What drives social in-group biases in face recognition memory? ERP evidence from the own-gender bias

Nicole Wolff,1,2 Kathleen Kemter,1 Stefan R. Schweinberger,1,2 and Holger Wiese1,2
1Friedrich Schiller University of Jena, Jena, Germany and 2DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Jena, Germany

It is well established that memory is more accurate for own-relative to other-race faces (own-race bias), which has been suggested to result from larger perceptual expertise for own-race faces. Previous studies also demonstrated better memory for own-relative to other-gender faces, which is less likely to result from differences in perceptual expertise, and rather may be related to social in-group vs out-group categorization. We examined neural correlates of the own-gender bias using event-related potentials (ERP). In a recognition memory experiment, both female and male participants remembered faces of their respective own gender more accurately compared with other-gender faces. ERPs during learning yielded significant differences between the subsequent memory effects (subsequently remembered – subsequently forgotten) for own-gender compared with other-gender faces in the occipito-temporal P2 and the central N200, whereas neither later subsequent memory effects nor ERP old/new effects at test reflected a neural correlate of the own-gender bias. We conclude that the own-gender bias is mainly related to study phase processes, which is in line with sociocognitive accounts.

Keywords: own-gender bias; ERP; face processing; Dm effect; old/new effect

INTRODUCTION

Although humans can recognize a massive number of previously encountered faces, this ability varies with the category a specific face belongs to. For example, participants are more accurate at remembering faces of their own relative to other ethnic (for a review see Meissner and Brigham, 2001) or age groups (Bartlett and Leslie, 1986; Bartlett and Fulton, 1991; for a recent meta-analysis see Rhodes and Anastasi, 2012). Both these effects, commonly referred to as own-race and own-age biases, respectively, have been explained by differences in contact and/or perceptual expertise with own-group and other-group persons. For instance, most Caucasian participants in European countries spend more time with people from their own ethnic group than with people from another ethnic group. Similarly, most young adults have substantially more contact to their own age group relative to elderly persons. Accordingly, it was suggested that the face representational system (the so-called ‘face space’), being shaped by experience, is normally tuned to differentiate optimally between individual faces of a viewer’s own age and ethnic group, but less well for other-group faces (Valentine, 1991, 1992; Furl et al., 2002; Byatt and Rhodes, 2004). Such differences in the representation of own- and other-group faces have been suggested to underlie biases in face recognition memory.

Similar to these well-established own-race and own-age biases, an own-gender bias in face recognition memory has also been described. However, previous results have not been entirely consistent. Although women are commonly found to be more accurate at remembering female compared with male faces (Lewin and Herlitz, 2002; Lovén et al., 2011), results for male participants appear more variable, either showing (i) a corresponding own-gender bias with better memory for male than female faces (Wright and Sladden, 2003), (ii) similar performance for male and female faces (Lewin and Herlitz, 2002) or (iii) even enhanced memory performance for female relative to male faces (McKelvie et al., 1993). Although an own-gender bias in women was suggested to result from enhanced perceptual expertise (Lewin and Herlitz, 2002; Rehnman and Herlitz, 2006; Lovén et al., 2011), empirical support for this suggestion is lacking, as no measure of contact or expertise was reported in those studies. In contrast, and irrespective of the earlier-described discrepancies, it could be argued that any own-gender bias in face memory is unlikely to result from different perceptual expertise or daily-life contact, to the extent that most people in western societies regularly interact with persons from both genders.

Alternatively, biases in face recognition memory have been explained by sociocognitive accounts, emphasizing the importance of categorizing a face as belonging to either a social ‘in-group’ or ‘out-group’ (Sporer, 2001; Hugenberg et al., 2010). Accordingly, face memory biases may occur not because of differences in expertise or contact, but because faces perceived as belonging to a social ‘out-group’ are processed at a categorical level, whereas ‘in-group’ faces are individualized. Such individualization is assumed to result in more in-depth processing, and thus more accurate recognition memory. Supporting these ideas, it was demonstrated that even ‘in-group’ faces with respect to race and age, when arbitrarily labeled as belonging to the participants’ university, were remembered more accurately as similar faces when labeled as being affiliated to a different university (Bernstein et al., 2007). Importantly, sociocognitive accounts assume that face memory biases reflect differential processing of in-group vs out-group faces during learning. Relevant empirical support comes from observations that both the own-race bias (ORB) and biases based on labeling same-race faces as belonging to a social out-group disappeared when participants were briefed about memory biases, and were instructed to pay close attention to out-group faces during learning (Hugenberg et al., 2007; Young et al., 2010).

In sum, although an own-gender bias is not easily integrated into contact- or expertise-based accounts, it can be explained by sociocognitive accounts, given that other-gender faces are perceived as belonging to a social ‘out-group’. It also remains unclear whether similar or different processes underlie the own-race, own-age, and own-gender biases. Thus, even when the own-gender bias may not be related to expertise, it remains possible that the ORB is. It is therefore important to identify and compare cognitive and neural processes underlying...
ERP correlates of social in-group biases

We aimed at filling this gap by examining ERP correlates of recognition memory for female and male faces in both female and male participants. Although this study did not test different biases in a combined experiment, a comparison of such ERP effects with previously observed neural correlates of the ORB may help to clarify the extent to which these biases rely on similar processes. Although combining memory biases within the same experiment facilitates direct comparison between biases (Wiese, 2012), biases likely interact and are therefore not measured in their ‘purest’ form in combined

these effects. Event-related potentials (ERP) provide an excellent method to examine this issue.

Although behavioral measures can only capture the end-product of a sequence of mental sub-processes initiated after stimulus presentation, ERPs provide detailed chronometric information about those sub-processes. The first ERP component usually examined in face perception experiments is the P1, which is maximal over occipital sites at approximately 80–130 ms after stimulus onset. P1 is elicited by any visual stimulus, and its amplitude and latency are known to be highly sensitive to physical characteristics, including spatial frequency, luminance and contrast (Luck, 2005).

More importantly, face stimuli elicit a large negative deflection maximal at right occipito-temporal channels after approximately 170 ms. This so-called N170 (Bötzel et al., 1995; Bentin et al., 1996; Eimer, 2011) was suggested to reflect the structural encoding of faces (Bentin et al., 1996) or the detection of a face-like stimulus in the visual field (Schweinberger and Burton, 2003). Although several studies did not find differences for own-race vs other-race faces in N170 (Caldara et al., 2003; Wiese et al., 2009), a number of recent studies reported larger N170 for other-race compared with own-race faces (Stahl et al., 2008, 2010; Walker et al., 2008; Cabarel et al., 2011; Wiese, 2012). This discrepancy may be partly explained by task demands: Senholzi and Ito (2013) demonstrated larger N170 for other-race faces only when participants were asked to individuate the faces. A number of the earlier-cited studies used recognition memory paradigms, in which participants were to memorize faces, which likely induced a motivation to process these on an individual level. Other studies examined effects of face ethnicity on the central P200 (also typically peaking ~170 ms), which likely reflects the positive counterpart of the N170, the vertex positive potential (Jeffreys, 1989), and observed larger VPP amplitudes for other-race faces (Ito and Urland, 2003; Ito and Bartholow, 2009). Moreover, N170 has been found to be larger for old compared with young faces (Wiese et al., 2008, 2012). In contrast, previous studies on processing facial gender did not find differences in N170 amplitude for male vs female faces (Mouchetant-Rostaing et al., 2000; Mouchetant-Rostaing and Giard, 2003). Interestingly, however, female participants have been reported to demonstrate smaller N170 amplitudes in gender categorization (female/male) compared with orientation tasks (face to the left/right), whereas male participants yielded no comparable pattern (Sun et al., 2010). Together, although preliminary evidence suggests some (task-dependent) differences between male and female observers, there is no evidence for overall differences in N170 depending on facial gender.

The occipito-temporal P2 represents a positive deflection approximately 200–280 ms. P2 was found to be decreased for vertically stretched vs normal faces (Halit et al., 2000) and was suggested to reflect the processing of metric distances between facial features (Latinus and Taylor, 2006; Mercure et al., 2008). Larger P2 responses have been observed for own-race vs other-race faces (Stahl et al., 2008, 2010), for young vs old faces (Wiese et al., 2008), and for veridical vs spatially caricatured faces (Kaufmann and Schweinberger, 2012). Accordingly, P2 may reflect the perceived typicality of a face.

The subsequent negative component, the N250, is consistently larger for repeated relative to nonrepeated faces in immediate repetition priming experiments (the so-called N250r effect [Schweinberger et al., 1995, 2002]). More recently, N250 has also been associated with face learning, because more negative N250 amplitudes for learned vs novel faces have been reported (Tanaka et al., 2006; Kaufmann et al., 2009). The N250 may reflect the access to perceptual representations of individual faces, and thus the processing of facial identity. In addition, generally more negative amplitudes in the N250 time range were reported for other-race compared with own-race faces (Stahl et al., 2010; Wiese, 2012), and effects of stimulus repetition and face category have been shown to add up in the N250 (Wiese et al., in press). Accordingly, effects of face repetition and face category are independent and may represent different processes. More negative amplitudes in this time range have also been observed for emotional faces (Schupp et al., 2007), which has been interpreted as ‘tagging’ of particularly salient faces for further processing. Similar to effects in the N170, face category effects in N250 may be related to task demands: Ebner et al. (2011) found more negative N250 for own-age faces in a gender categorization task, whereas studies using recognition memory paradigms found more negative N250 for other-age faces (Wiese, 2012). To the best of our knowledge, no previous study systematically examined effects of face gender on P2 or N250.

In addition, several ERP effects related to episodic memory have been identified. For instance, in the learning phases of memory experiments, words later remembered at test elicit more positive amplitudes than later forgotten words (Paller et al., 1987, 1988). This so-called Difference due to subsequent memory (or Dm) effect is typically broadly distributed and sustained roughly between 300 and 1000 ms (Herzmann and Curran, 2011). Importantly, Dm effects were also described for faces (Sommer et al., 1991, 1995), and were recently observed to be larger for own-race compared with other-race faces (Lucas et al., 2011). In that study, a differential Dm effect for own- and other-race faces was detected in the occipito-temporal P2 (and polarity-reversed in the central N200), with more positive amplitudes for subsequently forgotten relative to subsequently remembered own-race faces, and with the opposite pattern for other-race faces. Herzmann et al. (2011) examined Dm effects for own- and other-race faces, which were either subsequently recollected (i.e. elicited recollection of study phase detail at test) (Yonelinas, 2002), or subsequently familiar (i.e. recognized at test without accompanying study phase detail). At some variance with the earlier-discussed findings, subsequently recollected other-race faces elicited more positive amplitudes than subsequently familiar other-race faces in Caucasian participants. In addition, no significant differences between subsequently recollected vs subsequently familiar own-race faces were detected. This was interpreted as representing similarly deep encoding for subsequently recollected and subsequently familiar own-race faces, but more elaborate learning of subsequently recollected other-race faces.

Finally, retrieval effects from episodic memory are typically detected in so-called ERP old/new effects during test phases of recognition memory experiments (Rugg and Coles, 1995). Old/new effects consist of more positivity for correctly remembered items (hits) compared with correctly rejected new items (correct rejections [CR]). They typically occur between 400 and 800 ms, with a left parietal maximum for verbal material (Rugg and Curran, 2007). Previous studies on own-group biases reported larger centroparietal old/new effects for own-race relative to other-race faces (Stahl et al., 2010; Herzmann et al., 2011; Wiese, 2012), and for own-age vs other-age faces in young participants (Wiese et al., 2012, 2008). Moreover, Guillem and Mograss (2005) found the old/new effect for faces to be reduced in male compared with female participants. However, to the best of our knowledge no previous study directly investigated neural correlates of the own-group bias.

We aimed at filling this gap by examining ERP correlates of recognition memory for female and male faces in both female and male participants. Although this study did not test different biases in a combined experiment, a comparison of such ERP effects with previously observed neural correlates of the ORB may help to clarify the extent to which these biases rely on similar processes. Although combining memory biases within the same experiment facilitates direct comparison between biases (Wiese, 2012), biases likely interact and are therefore not measured in their ‘purest’ form in combined
experiments. Accordingly, testing these biases in separate experiments represents important additional information.

We reasoned that both the absence of an effect that is consistently observed for other biases, and the observation of a novel ERP correlate specific to the own-gender bias would point to partly different processes underlying the respective memory phenomena. We considered the following specific hypotheses with respect to the ERP components under investigation: (i) As other-race faces elicit a larger N170 in recognition memory studies (Stahl et al., 2010), and as this effect is correlated with the ORB in memory (Wiese et al., in press), a similar finding for other-gender faces would suggest that face category effects in N170 are not based on perceptual expertise—as the processing of own- and other-gender faces likely does not differ with respect to expertise-based processing. (ii) P2 amplitude is larger for own-race vs other-race faces, but participants with enhanced contact toward other-race faces do not show this effect (Stahl et al., 2008; Wiese et al., in press). Accordingly, assuming that face category effects in P2 are based on perceptual expertise, no differential P2 was predicted for own-gender vs other-gender faces. (iii) N250 has been observed to be more negative for other-race compared with own-race faces, and during test this difference was shown to correlate with the ORB in memory (Wiese et al., in press). If a corresponding effect were observed for the own-gender bias, it would likely not reflect perceptual expertise-based processing. Finally, with respect to memory-related ERP effects the following predictions were tested: Dm effects during learning were found to be larger, or to differ in scalp distribution, for own- and other-race faces (Lucas et al., 2011). Moreover, old/new effects at test were found to be larger for own-race compared with other-race faces (Stahl et al., 2010; Herzmann et al., 2011). The presence of similar effects in the own-gender bias would point to an interpretation in terms of social categorization.

METHODS

Participants

Participants were 28 female (18–33 years, 22.89 ± 3.65) and 28 male (18–32 years, 23.82 ± 3.36) undergraduate students from the University of Jena. All values are expressed as mean ± s.d. All were right-handed according to a modified version of the Edinburgh Handedness Inventory (Oldfield, 1971), reported normal or corrected-to-normal vision, and received an honorarium of 5.00 €/h or course credits. None reported neurological or psychiatric disorders, or to receive central-acting medication.

Stimuli

Stimuli were 120 unfamiliar female faces aged 18–40 years (24.09 ± 3.16) and 120 male faces aged 19–40 years (25.70 ± 4.74). All images depicted color portraits with frontal views and neutral expressions. Stimuli were taken from the CAL/PAL (Minear and Park, 2004), the FACES (Ebner et al., 2010), the FERET (Phillips et al. 1998), and the GUFD databases (Burton et al., 2010). Using Adobe Photoshop, faces were cut out (i.e. no clothing or background visible) and pasted to a uniform black background. Stimuli were cropped to a frame of 299 × 380 pixels (10.55 × 13.41 cm), resulting in a visual angle of 6.7° × 8.5° at a viewing distance of 90 cm and presented on a computer monitor using E-Prime.

Procedure

Main experiment

Participants were seated in a dimly lit, electrically shielded and noise-attenuated chamber (400-A-CT-Special, Industrial Acoustics, Niederkürtchen, Germany) with their heads in a chin rest. The experiment consisted of a practice block (16 trials) and 4 experimental blocks. Each block was divided into a study and a test phase. In individual trials, a fixation cross was first displayed for 500 ms, followed by a face for 5000 ms (study phases) or for 2000 ms (test phases). Each trial ended with a blank screen (500 ms).

During each study phase, 30 faces (50% female) were presented, and participants were instructed to memorize and categorize the faces by gender. Study and test phases were separated by a 20 s break. During the ensuing test phase, all 30 faces from the directly preceding study phase were presented in randomized order together with 30 new faces (again 50% female). Participants decided via button presses with their left or right index finger whether a face had been presented in the directly preceding study phase (‘learned’) or not (‘new’). Speed and accuracy were emphasized. If participants classified a face as ‘learned’, a remember/know rating with an additional ‘guess’ option (Gardiner and Richardson-Klavehn, 2000) was conducted. Between blocks, participants were allowed a self-paced period of rest. Key assignment and allocation of stimuli to learned or non-learned conditions were balanced across participants.

Responses during test phases were sorted into four conditions, for female and male faces separately: hits (correctly identified learned faces), misses (learned faces incorrectly classified as new), Cr (new faces correctly identified as new), and false alarms (FA, new faces incorrectly classified as learned). Measures of sensitivity (d′ = z[hits] − z[FA]) and response bias (C = −0.5 [z(hits) + z(FA)]) were calculated according to signal detection theory (Green and Swets, 1966). Additionally, a memory bias score, consisting of the difference between d′ for female and male faces divided by the sum of the two measures, was calculated. Finally, percentages of ‘remember’ and ‘know’ responses for hits relative to the number of learned items were analyzed for male and female faces.

Face rating

After the experiment, each participant rated those 120 faces presented during learning for distinctiveness (from 1 = not distinctive to 6 = very distinctive) and attractiveness (from 1 = not attractive to 6 = very attractive). Faces were presented in randomized order, and participants were instructed to rate them spontaneously.

Contact questionnaire

Finally, all participants completed a questionnaire, which asked for the amount of contact (in hours/week), number of contact persons (per week) and contact quality (from 1 = very superficial to 4 = very intense) for female and male people in daily-life situations (such as job/university, meeting friends/spare time activities, family, and domestic circumstances (Wiese, 2012). Total scores were calculated for each participant by summing up (hours/week, number of persons/week) or averaging (contact quality) self-report measures from the different situations separately for contact toward male and female persons.

Electroencephalography recording and analysis

Electroencephalography (EEG) was recorded using a 32-channel BioSemi Active II system (BioSemi, Amsterdam, Netherlands). Active sintered Ag/AgCl-electrodes were mounted in an elastic cap with recording sites at Fz, Cz, Pz, Iz, FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, F9, F10, FT9, FT10, TP9, TP10, P9, P10, PO9, PO10, 11 and 12. EEG was recorded continuously (512 Hz sampling rate, DC to 155 Hz). Note that BioSemi systems work with a ‘zero-Ref’ setup with ground and reference electrodes replaced by a so-called CMS/DRL circuit (cf. to http://www.biosemi.com/faq/cms&drl.htm for further information).
Blink artefacts were corrected using the algorithm implemented in BESA 5.1 (Berg and Scherg, 1994). EEG was segmented from ~200 to 1000 ms relative to face onset, with 200 ms pre-stimulus baseline. Trials contaminated by non-ocular artefacts and saccades were rejected from further analysis using the BESA 5.1 tool, with an amplitude threshold of 100 μV and a gradient criterion of 75 μV. Artefact-free trials were recalculated to average reference, averaged according to experimental condition and digitally low-pass filtered at 40 Hz (12 dB/oct, zero phase shift). Learning phase data were averaged separately for subsequently remembered and subsequently forgotten male and female faces. Data from the test phases were averaged separately for hits and CR of male and female faces. An inclusion criterion of at least 16 trials per condition per participant was applied. For learning phase data, three participants (two females and one male, all exhibiting fewer than 16 trials in one of the ‘subsequently forgotten’ conditions) did not reach this criterion. To keep the design counterbalanced with respect to key and stimulus assignments to experimental conditions when excluding these participants, two additional female and three additional male participants were randomly selected and excluded. Thus, 24 male and 24 female participants entered the statistical analyses of ERP data from the learning phases. Analysis of test phase data was based on all 28 participants per group. The maximum number of trials to be achieved, during learning and test, per participant and condition was 60. More specifically, during learning, the mean numbers of attained trials contributing to each averaged ERP of female participants was for subsequently remembered female faces: 39.71 ± 8.09, for subsequently remembered male faces: 34.83 ± 10.42, for subsequently forgotten female faces: 20.25 ± 5.75 and for subsequently forgotten male faces: 23.79 ± 8.45. For male participants, the corresponding numbers were for subsequently remembered female faces: 35.13 ± 8.34, for subsequently remembered male faces: 36.63 ± 7.88, for subsequently forgotten female faces: 23.50 ± 7.32 and for subsequently forgotten male faces: 21.83 ± 7.43. During test, mean numbers of attained trials for female participants were for hits to female faces: 40.11 ± 6.21, for CR to female faces: 45.43 ± 9.21, for hits to male faces: 36.32 ± 8.56 and for CR to male faces: 46.43 ± 9.16. For male participants, the corresponding numbers were for hits to female faces: 36.50 ± 7.12, for CR to female faces: 48.46 ± 6.43, for hits to male faces: 38.71 ± 6.38 and for CR to male faces: 48.11 ± 7.51.

Early ERP components (P1 and N170) were measured at the electrodes of their respective maximum relative to a 200 ms prestimulus baseline. Mean P1 amplitudes were analyzed at O1 and O2 between 105 and 145 ms. Mean amplitudes of N170 and two following components (P2 and N250) were analyzed at P9 and P10 in consecutive time windows (N170: 155–195 ms, P2: 220–270 ms, N250: 270–400 ms). Finally, mean amplitudes for Dm effects in N200 (220–270 ms), the late Dm (450–1000 ms) and old/new effects (400–675 ms) were calculated at F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Separate mixed-model analysis of variance (ANOVAs) were performed for each component during both learning and test. When appropriate, degrees of freedom were corrected according to the Huynh–Feldt procedure.

RESULTS

Performance

Accuracies during learning were near ceiling (Table 1) and not statistically tested. To analyze reaction time (RT) data from the learning phases (Table 1), a mixed-model ANOVA with the within-subject factors ‘face gender’ (male faces and female faces) and the between-subjects factor ‘participant gender’ (male participants and female participants) was conducted, which yielded no significant results (all P > 0.05).

Memory performance during test phases was analyzed with a corresponding mixed-model ANOVA on d’ (Table 1 and Figure 1), which yielded a significant interaction of ‘participant gender’ (male participants and female participants) (F[1, 54] = 12.29, P < 0.01). Separate analyses for each participant group revealed superior performance for own-gender compared with other-gender faces for both female (F[1, 27] = 7.12, P < 0.05) and male participants (F[1, 27] = 5.19, P < 0.05).

A corresponding ANOVA on C revealed a significant interaction of ‘face gender × participant gender’ (F[1, 54] = 8.91, P < 0.01). Post hoc tests demonstrated female participants to respond more conservatively for male compared with female faces (F[1, 27] = 9.53, P < 0.01). Male participants showed no significant effect (F[1, 27] = 1.06, P > 0.05).

During test, analysis of remember-responses for hits revealed an interaction of ‘face gender × participant gender’ (F[1, 54] = 4.13, P < 0.05), reflecting more remember-responses for male relative to female faces in male participants (F[1, 27] = 7.53, P < 0.05), but no difference in female participants (P > 1). A mixed-model ANOVA on know-responses for hits yielded no significant effects (all P > 0.05).

Face ratings

Attractiveness

An analogous mixed-model ANOVA on attractiveness ratings yielded main effects of ‘face gender’ (F[1, 238] = 4.23, P < 0.05), indicating higher attractiveness ratings for female vs male faces (Table 2), and ‘participant gender’ (F[1, 238] = 7.48, P < 0.001), reflecting overall higher ratings by female vs male participants. Furthermore, a significant interaction of ‘face gender × participant gender’ (F[1, 238] = 13.68, P < 0.001) was observed. Post hoc tests revealed a significant main effect of ‘face gender’ for female participants (F[1, 119] = 10.02, P < 0.01), with higher ratings for female vs male faces, but no corresponding effect for male participants (P > 0.05).

Distinctiveness

A corresponding ANOVA on distinctiveness ratings yielded main effects of ‘face gender’ (F[1, 238] = 3.90, P < 0.05), indicating higher ratings for male vs female faces (Table 2), and ‘participant gender’ (F[1, 238] = 12.67, P < 0.001), reflecting overall higher ratings by male vs female participants. A significant interaction of ‘face gender × participant gender’ (F[1, 238] = 15.30, P < 0.001) was also observed. Post hoc tests revealed no significant effect of ‘face gender’ for female participants (P > 0.05), whereas male participants exhibited significantly higher distinctiveness ratings for male vs female faces (F[1, 119] = 13.82, P < 0.001).

Effects of attractiveness and distinctiveness on memory

To test for potential effects of the above-described differences between male and female participants on face memory, we conducted an ANCOVA on d’ with the within-subjects factor ‘face gender’ and the between-subjects factor ‘participant gender’, with the covariates ‘Δ attractiveness’ (female – male faces) and ‘Δ distinctiveness’ (female – male faces). Replicating the results described earlier, this analysis yielded an interaction of ‘participant gender × face gender’ (F[1, 52] = 9.94, P < 0.01), indicating higher d’ for own-gender compared with other-gender faces for both participant groups. Neither of the two covariates yielded a significant effect (both P > 0.05). Furthermore, in additional analyses neither ‘Δ attractiveness’ nor ‘Δ distinctiveness’ correlated significantly with the memory bias score (both P > 0.05). Accordingly, differences between groups in perceived distinctiveness or attractiveness of male and female faces can not explain the own-gender bias.
Table 1 Performance measures from study and test phases

<table>
<thead>
<tr>
<th></th>
<th>Female participants</th>
<th>Male participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female faces</td>
<td>Male faces</td>
</tr>
<tr>
<td>Study phases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT ± s.d.</td>
<td>843.11 ± 239.89</td>
<td>832.62 ± 210.97</td>
</tr>
<tr>
<td>ACC ± s.d.</td>
<td>0.99 ± 0.02</td>
<td>0.99 ± 0.02</td>
</tr>
<tr>
<td>Test phases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT ± s.d. (hits)</td>
<td>1053.48 ± 143.24</td>
<td>1076.38 ± 154.64</td>
</tr>
<tr>
<td>RT ± s.d. (FA)</td>
<td>854.64 ± 270.65</td>
<td>874.40 ± 348.18</td>
</tr>
<tr>
<td>Hits ± s.d.</td>
<td>0.74 ± 0.14</td>
<td>0.65 ± 0.16</td>
</tr>
<tr>
<td>FA ± s.d.</td>
<td>0.20 ± 0.14</td>
<td>0.20 ± 0.15</td>
</tr>
<tr>
<td>Remember (hits)</td>
<td>0.47 ± 0.22</td>
<td>0.47 ± 0.26</td>
</tr>
<tr>
<td>Know (hits)</td>
<td>0.21 ± 0.15</td>
<td>0.18 ± 0.15</td>
</tr>
<tr>
<td>Guess (hits)</td>
<td>0.02 ± 0.03</td>
<td>0.02 ± 0.03</td>
</tr>
<tr>
<td>d’ ± s.d.</td>
<td>1.69 ± 0.73</td>
<td>1.48 ± 0.67</td>
</tr>
<tr>
<td>C ± s.d.</td>
<td>0.13 ± 0.32</td>
<td>0.31 ± 0.42</td>
</tr>
</tbody>
</table>

Values are expressed as means ± s.d.

Fig. 1 Sensitivity (d’) values depicting recognition performance in the test phases, separate for female and male participants and female and male face gender. Error bars depict standard errors of the mean.

Contact measures

A mixed-model ANOVA on contact time (in hours/week) with the within-subjects factor ‘contact persons’ gender’ and the between-subjects factor ‘participant gender’ yielded no significant effects (all P > 0.05) (Table 3). A corresponding ANOVA on number of contact persons yielded a main effect of ‘participant gender’ (F[1, 54] = 12.52, P < 0.001), indicating that male participants reported higher numbers of contact persons than female participants. A mixed-model ANOVA on contact quality revealed a significant interaction of ‘contact persons’ gender × participant gender’ (F[1, 54] = 5.48, P < 0.05), reflecting more intense contact towards female vs male contact persons in female participants (F[1, 27] = 6.79, P < 0.05), but no significant effect of ‘contact persons’ gender’ in male participants (P > 0.05).

Importantly, the memory bias score did not correlate with contact as measured in hours per week, number of contact persons, or quality of contact (all r < 0.1, all P > 0.05).

ERP during learning

Please note that ERP effects reported here focus on those results directly relevant for the aims of this study (Figure 2). Analysis of P1 and additional results for all remaining ERP components can be found in the Supplementary Material.

Unless stated otherwise, learning phase ERPs were analyzed by calculating mixed-model ANOVAs with the within-subject factors ‘face gender’, ‘subsequent memory’ (subsequently remembered, subsequently forgotten), and ‘hemisphere’ (left, right), as well as the between-subjects factor ‘participant gender’. An ANOVA on N170 amplitudes at P9/P10 revealed a significant main effect of ‘participant gender’ (F[1, 46] = 7.37, P < 0.01), with larger N170 for female vs male participants, as well as a significant interaction of ‘hemisphere × face gender × participant gender’ (F[1, 46] = 4.29, P < 0.05). Post-hoc analyses for male participants yielded a significant effect of ‘face gender’ at electrode P9 (F[1, 23] = 5.48, P < 0.05), reflecting larger N170 amplitudes for female vs male faces, with no significant effect at P10 (F < 1). Female participants did not exhibit significant effects of ‘face gender’ in the N170 (both P > 0.05). The difference in N170 amplitude between female and male faces correlated neither with the differences between female and male faces in attractiveness or distinctiveness ratings, nor with the differences in contact measures toward female vs male persons (all P > 0.05).

A mixed-model ANOVA on P2 amplitudes revealed a significant three-way interaction of ‘subsequent memory × face gender × participant gender’ (F[1, 46] = 8.20, P < 0.01). Post-hoc tests for female participants revealed significantly less positive amplitudes for subsequently remembered male compared with subsequently forgotten male faces (t[23] = -2.16, P < 0.05), whereas no significant difference for subsequently remembered vs forgotten female faces was detected (P > 0.05). For male participants, the post-hoc analysis revealed significantly more positive amplitudes for subsequently remembered vs forgotten male faces (t[23] = 2.14, P < 0.05). Again, no significant difference was obtained for subsequently remembered vs forgotten male faces (P > 0.05). In addition, the P2 amplitude difference between female and male faces was found to be significantly correlated with the difference in rated distinctiveness between female and male faces (r = -0.31, P < 0.05) at electrode P10, reflecting more positive P2 amplitudes for female faces when female faces were rated as less distinctive, and more positive P2 amplitudes for male faces when male faces were rated as less distinctive. None of the remaining correlations were significant (all P > 0.05).

An ANOVA on N200 at left, midline, and right frontal, central and parietal electrodes, with repeated-measures factors ‘site’ (frontal, central, parietal) and ‘laterality’ (left, midline, right; replacing the ‘hemisphere’ factor), revealed an interaction of ‘subsequent memory × face gender × participant gender’ (F[1, 46] = 7.79, P < 0.01). Post hoc tests
Table 2  Face rating: attractiveness and distinctiveness

<table>
<thead>
<tr>
<th></th>
<th>Participants</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Attractiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faces</td>
<td>2.83 ± 0.59</td>
<td>2.51 ± 0.55</td>
<td></td>
</tr>
<tr>
<td>Mean attractiveness rating</td>
<td>2.56 ± 0.57</td>
<td>2.42 ± 0.52</td>
<td></td>
</tr>
<tr>
<td>Distinctiveness</td>
<td>2.69 ± 0.56</td>
<td>2.46 ± 0.50</td>
<td></td>
</tr>
<tr>
<td>Faces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean distinctiveness rating</td>
<td>3.20 ± 0.56</td>
<td>3.19 ± 0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values are expressed as means ± s.d.

Table 3  Contact data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants</th>
<th>Contact persons</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily life</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact time</td>
<td>50.20 ± 38.16</td>
<td>49.58 ± 26.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact numbers</td>
<td>14.46 ± 5.68</td>
<td>13.96 ± 5.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact quality</td>
<td>3.03 ± 0.62</td>
<td>2.03 ± 0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily life</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact time</td>
<td>62.04 ± 68.27</td>
<td>67.86 ± 94.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact numbers</td>
<td>22.21 ± 15.26</td>
<td>26.91 ± 16.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact quality</td>
<td>2.30 ± 0.56</td>
<td>2.27 ± 0.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values are expressed as means ± s.d.

in female participants revealed less negative amplitudes for subsequence

tly remembered vs subsequently forgotten male faces (t{23} = 2.64, P < 0.05), but no significant difference for female faces (P > 0.05). In
male participants, subsequently remembered faces were more negative than subsequently forgotten male faces (t{23} = -2.18, P < 0.05), whereas again no significant effect was observed for female faces (P > 0.05).

Analysis of N250 amplitudes at P9/P10 revealed no significant effects of interest (all P > 0.05). Similar to N170, the difference in N250 between female and male faces correlated neither with the difference between female and male faces in attractiveness or distinctiveness ratings, nor with the difference in contact measures toward female vs male persons (all P > 0.05).

Finally, between 450 and 1000 ms, a mixed-model ANOVA at left, midline, and right frontal, central and parietal electrodes revealed a clear Dm effect of ‘subsequent memory’ (F{1, 46} = 9.57, P < 0.01), with more positive amplitudes for subsequently remembered compared with subsequently forgotten faces. Importantly, there were no significant with ‘participant gender’ or ‘face gender’ (all P > 0.05).

ERP during test phases

Unless stated otherwise, test phase ERPs were analyzed using mixed-model ANOVAs with the within-subject factors ‘face gender’, ‘response type’ (hits, CR), and ‘hemisphere’ (left, right), as well as the between-subjects factor ‘participant gender’ (Figure 3). Analysis of N170 amplitudes yielded no significant effects (all P > 0.05). A mixed-model ANOVA on P2 amplitudes revealed a main effect of ‘face gender’ (F{1, 54} = 5.20, P < 0.05), with more positive amplitudes for female compared with male faces. There was an interaction of ‘hemisphere × response type × participant gender’ (F{1, 54} = 4.17, P < 0.05), although separate post-hoc tests at P9 and P10 for male and female participants revealed no significant effects of ‘response type’ (all P > 0.05).

A corresponding ANOVA on N250 amplitudes resulted only in a main effect of ‘response type’ (F{1, 54} = 5.43, P < 0.05), with more negative amplitudes for hits than CR (all other effects P > 0.05). Correlation analyses revealed no significant effects for the difference in ERP amplitudes between female and male faces in N170, P2 and N250 and the corresponding difference in attractiveness, distinctiveness or contact ratings.

An old/new effect, with more positive amplitudes for hits than CR, was observed between 400 and 675 ms (Figure 3). A mixed-model ANOVA with repeated measurements on ‘site’, ‘laterality’, ‘face gender’ and ‘response type’, as well as the between-subjects factor ‘participant group’, revealed a main effect of ‘response type’ (F{1, 54} = 62.40, P < 0.001), with more positivity for hits vs CRs. Importantly, a three-way interaction of ‘response type × face gender × participant gender’ (F{1, 54} = 4.04, P < 0.05) was additionally observed. A post hoc test indicated larger old/new effects for female compared with male faces (F{1, 27} = 5.12, P < 0.05) in female participants. Although a similar own-gender effect appears to be present in the waveforms of male participants at central electrodes (Figure 3), a post hoc test did not indicate significant differences in old/new effects for female and male faces in male participants (P > 0.05).

Finally, we calculated an analysis which aimed at examining recollection-based recognition memory in isolation. A corresponding mixed model ANOVA with ‘remember’-responses vs CR as levels of the ‘response type’ factor revealed a trend for a significant main effect of ‘response type’ (F{1, 54} = 3.24, P = 0.077).

DISCUSSION

This study is the first that examined neural correlates of an own-gender bias in face memory. We observed more accurate recognition performance for own-gender relative to other-gender faces, in both female and male participants. Analysis of ERP components related to perceptual face processing did not reveal strong evidence for different processing of own- and other-gender faces. More specifically, neither N170, nor P2 and N250 yielded systematic differences between own- and other-gender faces. However, the analysis of memory-related effects in the occipito-temporal P2 and the central N200 during learning revealed differential effects of subsequent memory for male faces in female and male participants, and therefore a potential neural correlate of the own-gender bias. Finally, old/new effects at test were larger for own-gender compared with other-gender faces in the female participant group. These findings are discussed later against previous results from the own-race and own-age bias.

The role of expertise vs social category membership for the OGB

As detailed in the introduction, the majority of previous studies found an OGB in female, but not male participants, who additionally demonstrated worse overall memory performance. In contrast, this study demonstrated a crossover interaction, with superior memory for own-gender faces in both participant groups, in the absence of overall group differences (for similar results see Wright and Sladden, 2003). We thus may conclude that the present OGB was unaffected by general stimulus or participant group effects, and therefore represents a true ‘in-group’ bias.

Both female and male participants reported similar amounts of contact (in h/week and number of contact persons/week). Moreover, although female participants reported more intense contact toward own-gender persons, this difference did not appear to affect memory
performance, as the memory bias towards faces of one’s own gender did not correlate with contact quality toward female vs male persons. Thus, the own-gender bias may largely rely on perceived in-group vs out-group membership.

In addition to this pattern, which held for recognition memory in general, an interesting gender difference was observed with respect to ‘remember’ responses specifically: the OGB in male but not female participants was accompanied by more remember responses, and thus increased recollection, for own-gender compared with other-gender faces. Similarly, previous studies on the ORB reported increased recollection for own-race relative to other-race faces (Meissner et al., 2005; Herzmann et al., 2011). These findings suggest that increased recollection may not exclusively be related to increased perceptual expertise (Herzmann and Curran, 2011), but may similarly occur for social categorization biases. Interestingly, male but not female participants in this study also rated male faces as more distinct-ive than female faces. We thus suggest that distinctiveness enhanced recollection of study phase detail.
ERP during learning

During learning, male participants exhibited a larger N170 for female vs male faces over the left hemisphere, whereas no corresponding difference was seen in female participants. This finding is somewhat discrepant to previous studies that did not detect effects of face gender on N170 (Mouchetant-Rostaing et al., 2000; Mouchetant-Rostaing and Giard, 2003; but see Ito and Urland, 2003 for larger VPP amplitudes for male relative to female faces). However, previous studies observed a small effect of adaptation to facial gender on the left N170, with smaller amplitudes after adaptation (Kovács et al., 2006; Kloth et al., 2010). A critical difference between the present study and previous work may concern task requirements: Although in previous studies participants were asked to exclusively process face gender, we additionally instructed our participants to memorize the facial stimuli, which may have led to stronger motivation to individuate the faces. In line with this interpretation, a very recent study observed larger N170 amplitudes for racial out-group faces only when facial identity was task relevant (Senholzi and Ito, 2013). Accordingly, it appears plausible that the effect of face gender on N170 would not have occurred in a passive viewing or simple categorization task. Of note, while previous studies did not investigate interactions between participant and face gender, we detected face gender effects in male participants only.
Interestingly, a functional magnetic resonance imaging study revealed increased activation of the left-hemispheric amygdala and anterior temporal cortex regions for female relative to male faces in male participants, with no corresponding effects in female participants (Fischer et al., 2004).

The finding of a left-hemispheric mechanism discriminating ‘in-group’ from ‘out-group’ faces in the N170 time range is reminiscent of a similar effect in a recent study on the ORB from our group (Wiese et al., in press). In this study, larger N170 amplitudes for other-race faces were observed, and the magnitude of this effect correlated significantly with the behavioral memory bias. In contrast to these results, the present N170 effect for male participants did not correlate significantly with the memory bias ($r=0.05$, $P=0.82$), which is unlikely to result from a lack of statistical power, given the complete absence of an effect despite the fact that the number of participants and trials was similar across studies. We thus conclude that there is hardly any evidence that the own-gender bias is moderated by participants and trials was similar across studies. We thus conclude that there is hardly any evidence that the own-gender bias is moderated by early perceptual processes as reflected in the N170. Effects of subsequent memory were detected in the occipito-temporal P2 (and polarity-reversed in the central N200). P2 for subsequently remembered vs forgotten male faces was more positive in male participants, whereas female participants demonstrated more positive amplitudes for subsequently forgotten relative to remembered male faces. The P2 has been interpreted to reflect perceived typicality of faces (Stahl et al., 2008; Kaufmann and Schweinberger, 2012). Accordingly, P2 has been found to be less positive for caricatures, which are more distinctive than veridicals, and similarly for naturally distinctive vs non-distinctive faces (Schulz et al., 2012). According to the multi-dimensional face space model (Valentine, 1991) distinctive faces are less densely clustered and therefore harder to categorize on dimensions such as race, age or gender than typical faces. In this study, female participants may have been more accurate at subsequently remembering particularly distinctive male faces, as a categorization of these faces as belonging to a social out-group might have been delayed, which could have resulted in individuation to some extent. Male participants, in contrast, may have been more accurate at subsequently remembering less distinctive or more typical male faces, perhaps because these are easier to categorize as in-group faces, which according to sociocognitive models should result in more in-depth processing. Although this interpretation is somewhat tentative at present, it is generally in line with the observed negative correlation between P2 amplitude and distinctiveness ratings, demonstrating an overall relationship between the two variables.

Finally, a prominent long-lasting Dm effect was observed, but this Dm did not differentiate between subsequently remembered own-gender and other-gender faces.

**ERP at test**

In the test phase of this study, a larger P2 was detected for female compared with male faces. Although no previous studies examined effects of face gender on P2, Ito and Urland (2003) observed a larger parietal N200 for female faces in a similar time range using a mastoid reference. This finding may be seen in line with the present results, as with an average reference (as used in this study) a parietal N200 effect may be observed with reversed polarity at occipito-temporal channels. Although the processes underlying this effect are not entirely clear at present, it is presumably not related to the OGB in memory, as it similarly occurred in both participant groups.

The first ERP component yielding significant differences between correctly remembered and correctly rejected new items was the N250, which was more negative for hits vs CR. This finding is generally consistent with previous studies that demonstrated effects of face learning on the N250 (Tanaka et al., 2006; Kaufmann et al., 2009). Moreover, both own-age and ORB studies demonstrated additional effects in the N250 range, with more negative amplitudes for other-age and other-race relative to own-group faces, independent of whether faces were learned or new items (Stahl et al., 2010; Wiese, 2012). Such effects may be related to the early posterior negativity and may represent the capture of attentional resources by perceptually salient stimuli (Schupp et al., 2007). The absence of a corresponding face gender effect in the N250 time range may reflect the substantial frequency with which other-gender faces are observed.

Finally, larger old/new effects were detected for own-gender faces in female participants. The old/new effect has been interpreted as reflecting conscious recollection, and is known to increase with the amount of study phase detail available (Vilberg and Rugg, 2007). Thus, the present results suggest that female participants not only remember more female compared with male faces, but also recollect more study phase detail for own-gender faces. Our finding of an own-gender bias in the amount of ‘remember’-responses in male but not female participants, however, challenges this interpretation. In this context, it should be noted that previous studies on face memory did not observe the typical distribution of old/new effects known from studies on memory for words (Rugg and Curran, 2007), with frontal effects representing familiarity and (left) parietal effects reflecting recollection-based recognition. Examining recollection and familiarity in face memory, MacKenzie and Donaldson (2007) found a posterior effect of familiarity (300–500 ms), while a later (500–700 ms) and more anterior effect was related to recollection. The time window used in the present study to analyze the old/new effect overlapped with both segments reported in this previous study, and could thus represent a combination of both processes. Consequently, an analysis on ‘remember’-responses vs CR was calculated, which more unambiguously reflects recollection-based recognition. This analysis, however, revealed no own-gender bias in the ERP old/new effect. The present ERP findings thus do not provide evidence for enhanced recollection-based recognition of own-gender faces.

**CONCLUSIONS**

Here, we observed superior memory for own-gender faces in both male and female participants. This behavioral effect was accompanied by differences due to subsequent memory for male faces in the P2/N200 during learning, which could reflect more efficient categorization of typical male faces as out-group in female, but less efficient categorization as in-group by male participants. As the own-gender bias was found to be unrelated to measures of contact, we suggest that these ERP effects do not reflect differences in perceptual expertise, but rather reflect neural correlates of the influence of perceived social category membership on face recognition memory. Overall, our findings suggest that the own-gender bias is mainly related to encoding processes, which is in line with sociocognitive accounts.

**SUPPLEMENTARY DATA**

Supplementary data are available at SCAN online.

**REFERENCES**


