Medial prefrontal and occipito-temporal activity at encoding determines enhanced recognition of threatening faces after 1.5 years

Xiqin Liu1 · Xinqi Zhou1 · Yixu Zeng1 · Jialin Li1 · Weihua Zhao1 · Lei Xu1 · Xiaoxiao Zheng1 · Meina Fu1 · Shuxia Yao1 · Carlo V. Cannistraci2,3 · Keith M. Kendrick1 · Benjamin Becker1

Abstract
Studies demonstrated that faces with threatening emotional expressions are better remembered than non-threatening faces. However, whether this memory advantage persists over years and which neural systems underlie such an effect remains unknown. Here, we employed an individual difference approach to examine whether the neural activity during incidental encoding was associated with differential recognition of faces with emotional expressions (angry, fearful, happy, sad and neutral) after a retention interval of > 1.5 years (N = 89). Behaviorally, we found a better recognition for threatening (angry, fearful) versus non-threatening (happy and neutral) faces after a delay of > 1.5 years, which was driven by forgetting of non-threatening faces compared with immediate recognition after encoding. Multivariate principal component analysis (PCA) on the behavioral responses further confirmed the discriminative recognition performance between threatening and non-threatening faces. A voxel-wise whole-brain analysis on the concomitantly acquired functional magnetic resonance imaging (fMRI) data during incidental encoding revealed that neural activity in bilateral inferior occipital gyrus (IOG) and ventro-medial prefrontal/orbitofrontal cortex (vmPFC/OFC) was associated with the individual differences in the discriminative emotional face recognition performance measured by an innovative behavioral pattern similarity analysis (BPSA). The left fusiform face area (FFA) was additionally determined using a regionally focused analysis. Overall, the present study provides evidence that threatening facial expressions lead to persistent face recognition over periods of > 1.5 years, and that differential encoding-related activity in the medial prefrontal cortex and occipito-temporal cortex may underlie this effect.

Keywords Behavioral pattern similarity analysis · Emotional expressions · Individual differences · Long-term face recognition · Medial prefrontal cortex · Occipito-temporal cortex

Introduction

For social species, the recognition of previously encountered conspecifics is vital for survival and successful interaction. In humans, faces are presumably the most important stimuli for subsequent recognition. Given the high evolutionary significance of these stimuli, cortical networks specialized in perceiving and recognizing faces develop already during early infancy (Powell et al. 2018; Cohen et al. 2019). Nevertheless, the ability to recognize faces varies greatly in the human population. While some individuals can recognize faces following a single exposure over years, others find it nearly impossible to recognize highly familiar faces (Russell et al. 2009; Tardif et al. 2019). In addition to individual differences, several characteristics of the facial stimuli can affect subsequent recognition including emotional expressions (Bruce and Young 1986; Haxby et al. 2000). From...
an evolutionary perspective, the emotional expressions may transmit important information such that threatening facial expressions (e.g., angry or fearful) can signal danger, and may thus relate to harm avoidance in the future (Darwin 1872; Staugaard 2010).

In support of this evolutionary hypothesis, some experimental studies have demonstrated a recognition advantage of threatening facial expressions across a variety of delays (Grady et al. 2007; Jackson et al. 2014; Pinabiaux et al. 2013; Sterinstrømer et al. 2016; Thomas et al. 2014; Wang 2013). For example, previous studies consistently reported that faces with threatening expressions (i.e., angry or fearful) are better remembered compared to non-threatening (e.g., neutral) faces in visual working memory (Jackson et al. 2014; Öhman et al. 2001; Sessa et al. 2011; Thomas et al. 2014; Vuilleumier 2002). Several studies on short-term memory also found a recognition advantage for threatening faces when the memory was tested immediately (i.e., minutes) after encoding (e.g., Grady et al. 2007; Pinabiaux et al. 2013; Wang 2013). Following a longer retention interval of 24 h, the recognition memory was better for fearful relative to neutral faces (Wang 2013). These findings broadly align with numerous studies indicating enhanced recognition for non-facial emotional stimuli (e.g., scenes or words), particularly high-arousing negative ones, which are more vividly and accurately remembered over retention intervals ranging from minutes to years (for a review, see Bowen et al. 2018). Meanwhile, other studies failed to observe memory enhancement of threatening faces across a variety of delays (minutes to 2 weeks, Anderson et al. 2006; Grady et al. 2007; Satterthwaite et al. 2009; Xiu et al. 2015). These contradictory results may be accounted for by factors such as the lack of statistical power (i.e., small number of subjects and face stimuli), different methodological approaches (e.g., memory performance indexes), and heterogeneity of sample characteristics such as age and gender which may relate to differences in face memory ability (Grady et al. 1995; Sommer et al. 2013). Particularly, various retention intervals and univariate analysis approaches employed in different studies may also have an effect on the mixed findings. Previous studies primarily used hypothesis-driven univariate approaches (e.g., ANOVA) to compare recognition performance for emotional relative to neutral faces. However, this approach discards the information related to response variability of the data due to averaging and dismisses individual differences that characterize face memory (Miendlarzewska et al. 2018; Tian et al. 2020). Also, the emotional expression effect on face memory after a long retention interval (i.e., years) has not been determined to date. Previous studies have reported better recognition performance for emotional scenes compared to neutral scenes selected from the International Affective Picture System (IAPS) 1 year after incidental encoding (e.g., Dolcos et al. 2005; Erk et al. 2010; Gavazzeni et al. 2012), and suggested that the memory advantage of emotional, particularly negative, materials is mediated by effects on consolidation facilitating a better remembering of emotional relative to neutral materials for time periods ranging from 24 h to several months (Yonelinas and Ritchey 2015). Based on these findings, we expected that this long-term beneficial effect of negative emotion on memory would also present in face stimuli which may serve the adaptive needs. Thus, the first aim of the present study was to systematically investigate the emotional expression effects on face recognition over an extended retention interval (> 1.5 years) in a large sample using both univariate and data-driven multivariate analysis approach.

A large body of neuroimaging studies suggests that long-term memory advantage for emotional material—primarily negative visual stimuli like threatening scenes—relative to neutral material is mediated by medial temporal lobe (MTL) regions, in particular the interaction between the amygdala and hippocampal formation during encoding and recognition (Dolcos et al. 2004; Mackiewicz et al. 2006; Ritchey et al. 2008; for a review, see Yonelinas and Ritchey 2015). In contrast, findings on the neural substrates of long-term emotional face recognition have not been well characterized. Human functional magnetic resonance imaging (fMRI) investigations suggest that face processing is supported by distributed neural systems (Haxby et al. 2000), including brain regions processing facial features and identity such as the face fusiform area (FFA) (Kanwisher et al. 1997; McCarthy et al. 1997) and the inferior occipital gyrus (IOG)/occipital face area (OFA) (Gauthier et al. 2000), and those processing social and emotional information (e.g., facial expression) such as the amygdala (Gobbini and Haxby 2007), inferior frontal gyrus (IFG) (Ishai et al. 2002) and orbitofrontal cortex (OFC) (Ishai 2008). Initial studies examining the subsequent memory effects for emotional faces have shown that activation associated with the successful encoding (i.e., remembered > forgotten) of emotional (fearful and happy) versus neutral faces were centered on prefrontal regions such as the IFG, dorsolateral prefrontal cortex (dLPFC) and OFC (Sergerie et al. 2005). More recent fMRI studies using individual difference approaches have additionally revealed that individual variations in immediate memory performance for negative, positive and neutral faces are associated with encoding-related differences in connectivity between occipito-temporal areas (e.g., FFA, IOG) and regions engaged in social-emotional processes such as the OFC (Geiger et al. 2016; Xiu et al. 2015). However, the encoding-related neural basis underlying long-term emotional face recognition after a lengthy retention interval (i.e., years) has not been systematically examined, especially using an individual difference approach. Compared to subsequent memory paradigms (Paller and
discriminative long-term recognition performance for faces with different emotional expressions.

Based on previous findings, we expected differential memory performance between faces with threatening expressions (i.e., angry and fearful) and faces with non-threatening expressions after a retention interval > 1.5 years in both univariate and multivariate analyses. On the neural level, we expected that individual differences in the emotional expression effects on long-term face recognition were accompanied by variations in encoding-related activity in brain regions implicated in emotional memory formation and face processing (e.g., OFC, amygdala, and visual cortex).

Materials and methods

Subjects

A total of 102 (53 males, age range 20–32) healthy, young right-handed Chinese students participated in this study which was part of a large-scale fMRI project (e.g., Li et al. 2019; Liu et al. 2020; Xu et al. 2020; Zhou et al. 2020). Due to incomplete behavioral and fMRI data (N = 7), extremely low hits and false alarms (hits < 1 and false alarms < 1, N = 4), or excessive head motion during fMRI scanning (N = 2), data from 13 subjects were excluded from both behavioral and fMRI analyses, resulting in N = 89 subjects (44 males, mean age = 23.80 ± 2.39 years) in the final analyses. Details on recruitment protocols and quality assessments are provided in the Supporting Information. The study was approved by the local ethics committee at the University of Electronic Science and Technology of China and in accordance with the latest revision of the Declaration of Helsinki. Written informed consent was obtained from each subject.

Stimuli

A total of 150 face stimuli were selected from two validated Asian facial expression databases: Chinese Facial Affective Picture System (Gong et al. 2011) and Taiwanese Facial Expression Image Database (TFEID) (Chen and Yen 2007) and were evenly divided into three sets. Facial expressions included angry, fearful, sad, happy and neutral (each from 30 different individual actors, 15 males). All facial stimuli were gray-scaled and covered with an oval mask to remove individual features (e.g., hair). The three face sets were used for incidental encoding (set 1), immediate recognition (set 2) and delayed recognition (set 3), respectively. The arousal and valence of each face set (50 faces, 10 for each expression category) was rated by an independent sample (n = 20, 10 males, mean age = 21.2 ± 0.70 years) before the experiment and the arousal ratings of emotional faces (angry, fearful, happy, sad) were higher as compared to neutral faces (all
ps < 0.001), whereas arousal ratings between emotional faces did not differ (all ps > 0.05) for each face set.

**Experimental procedure**

The present study employed a multiple-stage procedure including an incidental encoding phase and a subsequent memory phase (Fig. 1). All subjects initially underwent an event-related fMRI paradigm using an emotional face processing task (i.e., incidental encoding) between August, 2016 and October, 2017 (Time 1, T1). Fifty facial stimuli (set 1) were repeatedly presented over two subsequent runs with different pseudorandom sequence, balanced for facial expression and gender (5 min 12 s per run). Stimuli were shown for 2500 ms during which the subjects were required to judge the gender of the face by button press. After each trial, a jittered fixation cross was presented for 2000–5600 ms (mean ITI = 3800 ms, see Fig. 1). Stimuli were presented via E-prime 2.0 (Psychology Software Tools, USA, http://www.pstnet.com/eprime.cfm). Twenty minutes after the fMRI acquisition, subjects were asked to complete a surprise recognition memory test (immediate test) outside the scanner in which the 50 previously presented faces (set 1, targets) from the fMRI paradigm were intermixed with 50 new faces (set 2, lures). Subjects were instructed to indicate whether each face had been shown during the fMRI acquisition (forced choice: old versus new). Emotional arousal ratings for each face were additionally assessed after the immediate old/new recognition test using a 9-point Likert scale (1 = very weak to 9 = very strong). After a retention interval of > 1.5 years (interval range: 653–1113 days), 102 subjects agreed to participate in a surprise recognition test (delayed test) between July, 2019 and August, 2019 (Time 2, T2) in which target faces were intermixed with another set of 50 new faces (set 3, lures). In the delayed test, subjects were asked to rate their recognition confidence on a six-point scale (old vs new; 1 = definitely new to 6 = definitely old, see Fig. 1). The confidence rating approach was employed in the delayed test given that it reflects the strength and quality of the memory more precisely (Aly and Turk-Browne 2016; Stretch and Wixted 1998), and thus was more sensitive as compared to the categorical old/new judgement approach. Moreover, this allowed us to conduct multivariate analysis on the delayed test data with increased power. The delayed recognition memory test was carried out online via SurveyCoder 3.0 (https://www.surveycoder.com/).

**MRI data acquisition**

MRI data were obtained on a 3 T GE MRI system (General Electric, Milwaukee, WI, USA). Functional images were acquired with a gradient echo-planar imaging pulse sequence (39 slices; repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; slice thickness = 3.4 mm; spacing = 0.6 mm; field of view (FOV) = 240 × 240 mm2; flip angle = 90°; matrix size = 64×64). Each run of the emotional face processing task consisted of 173 volumes. High-resolution whole-brain T1-weighted images were additionally acquired to improve normalization of the functional images (spoiled gradient echo pulse sequence; 156 slices; TR = 6 ms;
TE = 1.964 ms; thickness = 1 mm; FOV = 256 × 256 mm2; flip angle = 9°; matrix size = 256 × 256).

**Behavioral data analysis**

**Univariate approach**

To assess whether subjects generally remembered faces immediately after encoding as well as > 1.5 years later, general sensitivity index A-prime (A′) and d-prime (d′) were initially computed and compared with their respective chance levels (for details, see Supplementary Methods). Hit rates were subjected to a 2 × 5 repeated-measures ANOVA with the factors time of assessment (immediate vs delayed) and emotional expression (angry vs fearful vs happy vs neutral vs sad) to examine the interaction effects between retention interval and facial expression. For both tests, hit rates were defined as the ratio of target faces correctly identified as old. For the delayed test, ratings of 4, 5 and 6 were considered as correctly identified. Post hoc tests for significant interactions with Holm correction were conducted to examine the facial expression effects within the immediate or delayed test, and planned two-tailed t tests were performed to determine changes of memory between immediate and delayed recognition in each facial expression condition. Notably, the analysis focused primarily on hit rates because it allowed a direct comparison between immediate and delayed test which included different sets of lures. We further compared the false alarm rates with chance level as well as among the expression conditions with each test to control for effects of a higher tendency to judge lure items as previously seen. ANOVA and post hoc analyses were analyzed using the R packages afex (Singmann et al. 2015) and emmeans (Lenth 2019).

**Principal component analysis (PCA)**

Given that the hypothesis-driven univariate analysis (e.g., ANOVA) might dismiss individual differences related to response variability (see Fig. 2A) due to averaging, we further investigated the robustness of long-term emotional expression effects using a data-driven unsupervised multivariate approach. PCA was applied given that it is one of the most widely used unsupervised multivariate machine learning algorithms for exploring the hidden pattern of multidimensional data (Ringnér 2008), which allowed us to detect and distinguish the factors that influence the memory performance in an unsupervised manner and visualize similarities and differences between individual behavioral responses (Fig. 2B). Specifically, the confidence ratings of all target face trials in the delayed test were plotted in a reduced two-dimensional space composed by principal component 1 (PC1) and PC2 (details see Supplementary Methods). PC1 explains the highest variance and represents the most discriminative dimension. We, thus, expected that trials with the same facial expression (color coded) would dominate separate regions within the reduced space along the PC1 axis, representing a discriminative expression-specific face memory pattern that threatening expressions (i.e., angry and fearful) would separate from non-threatening expressions (particularly happy and neutral, Fig. 2B). To further assess the significance of the segregation, we next employed a recently proposed measure ‘trustworthiness’ which compares the indicator of PC1 segregation (Area Under the ROC Curve, AUC) with a null distribution of AUC values derived from 1000 permutations of shuffled emotion labels (minimum two-tailed p value: 0.05, Fig. 2B) (details on this approach see Durán et al. 2021). Finally, Wilcoxon signed-rank tests (two tailed) with Benjamini–Hochberg-adjusted correction were performed to compare responses between expression conditions. Multivariate analyses including PCA, trustworthiness test of significance and non-parametric Wilcoxon Signed-Rank Tests were implemented in PC-corr MATLAB code (https://github.com/biomedical-cybernetics/PC-corr_net) which has been used in previous studies to successfully discriminate behavioral and omic patterns (Miedlarzewska et al. 2018; Ciucci et al. 2017). The multivariate PCA facilitated a sensitive determination of emotional expression-specific face recognition memory over an extended retention interval. In further exploratory analyses, the PCA was also applied to data from the immediate recognition test to provide a comparison with the delayed test.

**Behavioral pattern similarity analysis (BPSA)**

To quantify individual differences in representing a discriminative emotional face memory pattern in the delayed test, we devised a behavioral pattern similarity analysis (BPSA) approach by correlating the individual response pattern with the PC1 pattern (Fig. 2). The BPSA integrates PCA results with pattern similarity analysis which was proposed for multivariate fMRI data analysis (Haxby et al. 2001; Haxby 2012), providing a data-driven alternative for characterizing individual differences in discriminating complex conditions with high sensitivity. The BPSA includes the following procedures: (1) select a method for unsupervised dimensionality reduction of the multivariate data (without loss of generality in this study we consider PCA, Fig. 2A, B) and consider a data dimension as the response template that proves to offer a discriminative pattern between the samples (Fig. 2C, right). In our study, the projection of the samples on the first dimension of PCA embedding (which are the PC1 scores of the PCA on confidence ratings for targets of the delayed test) was extracted as a response template, which represents a discriminative variability that accounted for the facial expression conditions; (2) a measure of similarity (e.g., Pearson’s
correlation coefficients) between each subject’s confidence rating pattern and the PC1 score template was calculated to evaluate the extent to which the discriminative group-level emotional expression-specific face representation manifested at the individual level (Fig. 2C), and the similarity scores were then used to be associated with the brain activations. This BPSA method, thus, provided us with similarity score of each subject which could be used to characterize the individual emotional face representation and could inform an individual difference informed fMRI analyses to uncover the neural mechanisms underlying the long-term emotional expression effect for face recognition. To avoid inflation of the similarity calculation, the PC1 template was constructed employing a leave-one-subject-out approach, and specifically the individual whose correlation was calculated was excluded from constructing the template. For further exploratory analysis on the level of the fMRI data, we additionally capitalized on these values to split the sample into two groups with different behavioral patterns. To this end, we separated the subjects based on the similarity such that those whose confidence rating patterns showed significant correlations ($p < 0.05$, Pearson correlation) with the PC1 template were considered discriminators exhibiting a discriminative expression-specific face memory pattern and those whose similarity did not reach significance ($p > 0.05$) were considered non-discriminators (Fig. 2D).

**fMRI data analysis**

**Image preprocessing**

The functional MRI data were preprocessed and analyzed using SPM12 (Statistical Parametric Mapping, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The first ten volumes were discarded to allow for MR equilibration. The remaining functional images were realigned to correct for

---

**Fig. 2** Schematic presentation of the major steps in the multivariate investigation of the long-term emotional face recognition memory. A Representation of individual behavioral response patterns for all subjects. Each column represents the confidence rating pattern for one subject. Rows from top to bottom indicate face images with expressions angry, fearful, sad, neutral and happy. Color blue to red designates the confidence rating responses from 1 to 6. B PCA dimensionality reduction and expected discriminative pattern of the facial expression conditions with sample plotted in the 2D reduced space. The inset shows the assessment of significance of segregation using a non-parametric permutation test (i.e., trustworthiness). C Behavioral pattern similarity analysis (BPSA) between each subject’s response pattern and the PC1 score pattern (template) derived from PCA. The PC1 template was constructed excluding the individual whose correlation was calculated with the template for $N = 89$ times. D Grouping the subjects based on the significance of similarity scores. The subgroup of subjects exhibiting significant discriminative response pattern for facial expression conditions were called discriminators, and the subgroup not exhibiting such pattern were called non-discriminators.
head motion, co-registered with the T1-weighted structural images and normalized to Montreal Neurological Institute (MNI) standard template using a two-step procedure including segmentation of the brain structural images and application of the resultant transformation matrix to the functional time-series. The resampled voxel size of functional data was 3 × 3 × 3 mm. Finally, the images were spatially smoothed using a Gaussian kernel with full-width at half-maximum (FWHM) of 8 mm.

**Statistical analyses**

To identify the neural substrates associated with the individual differences in long-term emotional expression effects on face memory, a voxel-wise brain–behavior correlation analysis was conducted on the whole-brain level. To this end, the first-level contrast of interest (i.e., threatening vs non-threatening) was modeled using separate onset regressors for all trials with threatening vs non-threatening faces in accordance with the behavioral results, and convolved with the conventional hemodynamic response function (HRF). Six motion parameters were added in the design matrix to control for movement-related artifacts. Next, the first-level contrast images were subjected to a simple regression model with similarity score as a regressor and interval day as a nuisance covariate using the Statistical nonParametric Mapping toolbox (SnPM13, http://warwick.ac.uk/snpm) based on 5000 permutations to identify the regions in which the activity was correlated with individual differences in emotional face recognition representational pattern. As a parallel exploratory—yet not independent—approach, the association analysis was accompanied by a between-group differences approach using a whole-brain two-sample t test on two subgroups of subjects separated by the significance of the similarity scores (i.e., discriminators vs non-discriminators) to further describe the group difference between subjects with “qualitatively different” emotional face presentation (this analysis represents a related analysis and, thus, details are presented in Supplementary Methods). Significant clusters in the whole brain were determined using a height threshold of \( p < 0.001 \) (two tailed) and an extent threshold of \( p < 0.05 \) (two tailed) with cluster-based familywise error (FWE) correction (Eklund et al. 2016; Slotnick 2017).

Given that the amygdala has been strongly implicated in long-term (i.e., ranging from 1 to 4 weeks) emotional memory (Hamann et al. 1999; Mackiewicz et al. 2006; Ritchey et al. 2008), as well as that the fusiform gyrus plays a pivotal role in face recognition (Kanwisher et al. 1997; McCarthy et al. 1997), we examined effects in the amygdala and fusiform gyrus with increased sensitivity using a priori region-of-interest (ROI) analysis. Bilateral amygdala masks were created from the Brainnetome atlas (http://atlas.brainnetome.org/download.html, Fan et al. 2016). Bilateral fusiform masks were obtained from the probabilistic activation map (PAM) for faces (face vs objects) from the brain activity atlas (http://www.brainactivityatlas.org, Zhen et al. 2015). Small volume correction (SVC) was performed using FWE correction with a voxel-level threshold of \( p < 0.05 \).

**Results**

Overall, subjects successfully discriminated target faces (all expression conditions) from lure faces in both the immediate test (mean \( A' = 0.65, t_{88} = 15.74, p < 0.001 \), Cohen’s \( d = 1.67 \); mean \( d' = 0.50, t_{88} = 14.15, p < 0.001 \), Cohen’s \( d = 1.50 \)) and the delayed test (mean \( A' = 0.54, t_{88} = 3.70, p < 0.001 \), Cohen’s \( d = 0.40 \); mean \( d' = 0.11, t_{88} = 3.65, p < 0.001 \), Cohen’s \( d = 0.39 \)). Paired t test further indicated that \( A' \) and \( d' \) in the delayed test were significantly lower than those in the immediate test (\( A': t_{88} = -9.72, p < 0.001 \), Cohen’s \( d = 1.03 \); \( d': t_{88} = -9.97, p < 0.001 \), Cohen’s \( d = 1.06 \)), suggesting that the general recognition performance decreased with time.

**Long-term emotional expression effects on face memory**

**Univariate results**

The ANOVA revealed a significant main effect of facial expression (\( F_{(4,85)} = 7.61, p < 0.001, \eta^2 = 0.26 \)) and retention interval (\( F_{(1,88)} = 27.35, p < 0.001, \eta^2 = 0.24 \)) on the hit rates, as well as a significant interaction effect (\( F_{(4,85)} = 4.17, p < 0.005, \eta^2 = 0.16 \), Fig. 3A). Post hoc tests indicated that facial expression modulated delayed memory recognition (hit rate: \( F_{(4,85)} = 10.18, p < 0.001, \eta^2 = 0.32 \), but not immediate memory recognition (hit rate: \( F_{(4,85)} = 2.17, p = 0.08 \)). Within the delayed recognition, post hoc test further suggested that hit rates for faces with both threatening facial expressions (angry or fearful) were significantly higher as compared to faces with non-threatening expressions (sad, happy or neutral, respectively) (angry vs. sad: \( t_{88} = 3.31, p < 0.01, \) angry vs. happy: \( t_{88} = 5.05, p < 0.001, \) angry vs. neutral: \( t_{88} = 4.84, p < 0.001, \) fearful vs. sad: \( t_{88} = 2.79, p < 0.05, \) fearful vs. happy: \( t_{88} = 5.43, p < 0.001, \) fearful vs. neutral: \( t_{88} = 4.97, p < 0.001, \) all Holm corrected). In contrast, no significant differences between recognition performance of angry versus fearful (\( t_{88} = 1.09, p = 0.558, \) Holm corrected) as well as neutral versus happy faces (\( t_{88} = 0.10, p = 0.923, \) Holm corrected) were observed, whereas sad faces were significantly better remembered than neutral faces (\( t_{88} = -2.63, p = 0.04, \) Holm corrected) and marginally significantly better remembered than happy faces.
To rule out the possibility that this long-term emotional memory advantage of threatening faces was influenced by variations in the retention interval (ranging from 653 to 1113 days), repeated-measures ANOVA on hit rate of the delayed test was recomputed including interval day as a covariate and results remained stable ($F(4,84) = 2.43, p = 0.05, \eta^2 = 0.10$).

Moreover, correlation analyses were conducted between retention interval day and hit rate in each condition and these correlations were not significant (all $p$s > 0.05, Supplementary Results). To test whether this long-term enhanced memory for threatening faces is a result of higher tendency to respond “old” to threatening faces, two control analyses were conducted. First, we compared the false alarm rates in delayed test with the chance level and the results showed that the false alarm rates for both angry (one-sample $t$ test: $t_{88} = 0.70, p = 0.49$) and fearful (one-sample $t$ test: $t_{88} = 1.61, p = 0.11$, for other expression conditions, see Supplementary Results) expression conditions were at chance level, whereas the respective hit rates were higher than chance level (angry: one-sample $t$ test, $t_{88} = 3.00, p < 0.005$; fearful: one-sample $t$ test, $t_{88} = 2.16, p < 0.05$; for other expression conditions, see Supplementary Results). Correspondingly, the comparisons between false alarm rates and chance level in immediate test are provided in Supplementary Results. Moreover, the false alarm rates between expression conditions within the delayed test were not significantly different after including interval as a covariate ($F(4,84) = 1.08, p = 0.37$; although significant difference without controlling for interval days, $F(4,85) = 9.85, p < 0.001, \eta^2 = 0.32$), whereas the false alarms rates during immediate recognition showed an emotion-specific pattern ($F(4,83) = 22.15, p < 0.001, \eta^2 = 0.51$; for post hoc test results, see Supplementary Results). Further control analysis was performed on arousal of the face expressions. Repeated-measures ANOVA showed that the arousal ratings of emotional faces (angry, fearful, happy, sad) were significantly higher relative to neutral faces, and angry, fearful and happy faces elicited greater arousal than sad faces ($F(4,85) = 42.01, p < 0.001, \eta^2 = 0.66$, post hoc test results see Supplementary Fig S1) in the current sample. Nevertheless, the hierarchical regression analysis suggested that the interaction between facial expression and arousal did not reveal a significant effect ($\Delta R^2 = 0.068, p = 0.310$) on the hit rate. Together, these results suggested that the long-term recognition memory advantage for threatening faces was not influenced by interval days, false alarm rate, and arousal.

To determine changes of memory in each facial expression condition over > 1.5 years, pairwise comparisons on hit rates between the immediate and delay test for each facial expression condition were conducted. Recognition performance for happy, neutral and sad faces (two-tailed paired $t$ test: happy: $t_{88} = -5.25, p < 0.001$, Cohen’s $d = 0.56$; neutral: $t_{88} = -6.10, p < 0.001$, Cohen’s $d = 0.65$; sad: $t_{88} = -3.21, p < 0.01$, Cohen’s $d = 0.34$, Fig. 3A) significantly declined...
during the 1.5-year retention interval, whereas recognition for angry ($t_{88} = −1.65, p = 0.104$) and fearful ($t_{88} = −2.04, p = 0.09$) faces remained unchanged after Holm correction.

Together, the results indicated a long-term face recognition advantage of threatening expressions (i.e., angry and fearful) and this advantage was driven by decreased recognition of faces with non-threatening expressions including happy, sad and neutral following a retention interval of 1.5 years.

**PCA results**

Initial inspection of the response patterns (color-coded trial-wise response with blue to red as the confidence increased, see Fig. 2A) revealed strong individual variations in confidence ratings for expression-specific face images in the delayed test. PCA was then applied to map the confidence ratings of all 50 target face images in the 2D geometrical space of PC1 and PC2. As expected, PC1 had a discriminative variability that accounted for facial expression conditions with polarity: angry and fearful faces (generally negative scores) versus happy, neutral, sad faces (generally positive scores) (Fig. 3B). In particular, considering the localization of each face trial along the PC1 axis as visual reference, angry and fearful were separated from neutral and happy, while sad was located in between. PC1 explained 12.40% of the variance. The segregation pattern emerging from the amount of variance explained by PC1 was significant compared with the null distribution (trustworthiness/$p < 0.001$). In line with the visual presentation, pairwise non-parametric Wilcoxon Signed-Rank tests showed that the confidence ratings were not significantly different between angry vs fearful faces (Wilcoxon sign-rank $z = −1.33, p = 0.18$) and neutral vs happy faces (Wilcoxon sign-rank $z = −0.11, p = 0.91$). On the other hand, there were significant differences between angry vs. happy (Wilcoxon sign-rank $z = −5.25, p < 0.001$), angry vs. neutral (Wilcoxon sign-rank $z = −5.05, p < 0.001$) angry vs. sad (Wilcoxon sign-rank $z = −3.82, p < 0.001$), fearful vs. happy (Wilcoxon sign-rank $z = −5.20, p < 0.001$), fearful vs. neutral (Wilcoxon sign-rank $z = −5.32, p < 0.001$), fearful vs. sad (Wilcoxon sign-rank $z = −2.75, p = 0.006$), happy vs. sad (Wilcoxon sign-rank $z = −2.11, p = 0.035$) and neutral vs. sad (Wilcoxon sign-rank $z = −2.55, p = 0.011$). The results remained stable after Benjamini–Hochberg adjustment for multiple comparison testing. The same PCA procedure was also applied to the immediate test data and failed to detect a discriminative memory pattern along the PC1 axis (see Supplementary Fig S2). The lack of separation according to facial expression in the PCA was in line with findings from the conventional univariate non-parametric approach (Wilcoxon Signed-Rank tests) which did not reveal significant interaction effects between emotion and recognition performance for the immediate test (Supplementary Results).

To summarize, our univariate analyses suggested a long-term memory advantage of threatening (i.e., angry and fearful) versus non-threatening (i.e., particularly happy and neutral faces and to a lesser extent sad faces), and this emotional expression effect was further supported by the results from the data-driven multivariate approach.

**Neural basis of individual differences during encoding**

Given that the PC1 scores derived from the previous PCA revealed a group-level emotional expression dimension with polarity and there existed apparent individual differences in the emotional recognition representation pattern, we next conducted the BPSA by correlating the individual confidence rating pattern with the PC1 score pattern (i.e., template) in which a higher similarity score represents stronger memory difference between threatening versus non-threatening faces. We then used the similarity score as regressor to illustrate the neural basis at encoding that underlies the individual differences in the emotional face memory representation. To do this, the first-level contrast of interest was modeled as threatening (angry + fearful) vs non-threatening (happy + neutral) expressions in accordance with the data-driven behavioral results (sad condition was excluded from the contrast because of the ambiguous behavioral results and to match the trial numbers for the contrast). The second-level voxel-wise whole-brain analyses controlling for interval days revealed significant negative associations between memory performance and neural activity in clusters in the left and right inferior occipital gyrus extending to occipital pole (left IOG: $k = 140$, peak MNI coordinates: $−18, −94, −16, t = −4.65$, two-tailed $P_{\text{cluster-FWE}} < 0.05$; right IOG: $k = 280$, peak MNI coordinates: $30, −82, 2, t = −4.33$, two-tailed $P_{\text{cluster-FWE}} < 0.05$; all whole-brain corrected, Fig. 4A) and left and right vmPFC/OFC (left vmPFC/OFC: $k = 114$, peak MNI coordinates: $−39, 23, −22, t = −4.60$, two-tailed $P_{\text{cluster-FWE}} < 0.05$; right vmPFC/OFC: $k = 148$, peak MNI coordinates: $39, 26, −19, t = −4.97$, two-tailed $P_{\text{cluster-FWE}} < 0.05$; all whole-brain corrected, Fig. 4A). Because there were no significant differences in the brain–behavior correlations between the left and right IOG (Steiger’s $Z = −0.11, p = 0.916$) and between the left and right vmPFC/OFC (Steiger’s $Z = 0.88, p = 0.379$), neural responses in the left and right IOG and vmPFC/OFC were collapsed across hemispheres and used for illustration of the correlations (for scatter plots, see Fig. 4B). ROI analysis further indicated that neural responses to threatening vs non-threatening faces in the left fusiform face area (FFA) were negatively correlated with the similarity scores ($k = 59$, peak MNI coordinates: $−39, −55, −25, t = −4.17$,
two-tailed $p_{\text{voxel-FWE}} < 0.01$, SVC; Fig. 4C, D). Against our expectations, the examination of the bilateral amygdala with SVC did not reveal a significant brain–behavior association. These results suggested that differential reactivity to threatening vs non-threatening faces in bilateral IOG, bilateral vmPFC/OFC and left FFA during encoding might contribute to the individual differences in differential recognition threatening vs. non-threatening faces.

**Exploratory analysis: comparison of discriminators and non-discriminators**

To further explore the pattern of effects, an exploratory analysis aimed at determining the neural activation differences between ‘discriminators’ and ‘non-discriminators’ was conducted based on their behavioral pattern and brain activity. The BPSA approach successfully separated the subjects into discriminators ($N=43$, 21 males, $r_s > 0.28$, $p_s < 0.05$) and non-discriminators ($N=46$, 23 males, $r_s < 0.28$, $p_s > 0.05$), and a series of additional analyses further confirmed the separable emotional face representations between the two subgroups (details see Supplementary Results). To explore neural differences between the groups, a two-sample t test between the discriminators and non-discriminators was performed on the contrast (angry/fearful vs happy/neutral), revealing significant interaction effects in the left IOG ($k = 127$, peak MNI coordinates: $-18$, $-82$, $-19$; $t = -4.65$; two-tailed $p_{\text{cluster-FWE}} < 0.05$, whole-brain corrected) and right vmPFC/OFC ($k = 131$, peak MNI coordinates: $27$, $32$, $-16$; $t = -3.90$; two-tailed $p_{\text{cluster-FWE}} < 0.05$, whole-brain corrected; Fig. 5A), consistent with the brain–behavior correlation results with the overlap coefficients (defined as the size of intersection divided by the smaller of the two clusters, Saygin et al. 2016) of 0.60 and 0.53 in the left IOG and right vmPFC/OFC, respectively. Subsequent extraction of parameter estimates from independent masks from Brainnetome atlas suggested a distinct pattern of neural responses in the left IOG and right vmPFC/OFC, such that in the left IOG, the non-discriminators showed higher reactivity towards threatening faces compared to non-threatening faces.
while there was no difference in discriminators ($t_{42} = -0.49$, $p = 0.63$, Fig. 5B); whereas in the right vmPFC/OFC, the discriminators responded less to threatening faces compared to non-threatening faces ($t_{42} = -3.66$, $p < 0.01$, Holm corrected, Cohen’s $d = 0.56$), whereas the non-discriminators showed no difference ($t_{45} = 1.72$, $p = 0.09$, Fig. 5C, for further details, see also Supplementary Results and Figs S3 and S4).

Taken together, the findings from the two convergent analytic approaches suggest that the individual differences in long-term threatening vs. non-threatening recognition might be preceded by different encoding-related neural activity towards threatening vs non-threatening faces in the medial prefrontal and occipito-temporal cortex.

### Discussion

The present study systematically examined the impact of emotional facial expressions during initial encounters on face recognition following a retention interval of > 1.5 years. In line with our hypothesis, we found evidence that individuals recognized threatening faces better, particularly angry and fearful ones, as compared to non-threatening faces following the long-term retention interval. Moreover, the emotional expression-specific recognition advantage was not present directly after encoding and the long-term advantage of threatening faces was driven by decreased recognition of non-threatening faces over the retention period. Multivariate analyses further supported this finding by showing a separation according to emotional face expressions following the long-term retention interval but not during immediate.
recognition. Interestingly, the expression-specific face recognition pattern exhibited considerable inter-subject variation and a data-driven BPSA demonstrated that approximately half of the subjects demonstrated discriminative emotional face representation (discriminators) on the behavioral level, while the other half did not (non-discriminators). Examination of neural activation differences during encoding revealed that the different activation patterns in response to threatening vs non-threatening faces in bilateral IOG extending to occipital pole and vmPFC/mOFC on the whole-brain level, and left FFA on the ROI level was associated with the individual differences in the threatening vs non-threatening face recognition pattern after a delay of > 1.5 years. Two-sample t test on the subgroups of subjects separated by significance of similarity scores further confirmed differential neural activity in the medial prefrontal and occipito-temporal cortex during encoding. Together, the present findings demonstrated that threatening expressions during incidental encounters might facilitate long-term face recognition and that differential encoding in the occipito-temporal cortex and vmPFC/mOFC might contribute to expression-associated recognition pattern differences.

In contrast to previous studies using relatively short retention intervals ranging from minutes to weeks with univariate approach (e.g., Anderson et al. 2006; Wang 2013; Xiu et al. 2015), our study for the first time examined enhanced emotional expression effect for face recognition after an extensive delay of at least 1.5 years with both univariate and multivariate approaches. The enhanced/discriminative long-term memory for faces with threatening expressions (i.e., angry/fearful) relative to faces with non-threatening expressions (i.e., happy/neutral) in the present study was consistent with previous findings showing that the recognition and recollection for negative (e.g., fearful) faces was better than that for neutral faces after a delay of 24 h (Wang 2013). No differences between facial expression conditions were observed in the immediate test, which is consistent with some previous studies showing a lack of emotional facial expression modulation during immediate memory (Anderson et al. 2006; Grady et al. 2007; Satterthwaite et al. 2009; Xiu et al. 2015). The time-dependent effects of emotional expression on recognition observed in our study (i.e., enhanced emotional memory after a long-term delay but not immediately after encoding) were in line with prior studies using non-face emotional stimuli such as words or scenes (Sharot and Phelps 2004; Sharot and Yonelinas 2008), which indicated that the recognition for emotional materials was comparable to neutral ones immediately after encoding and that the recognition of neutral stimuli decreased over time, while recognition of negative stimuli remained the same. More importantly, our study extends previous findings by demonstrating that the year-long emotional enhancement effect of memory was not limited to emotional scenes (Dolcos et al. 2005; Erk et al. 2010; Gavazzeni et al. 2012) and suggests that threatening expressions slow forgetting of faces via mechanisms that enhance long-term consolidation (Cahill and McGaugh 1998; Talmi 2013; Yonelinas and Ritchey 2015). From an evolutionary perspective, maintaining recognition of threatening faces over long intervals may represent an adaptive and survival-relevant mechanism (Staugaard 2010), whereas faces with non-threatening expressions are of lower significance for future encounters, and thus were prone to be forgotten over time (Dunsmoor et al. 2015). Future studies could further determine the exact interval after which the memory enhancing effect of threatening expressions would emerge through consolidation by including repeated assessments in separate cohorts.

Some previous studies reported better memory for negative faces during immediate face recognition (Grady et al. 2007; Pinabiaux et al. 2013; Wang 2013). Methodological differences between experiments may account for this discrepancy. For example, the specific emotional expression tested and the number of face stimuli per emotional category varies between studies (Anderson et al. 2006; Xiu et al. 2015; but see Grady et al. 2007; Wang 2013). Besides, previous studies only tested a small number of subjects (N < 50) which may result in a lack of statistical power. Moreover, the mixed findings from previous studies might also attribute to the heterogeneity of sample characteristics such as age and gender which has been reported to be related to general face recognition memory (e.g., Grady et al. 1995; Sommer et al. 2013). Additionally, different analytical approaches may have contributed to the divergent results such that the previous studies generally employed univariate analytic approaches which might cause a loss of information in response variability and individual differences (Miendlarzewska et al. 2018; Tian et al. 2020). Indeed, the individual behavioral response pattern (Fig. 2A) indicated individual differences in emotional expression representation, consistent with individual variations that have been reported in face recognition (Tardif et al. 2019; Wang et al. 2012), emotional expression processing (Calder and Young 2005; Le Grand et al. 2006) and threat-related information processing (Liu et al. 2021a). Our study partly overcame some of the previous limitations by incorporating data-driven multivariate analyses in examining emotional expression effects in a large sample of young adults with balanced gender.

Additional control analyses suggest that the long-term memory-enhancing effect of threatening facial expressions could not completely be explained by factors such as an expression-specific tendency to respond items as previously seen or arousal at encoding. While an expression-specific variation in the false alarm rates was observed at both, immediate and delayed recall, false alarm rates were similar across the emotion expressions when variations in the retention interval were controlled for and false alarm rates
for threatening faces were at chance level in the delayed recognition test. Previous studies investigating emotional memory advantage emphasized the role of arousal (Bradley et al. 1992; Hamann 2001; LaBar and Phelps 1998; Sharot and Phelps 2004). Indeed, arousal may bias processing toward salient information that gains processing priority, and thus contribute to enhanced consolidation (Mickley Steinmetz et al. 2012; Ritchey et al. 2008). However, the enhanced memory for threatening faces in our study is unlikely explained by arousal based on three lines of evidence. First, a moderation analysis confirmed the lack of arousal effects by showing a non-significant interaction of facial expression category by arousal on memory performance. Second, arousal ratings for each expression at encoding between discriminators and non-discriminators were not significantly different (see Supplementary Results). Third, on the neural level, we did not observe significant activity in brain regions that have been shown to be engaged in processing emotional arousal such as the amygdala (Adolphs et al. 1999) or in unspecific salience processing such as the insula (Menon and Uddin 2010), which together suggests that the findings in the present study are unlikely to reflect unspecific arousal or salience effects of emotional stimuli.

The individual differences in emotional face recognition pattern not only substantiated the multivariate PCA analysis but also allowed us to capture the corresponding neural correlates by employing an individual difference approach. With a multivariate measure capturing individual differences based on BPSA, we observed that subjects with higher recognition differences between threatening versus non-threatening faces exhibited lower activity in response to threatening (i.e., angry/fearful) faces versus non-threatening (i.e., happy/neutral) faces in the bilateral IOG extending to occipital pole and vmPFC/OFC. ROI analysis further suggested that the left FFA was also engaged in the long-term emotional expression effects on memory. These identified regions align with the previously reported face network and resemble previous findings on emotional face recognition and emotion-related learning (Kumfor et al. 2013; LaBar and CabeZa 2006; Rolls 2019; Sergerie et al. 2005; Xiu et al. 2015). IOG and FFA are two key regions in the core face network (CFN) which usually support face identity recognition (Haxby et al. 2000; Liu et al. 2002; Tsantani et al. 2021), yet were also involved in processing emotional facial expression (Cohen Kadosh et al. 2010; Harry et al. 2013), whereas occipital pole is modulated by salient visual stimuli and has been involved in processing emotional faces in schizophrenia (Taylor et al. 2012). OFC is a part of extended face network (EFN) which processes social-emotional aspects of face processing, specifically value-based emotion and reward information (Ishai 2008; O’Doherty et al. 2003). Lesion studies have demonstrated that IOG and OFC lesion impairs emotional memory and learning (Calder and Young 2005; Kumfor et al. 2013; Rossion et al. 2003). Damage to inferior fronto-occipital fasciculus (IFOF), which connects the occipital cortex to OFC, has, moreover, been reported to predict facial emotion recognition impairment including anger, fear and sadness (Philippi et al. 2009). Our findings suggest a behavioral significance of encoding-related activity in these regions in predicting long-term emotional face memory patterns. Previous studies measuring the correlations between neural activation during fMRI and subsequent face (neutral expression) recognition performance employing independent task paradigms and stimuli have shown that higher face selectivity in FFA and IOG was associated with better face recognition ability in both normal subjects (Huang et al. 2014) and developmental prosopagnosia (Liu et al. 2021b), suggesting that the face-selective response in these regions contributes to individual differences in face recognition. However, these studies did not permit an investigation of encoding-related activity in predicting subsequent memory. One study that used an individual difference approach to examine encoding-related neural correlates of emotional face memory reported that the connectivity from IOG to OFC is negatively correlated with immediate subsequent memory performance of faces with different emotional expressions (i.e., negative, neutral, positive, respectively) (Xiu et al. 2015), suggesting an emotion-specific modulation of the face network which aligns with our findings.

Interestingly, the present results indicated a negative correlation between activity during encoding and long-term emotional face representation pattern, with higher activity for angry/fearful vs happy/neutral activity being associated with less discriminative recognition performance between faces with different emotional expressions (i.e., more similar performance for threatening vs non-threatening faces). Group comparison based on the similarity (discriminators vs. non-discriminators) further revealed distinct mechanisms in IOG and vmPFC/OFC, that is, the threatening vs non-threatening activation difference in IOG was driven by enhanced neural response to threatening faces in non-discriminators, whereas the threatening vs non-threatening activation difference in vmPFC/OFC resulted from suppressed neural response to threatening faces in discriminators. The differential mechanisms may correspond to the different function of these regions. That is, the preferential early visual processing of threatening faces in visual cortex did not lead to better recognition, while lower implicit threat evaluation and regulation at encoding in prefrontal regions engaged in emotional processing (i.e., vmPFC/OFC) might link to the enhanced recognition of threatening faces. The distinct role of IOG and vmPFC/OFC in predicting long-term emotional face memory remains to be finely characterized in future studies.

The lack of amygdala effect was consistent with the behavioral results suggesting that there was no modulation
of arousal on memory performance and might be partly explained by previous observations that the amygdala responds equally to positive, negative and neutral faces during encoding (Adolphs 2010; Ball et al. 2009). Although previous studies have indicated that amygdala activation at encoding predicted greater subsequent memory for emotional compared to neutral scenes after a 1-year delay (Dolcos et al. 2005), the findings are based on the neural activity of remembered versus forgotten items for emotional versus neutral scene pictures. Likewise, the present study did not reveal a role of the insula in emotional face memory, suggesting that the emotional face memory was not due a higher externally guided attention towards emotional faces during encoding (Yao et al. 2018).

The present study used multivariate PCA and BPSA for long-term emotional face recognition performance in a large sample, which differentiates our study from most previous ones comparing the experimental conditions by averaging across all stimulus items. The univariate method relies on the assumption that all stimulus items within each condition are homogeneous, and the performance differences between items reflect noise fluctuations. The data-driven multivariate PCA allowed us to detect hidden pattern in the confidence ratings in a hypothesis-free manner by capitalizing on the trial-wise responses, which not only complemented the traditional univariate analysis but also informed fMRI analysis using an individual difference approach. Although the first component retained only 12.4% of the original variance which might be due to the noise associated with the long retention interval, the PCA did generate separate clusters (see Fig. 3B) and the PC1 separation was statistically significant according to a non-parametric trustworthiness test, suggesting that this approach successfully identified distinguishable latent groups of facial expressions based on the recognition confidence ratings. More importantly, the extracted scores of PC1, representing a discriminative variability that accounted for the facial expression categories, promoted the BPSA by correlating the PC1 scores with each subject’s confidence rating pattern. This multivariate method allowed us for the first time to capture individual differences in emotional face representation patterns by relating the face representation (i.e., multi-item confidence rating pattern) of each subject to a group-level emotion-specific discriminative memory pattern (indexed by PC1 scores). Importantly, the present BPSA took the principal component score pattern as a response template, whereas previous studies usually used mean score pattern as a template (e.g., Tian et al. 2020).

Thus, our method could provide a novel way to distinguish the memory performance for faces with different emotional expression conditions and has the potential to support a broader examination of dimensional data (e.g., behavioral, gene or fMRI data) comprising of complex conditions. Moreover, by calculating the association between individual feature pattern and the pattern template generated by unsupervised dimensional reduction (e.g., PCA, minimum curvilinear embedding, Miendlarzewska et al. 2018), one can characterize the individual differences in the representation pattern and associated it with the neural systems.

Several limitations need to be addressed. First, the comparably low number of hits for each expression condition after a long retention interval did not permit a reliable fMRI analysis of subsequent memory effects comparing remembered versus non-remembered items as employed in previous studies (e.g., Becker et al. 2017; Dolcos et al. 2005). Future studies that aim to determine facial expression effects over very long retention intervals may enhance the processing depth of the stimuli or increase the number of stimuli per expression condition as well as the number of subjects (Steele, et al. 2016) are invited to further explore the neural basis underlying long-term emotional face memory. Moreover, to examine the long-term memory effect with higher sensitivity, a dimensional confidence rating was used at the long-term retention interval while the immediate recognition test initially implemented a forced choice response. Although previous studies used a similar approach and converted dimensional confidence ratings into binary responses (Weymar et al. 2011; Xiu et al. 2015), the results of direct comparison between immediate and delayed recognition should be interpreted with caution. Furthermore, previous studies investigating the immediate recognition of neutral faces that were previously presented with threatening (angry or fearful) or non-threatening (happy or sad) expressions and found no differences in recognizing neutral faces with threatening versus non-threatening expressions (Satterthwaite et al. 2009), suggesting that the “first impression” created by facial expression did not affect subsequent recognition of face identity. However, whether such a “first impression” has an effect on long-term face identity recognition remains to be addressed in future studies. Despite comparable recognitions for angry and fearful faces, the grouping of angry and fearful faces into threatening faces may obscure encoding difference between angry and fearful faces. Future studies are needed to investigate the neural differences within categories of threatening and non-threatening faces for subsequent recognition. Notably, on the behavioral level, no differences were observed during immediate recognition, suggesting that differential encoding-related activation in these regions may be associated with individual variation during memory consolidation. Further studies are necessary to explore neural basis during consolidation to provide a more comprehensive understanding of the formation of long-term emotional face memory. Moreover, given the exploratory nature of the fMRI analysis in the present study, we focused on univariate fMRI analyses to identify specific brain regions associated with individual differences. Combining suitable fMRI and sample designs with multivariate fMRI analyses (e.g., Zhou et al. 2020, 2021) may offer additional insights into the neurofunctional representation of...
long-term face memory in future studies. Finally, in line with our research goal, the neural basis of discriminative face recognition pattern was determined on the level of emotion categories. Future studies may aim to further disentangle neural systems of item-specific face memory within each emotional expression category or across emotions to explore different encoding patterns for face stimuli that are on average highly or poorly memorable.

Conclusions

Our study provides evidence for a recognition advantage of threatening faces after a long-term interval of > 1.5 years. fMRI analyses further suggested that individual differences in the emotional face memory pattern were associated with differential encoding of threatening versus non-threatening faces in the IOG/occipital pole, FFA and vmPFC/mOFC, suggesting that encoding-related activity in the occipito-temporal cortex and medial prefrontal cortex may play a role in the formation of long-term emotional face memory. These findings extend the theory of long-term emotional memory towards facial stimuli and shed new light on the encoding-related neural basis of preserved memory for faces with threatening expressions.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s00429-022-02462-5.

Acknowledgements  We thank all subjects who participated in this study.

Funding  This work was supported by the National Key Research and Development Program of China (Grant No. 2018YFA0701400), National Natural Science Foundation of China (NSFC, No. 91632117, 31700998, 31530032); Fundamental Research Funds for Central Universities (ZYGX2015Z02), Science, Innovation and Technology Department of the Sichuan Province (2018JY0001).

Data availability  Behavioral data are available from the corresponding author upon reasonable request. The code for the multivariate data analysis PC-corr MATLAB code is available via https://github.com/biomedical-cybernetics/PC-corr_net. Statistical images of main analyses can be found at https://neurovault.org/collections/9071/.

Declarations

Conflict of interest  The authors declare no competing financial interests.

References

Adolphs R (2010) What does the amygdala contribute to social cognition? Ann N Y Acad Sci 1191(1):42–61. https://doi.org/10.1111/j.1749-6632.2010.05445.x

Adolphs R, Russell JA, Tranel D (1999) A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. Psychol Sci 10(2):167–171. https://doi.org/10.1111/1467-9280.00126

Aly M, Turk-Browne NB (2016) Attention promotes episodic encoding by stabilizing hippocampal representations. Proc Natl Acad Sci 113(4):E420–E429. https://doi.org/10.1073/pnas.1518931113

Anderson AK, Yamaguchi Y, Grabski W, Lacka D (2006) Emotional memories are not all created equal: evidence for selective memory enhancement. Learn Mem 13(6):711–718. https://doi.org/10.1101/lm.388906

Ball T, Derix J, Wentlandt J, Wiekhorst B, Speck O, Schulze-Bonhage A, Mutschler I (2009) Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. J Neurosci Methods 180(1):57–70. https://doi.org/10.1016/j.jneumeth.2009.02.022

Becker B, Steffens M, Zhao Z, Kendrick KM, Neumann C, Weber B, Hurlemann R et al (2017) General and emotion-specific neural effects of ketamine during emotional memory formation. Neuroimage 150:308–317. https://doi.org/10.1016/j.neuroimage.2017.02.049

Bowen BJ, Kark SM, Kensing EA (2018) NEVER forget: negative emotional valence enhances recapitulation. Psychon Bull Rev 25(3):870–891. https://doi.org/10.3758/s13420-017-1313-9

Bradley MM, Greenwald MK, Petry MC, Lang PJ (1992) Remembering pictures: pleasure and arousal in memory. J Exp Psychol Learn Mem Cogn 18(2):379. https://doi.org/10.1037/0728-7393.18.2.379

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77(3):305–327. https://doi.org/10.1111/j.2404-8929.1986.tb02199.x

Cahill L, McGaugh JL (1998) Mechanisms of emotional arousal and lasting declarative memory. Trends Neurosci 21(7):294–299. https://doi.org/10.1016/S0166-2236(97)00289-2

Calder AJ, Young AW (2005) Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci 6(8):641–651. https://doi.org/10.1038/nrn1724

Chen LF, Yen YS (2007) Taiwanese facial expression image database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei

Ciucci S, Ge Y, Durán C, Palladini A, Jiménez-Jiménez V, Martínez-Sánchez LM, Cannistraci CV et al (2017) Enlightening discriminative network functional modules behind principal component analysis separation in differential-omic science studies. Sci Rep 7(1):1–24. https://doi.org/10.1038/srep43946

Cohen AL, Soussand L, Corrow SL, Martinaud O, Barton JJS, Fox MD (2019) Looking beyond the face area: lesion network mapping of prosopagnosia. Brain 142(12):3975–3990. https://doi.org/10.1093/brain/awz332

Cohen Kadosh R, Soskic S, Iuculano T, Kanai R, Walsh V (2010) Modulating neuronal activity produces specific and long-lasting changes in numerical competence. Curr Biol 20(22):2016–2020. https://doi.org/10.1016/j.cub.2010.10.007

Darwin C (1872) The expression of the emotions in man and animals. Murray, London

Dolcos F, LaBar KS, Cabeza R (2004) Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. Neuron 42(5):855–863. https://doi.org/10.1016/S0896-6273(04)00289-2

Dolcos F, LaBar KS, Cabeza R (2005) Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. Proe Natl Acad Sci USA 102(7):2626–2631. https://doi.org/10.1073/pnas.0409848102

Dunsmoor JE, Murty VP, Davachi L, Phelps EA (2015) Emotional learning selectively and retroactively strengthens memories for related events. Nature 520(7547):345–348. https://doi.org/10.1038/nature14106

Durán C, Ciucci S, Palladini A, Ijaz UZ, Zippo AG, Sterbini FP, Cannistraci CV et al (2021) Nonlinear machine learning pattern recognition and bacteria-metabolite multilayer network analysis of...
Damage to association fiber tracts impairs recognition of the facial expression of emotion. J Neurosci 29(48):15089–15099. https://doi.org/10.1523/JNEUROSCI.0796-09.2009

Pinhaiaux C, Hertz-Pannier L, Chiron C, Rodrigo S, Jambuâque I, Noulhiane M (2013) Memory for fearful faces across development: specialisation of amygdala nuclei and medial temporal lobe structures. Front Hum Neurosci 7:901. https://doi.org/10.3389/fnhum.2013.00901

Powell LJ, Kosakowski HL, Saxe R (2018) Social origins of cortical face areas. Trends Cogn Sci 22(9):752–763. https://doi.org/10.1016/j.tics.2018.06.009

Ringnér M (2008) What is principal component analysis? Nat Biotechnol 26(3):303–304. https://doi.org/10.1038/nbt0308-303

Ritchey M, Dolcos F, Cabeza R (2008) Role of amygdala connectivity in the persistence of emotional memories over time: an event-related fMRI investigation. Cereb Cortex 18(11):2494–2504. https://doi.org/10.1093/cercor/bhm262

Rolls ET (2019) The orbitofrontal cortex and emotion in health and disease, including depression. Neuropsychology 128:14–43. https://doi.org/10.1016/j.neuropsychologia.2017.09.021

Rossion B, Caldara R, Seghier M, Schuller AM, Lazeyras F, Mayer E (2003) A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain 126(11):2381–2395. https://doi.org/10.1093/brain/awg241

Russell R, Duchaine B, Nakayama K (2009) Super-recognizers: people with extraordinary face recognition ability. Psychon Bull Rev 16(2):252–257. https://doi.org/10.3758/PBR.16.2.252

Satterthwaite TD, Wolf DH, Gur RC, Ruparel K, Valdez JN, Gur RE, Loughead J (2009) Frontolimbic responses to emotional face memory: the neural correlates of first impressions. Hum Brain Mapp 30(11):3748–3758. https://doi.org/10.1002/hbm.20803

Saygin ZM, Osher DE, Norton ES, Youssoufian DA, Beach SD, Tardif J, Morin Duchesne X, Cohan S, Royer J, Blais C, Fiset D, Gosselin F et al (2019) Use of face information varies systematically from developmental prosopagnosia to super-recognizers. Psychol Sci 30(2):300–308. https://doi.org/10.1177/0956797618111338

Taylor SF, Kang J, Brege IS, Tso IF, Hosanagar A, Johnson TD (2012) Meta-analysis of functional neuroimaging studies of emotion perception and experience in schizophrenia. Biol Psychiatry 71(2):136–145. https://doi.org/10.1016/j.biopsych.2011.09.007

Thomas PMJ, Jackson MC, Raymond JE (2014) A threatening face in the crowd: effects of emotional singletons on visual working memory. J Exp Psychol Hum Percept Perform 40(1):253–263. https://doi.org/10.1037/a0033970

Tian X, Wang R, Zhao Y, Zhen Z, Song Y, Liu J (2020) Multi-item discriminability pattern to faces in developmental prosopagnosia reveals distinct mechanisms of face processing. Cereb Cortex 30(5):2986–2996. https://doi.org/10.1093/cercor/bhz289

Tsantani M, Kriegeskorte N, Storrs K, Williams AL, McGgettigan C, Garrido L (2021) FFA and OFA encode distinct types of face identity information. J Neurosci 41(9):1952–1969. https://doi.org/10.1523/JNEUROSCI.1449-20.2020

Vuilleumier P (2002) Facial expression and selective attention. Curr Opin Psychiatry 15(3):291–300. https://doi.org/10.1097/00001504-200205000-00011

Wang B (2013) Facial expression influences recognition memory for faces: robust enhancement effect of fearful expression. Memory 21(3):301–314. https://doi.org/10.1080/09658211.2012.725740

Wang R, Li J, Fang H, Tian M, Liu J (2012) Individual differences in holistic processing predict face recognition ability. Psychol Sci 23(2):169–177. https://doi.org/10.1177/0956797611420575

Weymar M, Löw A, Hamm AO (2011) Emotional memories are resilient to time: evidence from the parietal ERP old/new effect. Hum Brain Mapp 32(4):632–640. https://doi.org/10.1002/hbm.21051

Xiu D, Geiger MJ, Kiaver P (2015) Emotional face expression modulates visual working memory maintenance for fearful versus neutral facial identities. Psychophysiology 52(2):187–197

Xu L, Bolh T, Nomi IS, Li J, Zheng X, Fu M, Uddin LQ et al (2020) Inter-subject phase synchronization differentiates neural networks underlying physical pain empathy. Soc Cogn Affect Neurosci 15(2):225–233. https://doi.org/10.1093/scan/nzaa025

Slotnick SD (2017) Cluster success: fMRI inference for spatial extent have acceptable false-positive rates. Cogn Neurosci 8(3):150–155. https://doi.org/10.1080/17588928.2017.1319350
Yonelinas AP, Ritchey M (2015) The slow forgetting of emotional episodic memories: an emotional binding account. Trends Cogn Sci 19(5):259–267. https://doi.org/10.1016/j.tics.2015.02.009

Zhen Z, Yang Z, Huang L, Kong XZ, Wang X, Dang X, Liu J et al (2015) Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. Neuroimage 113:13–25. https://doi.org/10.1016/j.neuroimage.2015.03.010

Zhou F, Li J, Zhao W, Xu L, Zheng X, Fu M, Becker B et al (2020) Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. Elife 9:1–27. https://doi.org/10.7554/ELIFE.56929

Zhou F, Zhao W, Qi Z, Geng Y, Yao S, Kendrick KM, Becker B et al (2021) A distributed fMRI-based signature for the subjective experience of fear. Nat Commun 12(1):1–16. https://doi.org/10.1038/s41467-021-26977-3

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.