



OPEN

# Behavioral and neural evidence of enhanced long-term memory for untrustworthy faces

Mathias Weymar<sup>1\*</sup>, Carlos Ventura-Bort<sup>1</sup>, Julia Wendt<sup>1,2</sup>  & Alexander Lischke<sup>2\*</sup> 

In daily life, we automatically form impressions of other individuals on basis of subtle facial features that convey trustworthiness. Because these face-based judgements influence current and future social interactions, we investigated how perceived trustworthiness of faces affects long-term memory using event-related potentials (ERPs). In the current study, participants incidentally viewed 60 neutral faces differing in trustworthiness, and one week later, performed a surprise recognition memory task, in which the same old faces were presented intermixed with novel ones. We found that after one week untrustworthy faces were better recognized than trustworthy faces and that untrustworthy faces prompted early (350–550 ms) enhanced frontal ERP old/new differences (larger positivity for correctly remembered old faces, compared to novel ones) during recognition. Our findings point toward an enhanced long-lasting, likely familiarity-based, memory for untrustworthy faces. Even when trust judgments about a person do not necessarily need to be accurate, a fast access to memories predicting potential harm may be important to guide social behaviour in daily life.

During social interactions, individuals automatically form impressions of others based on facial cues. These impressions, which mainly rely on valence evaluation, reflected by trustworthiness judgments<sup>1</sup>, occur rapidly<sup>2</sup> and unintentionally<sup>3,4</sup> and have significant consequences for social behavior. For instance, individuals with untrustworthy faces have lower chances to have their loans funded<sup>5</sup> and are more likely to be sentenced for a crime<sup>6</sup>. Hosts with trustworthy faces, on the other hand, more likely rent their assets with higher charges in the sharing economy domain<sup>7</sup>.

From a neuroscience perspective, it has been suggested that the detection of trustworthiness signals from faces may be mediated by a phylogenetically old survival circuit, which mobilizes the organism for approach and avoidance behavior in social interactions<sup>2,8</sup>. Critically, the amygdala, a bilateral structure from the medial lobe, seems to be involved by engaging neural systems that support attention, learning and memory processes<sup>9–11</sup>. Accordingly, perception of trustworthy and untrustworthy faces is associated with enhanced amygdala<sup>3,12–14</sup> (for meta-analyses see<sup>15,16</sup>), as well as occipital and temporal cortical region activation<sup>12</sup> to prioritize perceptual processing of socially salient information. This pattern of neural activity has also been found for positively and negatively valenced faces<sup>17,18</sup> suggesting that the processing of cues signaling trustworthiness is rather related to general valence processing. Therefore, these findings support the emotion face overgeneralization hypothesis<sup>2,19,20</sup>, i.e. that responding appropriately to emotional states is such an adaptive function that prioritized processing also extends to facial cues that only resemble certain emotional expressions.

Interestingly, despite a prominent role of the amygdala in extracting subtle cues of both positive and negative faces, a negativity bias has been often reported: For instance, individuals with bilateral amygdala damage perceive untrustworthy faces as more trustworthy<sup>21</sup>. Consistent with this finding, greater amygdala sensitivity for untrustworthy than trustworthy faces is also reported in healthy individuals<sup>2,12,14,22</sup>. This suggests that processing signals of untrustworthiness may have higher adaptive value than signals of trust<sup>2,23,24</sup>, which is also consistent with animal and human studies showing a generally stronger defense motivation for aversive cues and contexts than for cues triggering approach motivation<sup>2,25,26</sup>.

Perceived untrustworthiness not only affects perception but also learning and memory processes<sup>27,28</sup>. When individuals learn that another individual cannot be trusted, they avoid strangers who look similar to that individual<sup>27</sup>. Interestingly, this avoidance pattern toward a putative untrustworthy individual is also reflected in amygdala

<sup>1</sup>Department of Biological Psychology and Affective Science, University of Potsdam, Potsdam, Germany.

<sup>2</sup>Department of Biological and Clinical Psychology, University of Greifswald, Greifswald, Germany. \*email: [mathias.weymar@uni-potsdam.de](mailto:mathias.weymar@uni-potsdam.de); [alexander.lischke@uni-greifswald.de](mailto:alexander.lischke@uni-greifswald.de)

activation. Furthermore, recent recognition memory studies using immediate testing from our own group and others suggest that the mere exposure to untrustworthy faces result in better memory for untrustworthy than for trustworthy faces<sup>28–30</sup>. These findings indicate that subtle cues of untrustworthiness are automatically learned and remembered. This memory advantage for untrustworthy faces may serve an adaptive function to avoid potential exploitation and harm in future social interactions.

Following up on recent recognition memory studies<sup>28–30</sup>, the present study addressed two open research questions: *First*, because these studies only tested recognition memory using short retention intervals, it is not known whether the memory advantage for untrustworthy faces persists after longer delays or only occurs when retrieval is tested immediately. The use of a longer retention interval would also clarify whether memory for untrustworthy information undergoes deeper consolidation or is just attributed to enhanced perceptual processing during encoding<sup>22,31,32</sup> as recently suggested<sup>30</sup>. *Second*, it is widely agreed upon that recognition memory can be based on two qualitatively distinct cognitive processes: familiarity and recollection<sup>33–35</sup>. Familiarity is assumed to reflect a fast-acting and automatic process that is generally associated with the feeling of knowing an item without being able to recall specific details. Recollection, on the other hand, is considered a slower more elaborate memory process that includes the ability to recognize past events with additional spatial, temporal or contextual information<sup>36</sup>. Currently, it is not understood whether memory retrieval for untrustworthy faces is more relying on the process of familiarity or recollection. We therefore measured differences in brain potentials during viewing old and new items during a recognition memory test<sup>36–39</sup>. An early frontal Old/New difference (300–500 ms) is hypothesized to be sensitive to memory processes based on familiarity, while a later parietal Old/New difference (>500 msec) has been proposed as a neural correlate of recollection, as it is enhanced by depth of processing, as well as for correct source and “remember” judgments in a recognition task<sup>39</sup>.

In line with our behavioral study<sup>28</sup> we expected to find enhanced recognition memory for untrustworthy, compared to trustworthy faces. For ERPs, two hypotheses can be assessed predicting the influence of trustworthiness on the electrophysiological indicators of familiarity and recollection: Given that prior long-term memory ERP studies reported enhanced recollection-based retrieval for motivationally relevant stimuli<sup>37,38</sup>, memory performance and late parietal Old/New differences may be heightened for untrustworthy, compared to trustworthy faces. Alternatively, features of trustworthiness are only subtle and help to rapidly form impressions about potential harm or joy<sup>2</sup> that do not necessarily need elaborated (contextual) memory storage. Better memory for untrustworthy faces could therefore also be related to the process of familiarity, as reflected by early frontal ERP Old/New differences.

## Method

**Participants.** Participants were 32 German-speaking students (6 male, 26 female; mean age: 25.1 years; 4 left-handed) from the University of Potsdam who participated for course credits or financial compensation. Based on our prior recognition memory study<sup>28</sup> showing a medium-sized effect size (Cohen's  $d = 0.57$ ) difference in memory for untrustworthy and trustworthy faces effect size we estimated a sample size of  $n = 26$  for the present study ( $\alpha = 0.05$ ;  $\beta = 0.20$ ). Prior experience has shown a small number of drop-outs, for instance, due to bad ERP quality. We therefore increased the number of participants to the final sample size. Each individual provided written informed consent for a protocol approved by the Review Board of the German Psychological Society and in accordance with the Declaration of Helsinki. All participants had normal or corrected-to-normal vision.

**Stimuli.** The stimulus material consisted of 120 neutral faces with direct gaze, which were previously evaluated as high (30 female, 30 male faces) and low (30 female, 30 male faces) in trustworthiness by an independent sample (see<sup>31</sup> for details on set construction). Following an established procedure<sup>40</sup>, the faces were converted into greyscale, equalized in size, position and luminance and surrounded by an elliptic mask using Adobe Photoshop CS4 (Adobe Systems Inc., San Jose, CA) and Matlab 7.7 (MathWorks Inc., Natick, MA) to minimize the influence of expression-irrelevant features on face processing during the present task.

**Procedure.** The study consisted of two experimental sessions: an encoding session and a recognition memory session one week later. The sessions took place in a sound-attenuated dimly lit room. Procedure and stimuli were the same as in our prior studies<sup>28,31</sup>. Participants viewed 60 faces (30 trustworthy and 30 untrustworthy faces) in pseudorandom order. Participants were instructed to pay attention to the faces but were not informed that the faces differed in trustworthiness and that a recognition test would follow (incidental encoding). Each face was presented once for 3,000 ms, followed by an intertrial interval (ITI) of 5,500, 6,000, or 6,500 ms. A fixation cross was present in all faces trials and ITIs to ensure that participants fixated the center of the screen.

One week later, participants returned to the lab for the recognition memory session. After the electrode net was placed on the head, participants viewed the 60 old faces together with 60 (30 trustworthy and 30 untrustworthy faces) new faces in pseudorandom order. Each trial started with the presentation of a face for 3,000 ms. After picture offset, the question “Old/New?” was presented and participants had to decide whether the face had been viewed before during encoding (i.e., by pressing “Old”) or not (i.e., by pressing “New”). Responses were made on two different buttons on a keyboard and response buttons were counterbalanced across participants. The assignment of the faces to the old and new face set was counterbalanced across participants. After the recognition memory task, the participants rated one of the two sets (i.e., their encoding set) for subjective trustworthiness using a similar rating task as in our previous studies<sup>28,31</sup>. In this task, participants were encouraged to rely on their feelings and to rate the trustworthiness of the faces as fast as possible on a rating scale, ranging from 1 (*untrustworthy*) to 9 (*trustworthy*). In line with our prior work<sup>28,31</sup>, analyses of these ratings revealed that the preselected untrustworthy faces were rated as less trustworthy ( $M = 4.36$ ;  $SD = 1.21$ ) than the preselected trustworthy faces

	Trustworthy faces	Untrustworthy faces
Hits (H)	0.59 (0.15)	0.61 (0.14)
False Alarms (FAs)	0.38 (0.17)	0.32 (0.16)*
<b>Discrimination indices</b>		
Pr	0.21 (0.15)	0.29 (0.18)*
$d'$	0.61 (0.50)	0.82 (0.56)
<b>Response bias indices</b>		
Br	0.47 (0.20)	0.44 (0.17)
C	0.07 (0.44)	0.11 (0.35)

**Table 1.** Recognition memory performance. Numbers in parentheses indicate SD. Higher  $Pr$  and  $d'$  values indicate better discriminability between old and new faces.  $Br$  values higher than 0.5 indicate liberal response criteria; lower  $Br$  values suggest conservative response bias. Negative  $C$  values correspond to a liberal response bias; positive  $C$  values correspond to a conservative response bias. \* $p < 0.01$ .

( $M = 5.89$ ;  $SD = 0.96$ ),  $t(31) = 12.64$ ,  $p < 0.001$ , confirming the validity of the face set (see also supplementary information for further analysis showing significant differences in trustworthiness when controlled for face typicality and distinctiveness).

**Electrophysiological recording.** Electrophysiological signals were continuously recorded from 129-sensor nets using an Electrical Geodesic system (EGI, Eugene, OR) and digitized at a rate of 250 Hz, using the vertex sensor (Cz) as recording reference. Scalp impedance of all channels was kept below 50 k $\Omega$ , as recommended by the manufacturer's guidelines. Online, all channels were bandpass filtered (0.1–100 Hz). Offline, ElectroMagnetic EncaphaloGraphy software (EMEGS<sup>41</sup>) was used to preprocess the data for subsequent analyses, including low-pass filtering (40 Hz), artifact detection, sensor interpolation, baseline correction, and conversion to an average reference<sup>42</sup>. The MATLAB-based toolbox BioSig<sup>43</sup> was used for eye movement and blink artifacts corrections of the extracted epochs. This method is based on linear regression to reliably remove electrooculogram activity from the EEG<sup>44</sup>. Stimulus-synchronized epochs were extracted from 100 ms before to 1,200 ms after face onset and baseline corrected (100 ms prior to stimulus onset).

**Data analysis.** ERPs were computed for each sensor and participant. Only trials with correct responses were included in ERP averages. In consideration of previous research showing distinct ERP old/new effects<sup>37,39,45,46</sup> and based on visual inspection of the waveforms, mean ERP amplitudes were analyzed in an early window from 350 to 550 ms over frontal regions (EGI sensors: 5, 9, 10, 11, 12, 15, 16, 18, and 22), and in a late window from 550 to 750 ms over a centro-parietal cluster (EGI sensors: 7, 31, 55, 80, 106, and 129), where the difference between old and new conditions was maximal. To examine the effects of trustworthiness on recognition-related ERPs, data were analyzed in a two-way ANOVA using the factors *Memory* (old, new) and *Trustworthiness* (untrustworthy, trustworthy) as repeated measures for the early and late time window, separately. To note, we particularly focused on ERPs related to explicit recognition memory. ERP analyses related to face processing and repetition<sup>47</sup> did not reveal any trustworthiness  $\times$  memory interactions and were provided as supplementary information. Furthermore, ERP analyses related to face processing during encoding showing that the main findings of this study are not driven by differences in distinctiveness between trustworthy and untrustworthy faces (e.g., P200)<sup>47</sup> were also included as supplementary information.

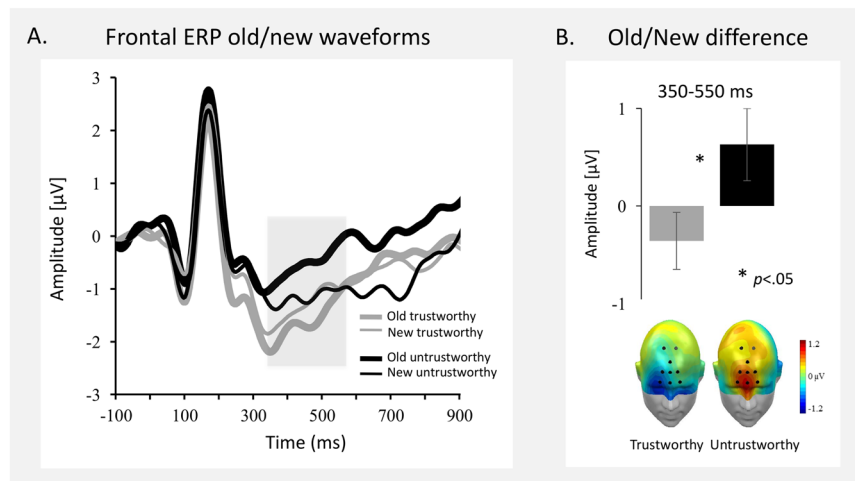
For behavioral performance, hit rate (H), false alarm rate (FA), recognition accuracy ( $Pr = H - FA$ ), and response bias [ $Br = p(FA)/p(1 - Pr)$ ] as recommended by Snodgrass and Corwin (two-high-threshold model)<sup>48</sup> were analyzed for trustworthy and untrustworthy faces. These indices were complemented by analyses of detection and bias parameters derived from signal detection theory ( $d' = z(H) - z(FA)$  and  $C = -0.5[(z(H) + z(FA))]$ ). T-tests for dependent means were computed for each of the behavioral indices to test for significant differences. All statistical analyses were conducted using IBM SPSS Statistics 24 and JMP 5.0.

## Results

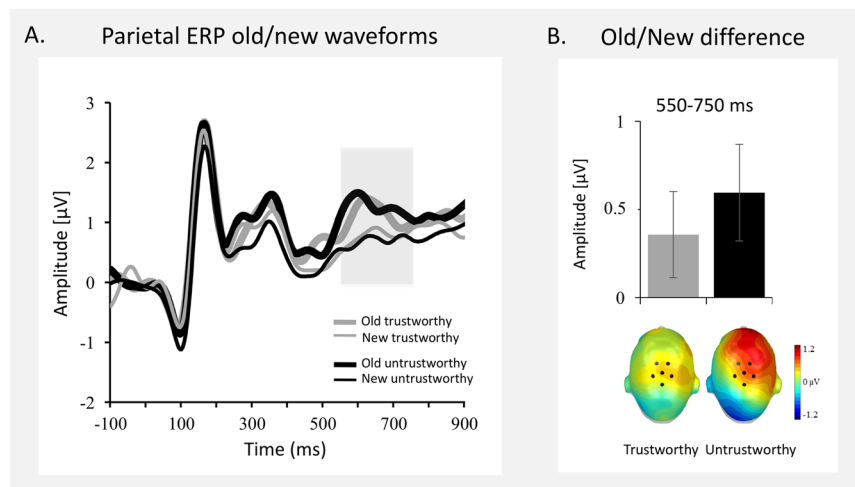
**Behavioral performance.** Table 1 lists participants' memory performance for trustworthy and untrustworthy faces. Hit rates did not differ between untrustworthy and trustworthy faces ( $t(31) < 1$ ). However, new untrustworthy faces were more easily recognized than new trustworthy faces, as indicated by significantly lower false alarm rates,  $t(31) = 3.56$ ,  $p = 0.001$ . As expected, recognition accuracy was higher for untrustworthy than for trustworthy faces as shown by larger  $Pr$ ,  $t(31) = 3.41$ ,  $p = 0.002$ , and  $d'$  values,  $t(31) = 2.75$ ,  $p = 0.01$ . Response bias for faces did not differ as a function of trustworthiness as indicated by similar  $Br$ ,  $t(31) = 1.24$ ,  $p = 0.23$ , and  $C$  values,  $t(31) < 1$ . Taken together, memory performance was better for untrustworthy than trustworthy faces when tested after a one week retention interval.

**ERPs.** *Frontal Old/New effect reflecting familiarity.* Figure 1A illustrates the grand average ERPs for correctly recognized old and new faces that differed in trustworthiness collapsed across a representative frontal sensor cluster.

At frontal regions, in the early time window (350–550 ms), a main effect of the factor *Trustworthiness* was found,  $F(1,31) = 6.90$ ,  $p = 0.013$ ,  $\eta_p^2 = 0.18$ , which interacted with the factor *Memory*,  $F(1,31) = 4.47$ ,  $p = 0.043$ ,



**Figure 1.** Frontal ERP Old/New effect reflecting familiarity. **(A)** Grand-averaged waveforms at representative frontal cluster for correctly recognized old and new trustworthy and untrustworthy faces. The shaded area represents the early (350–550 ms) time window used for the analyses. **(B)** ERP Old/New difference of the mean amplitudes and scalp topographies of the ERP difference (350–550 ms) for trustworthy and untrustworthy faces.



**Figure 2.** Centro-parietal ERP Old/New effect reflecting recollection. **(A)** Grand-averaged waveforms at representative centro-parietal cluster for correctly recognized old and new trustworthy and untrustworthy faces. The shaded area represents the late (550–750 ms) time window used for the analyses. **(B)** ERP Old/New difference of the mean amplitudes and scalp topographies of the ERP difference (550–750 ms) for trustworthy and untrustworthy faces.

$\eta_p^2 = 0.13$ . Follow-up testing revealed that the Old/New difference was larger for untrustworthy than for trustworthy faces,  $t(31) = 2.11$ ,  $p = 0.043$  (see Fig. 1B). In the later time window (550–750 ms), the ERP old/new difference continued, *Memory*:  $F(1,31) = 5.71$ ,  $p = 0.023$ ,  $\eta_p^2 = 0.16$ , but the interaction with *Trustworthiness* only reached trend level,  $F(1,31) = 3.75$ ,  $p = 0.062$ ,  $\eta_p^2 = 0.11$ . *Trustworthiness* did not reach significance ( $F < 1$ ).

*Parietal Old/New effect reflecting recollection.* Figure 2A illustrates the grand average ERPs for correctly recognized old and new faces that differ in trustworthiness collapsed across a representative centro-parietal sensor cluster.

No significant main effects or interactions (all  $F_s < 3.19$ ;  $p > 0.084$ ) were observed at centro-parietal regions in the early time window (350–550 ms). However, in the time window between 550 and 750 ms, old faces prompted larger ERP positivity than new faces, as reflected by a main effect of *Memory*,  $F(1,31) = 6.49$ ,  $p = 0.016$ ,  $\eta_p^2 = 0.17$  (see Fig. 2B). No other main effects or interactions reached significance in this later time window (all  $F_s < 1$ ).

Taken together, untrustworthy faces generated larger early ERP old/new effects over frontal electrodes than trustworthy faces, indicating facilitated familiarity-based recognition. No such differences were observed for the recollection-sensitive late centro-parietal ERP Old/New effect.

## Discussion

In the present study, ERPs were measured to assess the effects of facial trustworthiness on long-term memory. Consistent with prior behavioral studies using immediate recognition tests<sup>28–30</sup>, better memory discrimination was found for untrustworthy than for trustworthy faces after a one week retention interval. During memory retrieval, correctly recognized old faces evoked enhanced ERP positivity, relative to unseen novel faces. This ERP Old/New difference was most pronounced for untrustworthy faces over frontal electrode sites in an early 350–550 ms window. The results provide behavioral and neural evidence for a long-lasting memory advantage for untrustworthy faces.

Using ERPs, we were able to assess how memory for trustworthy and untrustworthy faces is formed and retrieved. In line with previous studies<sup>36,37,39</sup>, we found enhanced ERP positivity for old, relative to new items with specific spatio-temporal characteristics. While the late ERP Old/New effect reflecting the memory process of successful recollection did not vary as a function of trustworthiness, we found that particularly untrustworthy faces modulated the early Old/New effect, a putative correlate of familiarity-based remembering<sup>49</sup> (which may also involve implicit memory processes<sup>50</sup>). According to dual-process theory familiarity has been considered as a fast-acting, less effortful and relatively automatic memory process that works with greater efficiency than recollection<sup>35,51</sup>. While familiarity supports simply knowing that an event was encountered before, recollection is understood as retrieving additional (contextual) details (e.g., temporal, spatial, situational) that are associated with an event. For instance, prior ERP studies (for review see<sup>36</sup>) found that the early frontal old/new effect, compared to the late parietal old/new effect, is insensitive by depth of processing<sup>52</sup> and divided attention<sup>53</sup> during encoding. Furthermore, the early frontal old/new effect varies with familiarity strength<sup>54</sup> and memory confidence<sup>37</sup>, the parietal old/new effect, on the other hand is specific for recollection (e.g., sensitive to the amount of information recollected<sup>55</sup>, correct source judgments<sup>56</sup> and recollection based Remember judgments<sup>57</sup> and high confidence memory judgments<sup>37</sup>). The ERP finding that recognition of untrustworthy individuals is based on familiarity (reflected by early frontal old/new differences) fits with the emotion overgeneralization hypothesis<sup>2,20</sup>, which posits that face-based valence judgments are an adaptive mechanism, which helps in a rapid and efficient way to avoid upcoming harm. From this view, the potential cost of responding with approach behavior to an untrustworthy individual is higher than the potential cost of avoiding a trustworthy individual. This also implies that untrustworthy faces do not necessarily need to be stored in a more complex contextual fashion (e.g., recollection) when the evolutionary purpose is to quickly recognize socially relevant information. Interestingly, contrary to prior ERP studies using emotional scenes<sup>38,58</sup> or faces<sup>45</sup>, we did not find a modulation in the recollection-sensitive parietal old/new effect. However, emotional scenes and faces contain more contextual feature information and draw more attention<sup>59,60</sup> than neutral faces, which may have resulted in enhanced recollection-specific contextual binding<sup>61</sup> and enhanced parietal ERP old/new differences in these studies.

Enhanced acontextual familiarity-based memory for untrustworthy faces may also have serious consequences on social decision making. For instance, previous lab and more real life studies found that individuals whose faces are perceived as untrustworthy are less trusted<sup>62,63</sup>, are less likely to have their loans funded<sup>5</sup> and are more likely to be sentenced for crimes<sup>6</sup>. Given that face-based social inferences about a person's trustworthiness are not necessarily accurate and reliable in general<sup>19</sup>, familiarity-based remembering may come at costs. On the one hand, with limited additional (reputational) information about a person, individuals can instantaneously form first impressions that are stored in a less contextual manner (which likely engage less processing resources) that may help for quick decision making whether a person can be trusted or not (irrespective of its validity). Memory for untrustworthy faces may in turn also affect other social preferences for fairness and cooperation and contribute to less prosocial behavior and emotions, including altruism<sup>64</sup> and empathy<sup>65</sup>. On the other hand, familiarity-based memory for untrustworthy faces may even more strongly hinder to derive valid appearance-based impressions about other persons. Social decision making strongly relies on available contextual information (e.g., reputational information, social context etc., see<sup>66</sup>). Recognizing an untrustworthy face including remembrance of such contextual information may therefore also be adaptive (in the long term) in a sense that it helps to correct invalid appearance-based judgments. It must be noted, however, that our study was intended to investigate long-term recognition memory for trustworthy and untrustworthy looking faces that were incidentally encoded once, mimicking a brief social encounter in real life<sup>29</sup>. Future research need to explore whether additional reputational information and increased experience (e.g. through repetition of trials) may also enhance recollection memory.

Finally, we can only speculate about the underlying neural mechanism leading to enhanced familiarity-based remembering for untrustworthy faces in the present study. As suggested by Tsukiura<sup>67</sup>, valence-based memory for faces is mediated by the amygdala, which detects emotional properties of the faces (such as conveyed by subtle cues<sup>2</sup>) and interacts with the medial orbitofrontal cortex (OFC) and insular cortex (IC) to process positive and negative signals, respectively. These regions in turn may modulate a network consisting of hippocampus (HC) and fusiform faces area to enhance memory for these faces depicting affective signals<sup>67,68</sup>. Given that memory for untrustworthy faces was mediated by familiarity, it is likely, however, that the perirhinal cortex (PrC) rather than the HC, contribute to the memory-enhancing effect since the PrC has been specifically related representing item familiarity<sup>69–71</sup>.

Taken together, we found that the memory advantage for faces signaling untrustworthiness is remarkably stable over time. In support of the overgeneralization hypothesis, we found electrophysiological evidence that memory for untrustworthy faces is likely linked to familiarity-based recognition, a fast, less specific and automatic retrieval process, which helps to quickly recognize those who may cause potential harm in social interactions.

Received: 8 May 2019; Accepted: 2 December 2019;

Published online: 16 December 2019



## References

- Oosterhof, N. N. & Todorov, A. The functional basis of face evaluation. *Proc. Natl. Acad. Sci. USA* **105**, 11087–11092 (2008).
- Todorov, A. Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Ann. N. Y. Acad. Sci.* **1124**, 208–224 (2008).
- Freeman, J. B., Stolier, R. M., Ingbreten, Z. A. & Hehman, E. A. Amygdala responsivity to high-level social information from unseen faces. *J. Neurosci.* **34**, 10573–10581 (2014).
- Jessen, S. & Grossmann, T. Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia* **126**, 46–53 (2019).
- Duarte, J., Siegel, S. & Young, L. Trust and credit: The role of appearance in peer-to-peer lending. *Rev. Financ. Stud.* **25**, 2455–2484 (2012).
- Porter, S., ten Brinke, L. & Gustaw, C. Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychol. Crime Law* **16**, 477–491 (2010).
- Ert, E., Fleischer, A. & Magen, N. Trust and reputation in the sharing economy: the role of personal photos in Airbnb. *Tour. Manag.* **55**, 62–73 (2016).
- Slepian, M. L., Young, S. G. & Harmon-Jones, E. An approach-avoidance motivational model of trustworthiness judgments. *Motiv. Sci.* **3**, 91–97 (2017).
- Adolphs, R. What does the amygdala contribute to social cognition? *Ann. N. Y. Acad. Sci.* **1191**, 42–61 (2010).
- Lang, P. J. & Bradley, M. M. Emotion and the motivational brain. *Biol. Psychol.* **84**, 437–450 (2010).
- Weymar, M. & Schwabe, L. Amygdala and Emotion: The Bright Side of It. *Front. Neurosci.* **10**, 224 (2016).
- Engell, A. D., Haxby, J. V. & Todorov, A. Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *J. Cogn. Neurosci.* **19**, 1508–1519 (2007).
- Said, C. P., Baron, S. G. & Todorov, A. Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *J. Cogn. Neurosci.* **21**, 519–528 (2009).
- Winston, J. S., Strange, B. A., O’Doherty, J. & Dolan, R. J. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* **5**, 277–283 (2002).
- Mende-Siedlecki, P., Said, C. P. & Todorov, A. The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Soc. Cogn. Affect. Neurosci.* **8**, 285–299 (2013).
- Bzdok, D. *et al.* ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Struct. Funct.* **215**, 209–223 (2011).
- Sergerie, K., Chochol, C. & Armony, J. L. The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **32**, 811–830 (2008).
- Sabatinelli, D. *et al.* Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage* **54**, 2524–2533 (2011).
- Todorov, A., Olivola, C. Y., Dotsch, R. & Mende-Siedlecki, P. Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annu. Rev. Psychol.* **66**, 519–545 (2015).
- Zebrowitz, L. A. & Montepare, J. M. Social psychological face perception: Why appearance matters. *Soc. Personal. Psychol. Compass* **2**, 1497–1517 (2008).
- Adolphs, R., Tranel, D. & Damasio, A. R. The human amygdala in social judgment. *Nature* **393**, 470–474 (1998).
- Santos, S., Almeida, I., Oliveiros, B. & Castelo-Branco, M. The role of the amygdala in facial trustworthiness processing: a systematic review and meta-analyses of fMRI studies. *PLoS One* **11**, e0167276 (2016).
- Cosmides, L. & Tooby, J. Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind — Evolutionary Psychology and the Generation of Culture* (pp. 163–228). New York: Oxford University Press (1992).
- Rozin, P. & Royzman, E. B. Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psychol. Rev.* **5**, 296–320 (2001).
- Bradley, M. M. & Lang, P. J. Emotion and motivation. In J. T. Cacioppo, L. G. Tassinari & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 581–607). New York, NY, US: Cambridge University Press (2007).
- Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In: Koch, S. (Ed.): *Psychology: A study of a science* (Vol 2). New York: McGraw-Hill (1959).
- FeldmanHall, O. *et al.* Stimulus generalization as a mechanism for learning to trust. *Proc. Natl. Acad. Sci. USA* **115**, E1690–E1697 (2018).
- Wendt, J., Weymar, M., Junge, M., Hamm, A. O. & Lischke, A. Heartfelt memories: Cardiac vagal tone correlates with increased memory for untrustworthy faces. *Emotion* **19**, 178–182 (2019).
- Rule, N. O., Slepian, M. L. & Ambady, N. A memory advantage for untrustworthy faces. *Cognition* **125**, 207–218 (2012).
- Mattarozzi, K., Todorov, A. & Codispoti, M. Memory for faces: the effect of facial appearance and the context in which the face is encountered. *Psychol. Res.* **79**, 308–317 (2015).
- Lischke, A., Junge, M., Hamm, A. O. & Weymar, M. Enhanced processing of untrustworthiness in natural faces with neutral expressions. *Emotion* **18**, 181–189 (2018).
- Marzi, T., Righi, S., Ottonello, S., Cincotta, M. & Viggiano, M. P. Trust at first sight: Evidence from ERPs. *Soc. Cogn. Affect. Neurosci.* **9**, 63–72 (2014).
- Jacoby, L. L. A process dissociation framework: Separating automatic from intentional uses of memory. *J. Mem. Lang.* **30**, 513–541 (1991).
- Mandler, G. Recognizing: The judgment of previous occurrence. *Psychol. Rev.* **87**, 252–271 (1980).
- Yonelinas, A. P. The nature of recollection and familiarity: A review of 30 years of research. *J. Mem. Lang.* **46**, 441–517 (2002).
- Mecklinger, A. & Jäger, T. Episodic memory storage and retrieval: Insights from electrophysiological measures. In: Rösler, F., Ranganath, C., Röder, B., Kluwe, R. H. (Eds.) *Neuroimaging and psychological theories of human memory* (pp. 357–382). Oxford, UK: Oxford University Press (2009).
- Weymar, M., Löw, A., Melzig, C. A. & Hamm, A. O. Enhanced long-term recollection for emotional pictures: Evidence from high-density ERPs. *Psychophysiology* **46**, 1200–1207 (2009).
- Weymar, M., Löw, A. & Hamm, A. O. Emotional memories are resilient to time: Evidence from the parietal ERP old/new effect. *Hum. Brain Mapp.* **32**, 632–640 (2011).
- Rugg, M. D. & Curran, T. Event-related potentials and recognition memory. *Trends Cogn. Sci.* **11**, 251–257 (2007).
- Lischke, A. *et al.* Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. *Psychoneuroendocrinology* **37**, 475–481 (2012).
- Peyk, P., De Cesarei, A. & Junghöfer, M. ElectroMagnetoEncephalography software: Overview and integration with other EEG/MEG toolboxes. *Comput. Intell. Neurosci.* **2011**, 1–10 (2011).
- Junghöfer, M., Elbert, T., Tucker, D. M. & Rockstroh, B. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology* **37**, 523–532 (2000).
- Vidaurre, C., Sander, T. H. & Schlögl, A. BioSig: the free and open source software library for biomedical signal processing. *Comput. Intell. Neurosci.* **2011**, 1–12 (2011).
- Schlögl, A. *et al.* A fully automated correction method of EOG artifacts in EEG recordings. *Clin. Neurophysiol.* **118**, 98–104 (2007).
- Johansson, M., Mecklinger, A. & Treese, A. Recognition memory for emotional and neutral faces: An event-related potential study. *J. Cogn. Neurosci.* **16**, 1840–1853 (2004).

46. Ventura-Bort, C. *et al.* Binding neutral information to emotional contexts: Brain dynamics of recognition after a long retention interval. *Cogn. Affect. Behav. Neurosci.* **16**, 234–247 (2016).
47. Wiese, H., Altmann, C. S. & Schweinberger, S. R. Effects of attractiveness on face memory separated from distinctiveness: Evidence from event-related brain potentials. *Neuropsychologia* **56**, 26–36 (2014).
48. Snodgrass, J. G. & Corwin, J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *J. Exp. Psychol. Gen.* **117**, 34–50 (1988).
49. Tsivilis, D., *et al.* Old-new ERP effects and remote memories: the late parietal effect is absent as recollection fails whereas the early mid-frontal effect persists as familiarity is retained. *Front. Hum. Neurosci.* **9**, 532 (2015).
50. Voss, J. L., Lucas, H. D. & Paller, K. A. More than a feeling: Pervasive influences of memory without awareness of retrieval. *Cogn. Neurosci.* **3**, 193–207 (2012).
51. Brown, M. W. & Aggleton, J. P. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* **2**, 51–61 (2001).
52. Rugg, M. D., *et al.* Neural correlates of depth of processing effects on recollection: Evidence from brain potentials and positron emission tomography. *Exp. Brain Res.* **123**, 18–23 (1998).
53. Curran, T. Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia* **42**, 1088–1106 (2004).
54. Woodruff, C. C., Hayama, H. R. & Rugg, M. D. Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Res.* **1100**, 125–135 (2006).
55. Vilberg, K. L., Moosavi, R. F. & Rugg, M. D. The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Res.* **1122**, 161–170 (2006).
56. Wilding, E. L. & Rugg, M. D. An event-related potential study of recognition memory with and without retrieval of source. *Brain* **119**, 889–905 (1996).
57. MacLeod, C. A. & Donaldson, D. I. Investigating the Functional Utility of the Left Parietal ERP Old/New Effect: Brain Activity Predicts within But Not between Participant Variance in Episodic Recollection. *Front. Hum. Neurosci.* **11**, 580 (2017).
58. Wirkner, J., Schulz, P., Ventura-Bort, C., Hamm, A. O. & Weymar, M. Event-related potentials of emotional and neutral memories: The role of encoding position and delayed testing. *Psychophysiology* **55**, e13069 (2018).
59. Schupp, H. T. *et al.* The facilitated processing of threatening faces: An ERP analysis. *Emotion* **4**, 189–200 (2004).
60. Wangelin, B. C., Bradley, M. M., Kastner, A. & Lang, P. J. Affective engagement for facial expressions and emotional scenes: the influence of social anxiety. *Biol. Psychol.* **91**, 103–110 (2012).
61. Mather, M. & Sutherland, M. R. Arousal-biased competition in perception and memory. *Perspect. Psychol. Sci.* **6**, 114–133 (2011).
62. Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J. & Sanfey, A. G. Seeing is believing: Trustworthiness as a dynamic belief. *Cogn. Psychol.* **61**, 87–105 (2010).
63. Rezsescu, C., Duchaine, B., Olivola, C. Y. & Chater, N. Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE* **7**, e34293 (2012).
64. Adolphs, R. & Tusche, A. How attention to faces guides prosocial behavior. *Curr. Dir. Psychol. Sci.* **26**, 282–287 (2017).
65. Sessa, P. & Meconi, F. Perceived trustworthiness shapes neural empathic responses toward others' pain. *Neuropsychologia* **79**, 97–105 (2015).
66. Lee, V. K. & Harris, L. T. How social cognition can inform social decision making. *Front. Neurosci.* **7**, 259 (2013).
67. Tsukiura, T. Neural mechanisms underlying the effects of face-based affective signals on memory for faces: a tentative model. *Front. Integr. Neurosci.* **6**, 50 (2012).
68. Dolcos, F. *et al.* Emerging directions in emotional episodic memory. *Front. Psychol.* **8**, 1867 (2017).
69. Martin, C. B., McLean, D. A., O'Neil, E. B. & Kohler, S. Distinct familiarity-based response patterns for faces and buildings in perirhinal and parahippocampal cortex. *J. Neurosci.* **33**, 10915–10923 (2013).
70. Davachi, L. Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* **16**, 693–700 (2006).
71. Ranganath, C. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus* **20**, 1263–1290 (2010).

## Acknowledgements

We thank Viola Werling and Manon Giraudier for their help with the data collection. M.W. was supported by a grant provided by the German Research Foundation (DFG; WE 4801/3-1). J.W. was supported by a grant provided by the German Research Foundation (DFG; WE 5873/1-1). A.L. was supported by a grant provided by the German Research Foundation (DFG; LI 2517/2-1). We also acknowledge the support of the DFG and Open Access Publishing Fund of University of Potsdam for funding the publication fees.

## Author contributions

M.W. and A.L. designed the study. M.W. and C.V.B. collected the data. M.W. and C.V.B. analyzed the data. M.W. wrote the manuscript. M.W., C.V.B., J.W. and A.L. contributed to writing, reviewing and editing of the manuscript. All authors approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-55705-7>.

**Correspondence** and requests for materials should be addressed to M.W. or A.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019