seen these faces before”, were mixed with novel, never seen before faces. It is likely that subjects were searching for the previously seen faces and detecting them as “targets”. This enhanced response to familiar items is also consistent with the role of parietal regions in mediating the segmentation of old from new items [29,44]. Numerous recognition memory studies have further shown that posterior parietal cortex does not merely ‘detect old items’ but, rather, mediates higher order cognitive processes associated with memory retrieval [23,35,43]. Based on these empirical findings, three hypotheses have been recently suggested to account for the role of posterior parietal cortex in memory retrieval: attention to internal mnemonic representations; accumulation of sensory signals in order to reach a decision; and a memory buffer for stored information [42].

In prefrontal cortex, we found activation in the IFG, insula and ACC. Within these regions, responses during correct trials, regardless of stimulus type (familiar or novel faces) were stronger than responses during incorrect trials. Previous functional brain imaging studies have implicated the prefrontal cortex in memory formation [3] and monitoring during retrieval [2]. Enhanced activation in parietal and prefrontal regions in response to familiar items has been reported in numerous fMRI studies [20]. This “old-new” parieto-frontal effect is also consistent with the ERP correlates of recollection, namely a positive shift in waveforms elicited by correctly classified old items relative to the waveforms evoked by new items in the left parietal cortex, and a sustained positive shift elicited by old items in the right prefrontal cortex (reviewed in [34]). Importantly, most functional brain imaging studies of episodic memory retrieval have used words, and not pictures, as stimuli [4,33]. Moreover, our task required a simple “Yes–No” recognition, whereas other studies of episodic retrieval employed the “Remember–Know” procedure [36]. Finally, the ERP and fMRI techniques detect different signals, namely scalp electrical activity and the hemodynamic response, which reflect direct and indirect measures of neuronal activity,
respectively. Thus, hemispheric asymmetries should be carefully interpreted.

In the hippocampus, the previously seen faces evoked stronger responses than novel faces. Previous studies have reported activation in the hippocampus in various memory-related processes, including recognition memory [39], maintenance in working memory [32], source memory [8] and generation of visual images from short-term memory [16]. Based on our previous study [44], we predicted enhanced responses to the incorrectly recognized novel faces. We found that activation evoked during misses was very similar to the activation elicited during false alarms. Interestingly, activation evoked by false alarms to happy faces was stronger than activation elicited by false alarms to neutral faces, suggesting that novel, happy faces were more susceptible to be mistaken as familiar ones. Emotional faces therefore seem more likely to induce illusory memories than neutral faces. It is currently unclear to what extent the hippocampus mediates the recovery of true and false memory traces [5,38] and future studies will determine the role of valence in the formation of veridical and illusory memories.

In sum, our study shows that retrieval of newly learned faces, as compared with novel ones, results in enhanced activation within a distributed cortical network that includes visual, parietal and prefrontal regions. Within this network, stimulus-specific representations are stored in face-responsive regions in extrastriate cortex, parietal and prefrontal regions mediate the retrieval and classification of old and new items, and the hippocampus mediates the recovery of true memory traces.

Acknowledgement

This study was supported by the Swiss National Center for Competence in Research: Neural Plasticity and Repair.

References