

Werner Sommer^{1,2}, Krzysztof Kotowski³, Yang Shi², Adam Switonski⁴, Andrea Hildebrandt⁵, and Katarzyna Stapor³

¹Department of Psychology, Humboldt-University at Berlin

²Department of Psychology, Zhejiang Normal University

³Department of Applied Informatics, Silesian University of Technology

⁴Department of Graphics, Silesian University of Technology, Computer Vision and Digital Systems

⁵Department of Psychology and Research Center Neurosensory Science, Carl von Ossietzky Universität Oldenburg

January 19, 2023

Explicit Face Memory Abilities are positively related to the non-intentional Encoding of Faces: Behavioral and ERP Evidence.

Werner Sommer^{1,2}, Krzysztof Kotowski³, Yang Shi², Adam Switonski⁴, Andrea Hildebrandt⁵,
Katarzyna Stapor³

¹ Humboldt-University at Berlin, Department of Psychology, Germany

² Department of Psychology, Zhejiang Normal University, Jin Hua, China

³ Silesian University of Technology, Department of Applied Informatics, Gliwice, Poland

⁴ Silesian University of Technology, Department of Graphics, Computer Vision and Digital Systems, Gliwice, Poland

⁵ Department of Psychology and Research Center Neurosensory Science, Carl von Ossietzky Universität Oldenburg, Germany

Correspondence should be addressed to:

werner.sommer@cms.hu-berlin.de or katarzyna.stapor@polsl.pl

Abstract

Individual differences in face memory abilities have been shown to be related to individual differences in brain activity. The present study investigated brain-behavior relationships for the N250 component in event-related brain potentials, which is taken as a neural sign of face familiarity. We used a task in which a designated, typical target face and several (high- and low-distinctive) nontarget faces had to be distinguished during multiple presentations across a session. Separately, face memory/recognition abilities were measured with easy versus difficult tasks. We replicated an increase of the N250 amplitude to the target face across the session and observed a similar increase for the non-target faces, indicating the build-up of memory representations also for these faces. On the interindividual level, larger across-session N250 amplitude increases to low-distinctive non-target faces were related to faster face recognition as measured in an easy task. These findings extend the present knowledge about brain-behavior relationships in face memory/recognition and indicate that an advantage in non-intentional encoding of low-distinctive non-target faces into memory goes along with the swift recognition of explicitly learned faces.

1. Introduction

Individuals differ greatly in their face memory/recognition abilities (e.g., Wilhelm et al., 2010). For example, whereas super-recognizers (Russell et al., 2009) remember a vast number of faces, individuals with developmental or acquired prosopagnosia may have a hard time recognizing even the faces of their family members (Behrmann & Avidan, 2005). Although the general outlines of the functional neuroanatomy of face processing are understood (Haxby et al., 2000), research on the neural correlates of individual differences in face processing is still scarce. In the present study, we explored whether the amplitude of the N250 component in the event-related potential (ERP) is related to individual differences in face recognition and whether this relationship depends on the distinctiveness of the faces to be memorized.

In their influential face processing model describing its functional neuroanatomy, Haxby et al. (2000) distinguished between a core system and an extended system of facial information processing. The core system includes the fusiform face area (FFA), the occipital face area (OFA), and the posterior superior temporal sulcus. This core system is held to process facial features, face identity, and facial expressions, respectively. The extended system includes brain regions that encode biographical and semantic information gleaned from the face. In the ERP, various components in response to faces have been described. Best known is the N170, an occipito-temporal negativity peaking around 170 ms after stimulus onset, which is larger to faces than to most common objects and has therefore been interpreted as reflecting structural encoding of faces. Following the N170 and peaking between 200 to 300 ms after stimulus onset a repetition effect can be observed, consisting in a temporo-occipital negativity to repeated relative to non-repeated famous or personally familiar faces. This N250r has been related to a transient activation of facial representations for recognition (Schweinberger & Burton, 2003). Immediately thereafter, starting around 300 ms, a second repetition effect for faces has been described (e.g., Schweinberger et al., 1995). This has been related to the activation of

multimodal biographical and semantic person representation (for a review see Schweinberger & Neumann, 2016).

Tanaka et al. (2006) investigated the N250 component in the context of long-term repetition. A specific face was designated as a target and had to be distinguished from several non-target faces. The target face was repeatedly shown multiple times but widely spaced and interspersed between other (non-target) faces. The N250 amplitude to the target face increased from the first to the second half of the experimental session, indicative of the build-up of face representations. The findings of Tanaka et al. (2006) were replicated by Sommer et al. (2021). Gosling and Eimer (2011) reported that the N250 can distinguish already familiar faces from unfamiliar faces. Since the N250 increases after repeated presentations of a face also when different images are used across presentations (Andrews et al., 2017; Kaufmann et al., 2009), the evidence suggests that the N250 reflects facial representations at a certain level of abstraction (Wiese et al., 2021). The neural generators of the N250 and N250r have been reported to be in or near the fusiform gyrus (Kaufmann et al., 2009; Schweinberger et al., 2002; Schweinberger et al., 2007).

For assessing the neural correlates of face processing-related abilities, knowledge about the psychometric structure of these abilities at the behavioral level becomes crucial. Fairly comprehensive psychometric work based on multivariate test batteries applied to large participant samples and analyzed by confirmatory factor analyses has shown that face processing-related abilities are robustly separable from general intelligence and are multifaceted (Hildebrandt et al., 2010; Wilhelm et al., 2010). In difficult (accuracy) tasks, face perception and face memory abilities can be dissociated, but in easy (speed) tasks, there is only a single ability of face recognition speed. The separation of face perception and face memory has been also applied to the single-task Cambridge face memory test (Duchaine & Nakayama, 2006) versus the Cambridge face perception test (Duchaine et al., 2007).

Given the structure of face processing-related abilities derived from psychometric work, research on the neural correlates of individual differences in domain specific face processing abilities has recorded ERPs in independent testing sessions for EEG recordings versus psychometric measurements. Thus, in two independent large samples, Herzmann et al. (2010) and Kaltwasser et al. (2014) showed that both the accuracy of face perception and memory (in difficult tasks) and the speed of face processing (in easy tasks) are negatively associated with the latency of the N170 component and with the amplitude of the N250r. In other words, the shorter the N170 latency of an individual and the larger the priming effect in the N250r, the better are the mentioned face processing abilities. The negative association between the N170 latency and the accuracy of face perception and memory was further confirmed by Nowparast Rostami et al. (2017). These ERP studies were concerned with explicit memory tasks where individuals attempted to intentionally memorize and recognize faces during the EEG session as well as in most of the tasks administered in the psychometric test session.

Several studies with somewhat heterogeneous approaches and results were designed to test associations between face processing performance and fMRI data. Furl et al. (2011) compared a group of participants with developmental prosopagnosia and normal controls (total $n = 35$). Performance differences in a test battery of face identification were mirrored in the face selectivity of the fusiform gyrus during an incidental task; there was no relationship with repetition suppression (often likened to priming effects) in fMRI. Huang et al. (2014) used a localizer task in fMRI with passive viewing in two large samples. In both samples, the face selectivity of fMRI activation in both FFA and OFA were positively correlated with the accuracy of recognizing previously learned faces (controlling for the recognition performance for objects). Li et al. (2017) showed that face selectivity in the right and left FFA correlated with two different aspects of holistic face perception. Ramot et al. (2019) investigated resting state connectivity and related it to performance in the Cambridge Face Memory Test. They

found no relationship for the connectivity with the face network but with connectivity between the face network and hubs outside of this network.

Altogether, ERP and fMRI studies show brain-behavior relationships. For ERPs, the most consistent findings concern the N170 and N250r. Considering that both the N170 and the N250r have been suggested to be generated in the fusiform gyrus, these findings are largely in line with the reports that fusiform face selectivity in fMRI is predictive of individual differences in face memory/recognition.

An important variable in face processing tasks is the typicality or distinctiveness of a face relative to the average of the faces experienced by a person. This average face is viewed as a prototype in a multidimensional face space in which any known or perceived individual face can be located (e.g., Leopold et al., 2001; Valentine, 1991; Valentine et al., 2016). If a face is located close to the prototype it is “typical” and if it is far away from the prototype it is referred to as an untypical or distinctive face. Distinctive faces are usually easier to perceive and recognize than typical faces (e.g., Bartlett et al., 1984; Sommer et al., 1995; Valentine & Endo, 1992).

1.1. Aims of the present study

Our overarching aim was to investigate how changes of the N250 component in ERPs to faces across repeated exposures are associated with face memory/recognition abilities on the individual differences level. More specifically, we distinguished between an easy versus a difficult test of face memory, following the distinction established by Wilhelm et al. (2010). Due to limit testing time, we measured face memory/recognition on a manifest rather than a latent level. ERPs, in particular the N250, were derived in a paradigm first employed by Tanaka and colleagues (2006). In this paradigm, a designated target face is presented, randomly mixed with several initially unfamiliar nontarget faces and the own face of the participant. As a

novelty, the nontarget faces in the present study consisted of two sets of high- and low-distinctive unfamiliar faces, as determined by an independent sample of raters.

On the group level, we tested the N250 amplitudes for changes from the first to the second half of the EEG session, expecting to replicate the increase for target faces reported by Tanaka et al. (2006) and Sommer et al. (2021). We also tested this change across session halves in the N250 to the non-target faces and differentiated between high- and low-distinctive faces. On the individual differences level, we correlated the ERP differences between the first and second half of the session for the N250 to target faces, and to high- and low-distinctive non-targets faces with the performance estimates from the two memory tests. We expected more negative N250 amplitudes in the second half of the experiment to go along with better face memory/recognition. It was of special interest to study how this relationship depends on the target-status and the distinctiveness of the N250-eliciting stimuli and on the difficulty of the memory test.

2. Methods

The study consisted of a single session with two parts, a psychometric part, employing a test from the test battery of Hildebrandt et al. (2010) in a speed and an accuracy version and an EEG part modelled after the study of Tanaka et al. (2006). We always conducted the psychometric part first, followed by a break of 40 minutes before the EEG part started.

2.1. Participants

Participants were 40 female students of the Zhejiang Normal University, Jinhua, Zhejiang Province, China. Participants reported normal or corrected-to-normal vision. Because one participant had to be excluded from the data analysis due to noisy EEG, the final sample consisted of 39 women ($M_{\text{age}} = 20.41 \pm 1.30$ years; range: 18 – 23 years). The experimental protocol was approved by the ethical committee of Zhejiang Normal University.

2.2. Psychometrics

2.2.1. Materials

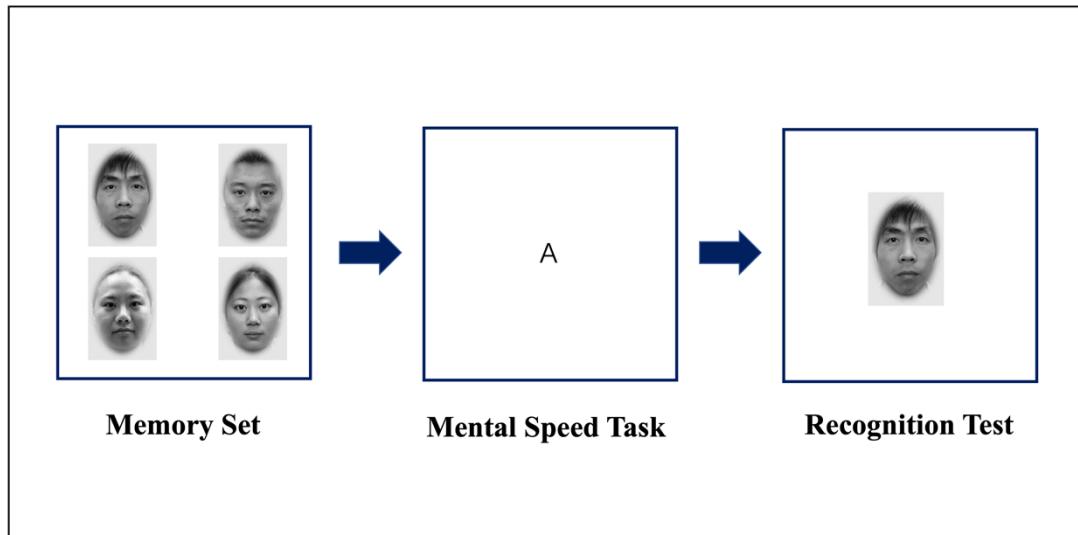
The face stimuli for the psychometric test were taken from Cepulic et al. (2018). They were grayscale frontal images of Caucasian face models taken from Hildebrandt et al. (2010) and Asian models taken from Gao et al. (2008) (cf. Fig.1). All images were fitted within an ellipse of 200×300 pixels and showed no external features or non-face cues, such as clothing, glasses, beards, hair, or piercings.

2.2.2. Procedure

The psychometric test procedure (Fig. 1), lasting about 30 min, was adapted from Cepulic et al. (2018) and conducted individually before the EEG experiment. After the overall test instruction, six practice trials with correctness feedback were provided. The tasks were designed for low versus high difficulty levels; either 25 or 40 s were available to memorize a matrix of either four or 15 target faces, respectively. All target faces in a given matrix were either all Asian or all Caucasian. Each of these four conditions was repeated once with different faces. The faces in these matrices served as target faces in the recognition phase of the task, which was conducted after an intermediate mental speed task aiming to hamper contributions of working memory to face recognition performance.

Figure 1.

Example of an easy condition in the face memory task with Asian faces. After memorizing a matrix with four previously unfamiliar faces and an intervening mental speed task, the memorized faces were presented one by one, in random order, and intermixed with novel distractor faces, requiring an old/new decision by pressing one of two buttons.



In the mental speed task, taking 1-1.5 min, one of four different 2-choice decisions to visual stimuli presented at the center of the screen were required by means of button presses about whether or not (1) a letter out of eight possible letters (A Q Z T S P R W) was an “A”, (2) a one-digit number out of eight was “3”, (3) a line-drawn face was smiling (out of smiling, neutral and sad expression), and (4) whether two symbols, letters or numbers shown side by side were identical.

In the recognition phase of the face memory task, single faces were presented that were to be classified as quickly and accurately as possible by choice-response button presses (without feedback) as to whether or not they had been present in the study matrix. In easy tasks, each of the four target faces was presented four times, randomly mixed with 16 unknown same-race distractor faces (32 trials per run). In difficult tasks, 15 target faces were mixed with 15 new same-race distractor faces (30 trials per run). The task order was fixed for all participants, as follows: (1) Asian, easy, Run 1, (2) Asian, easy, Run 2, (3) Asian, hard, Run 1, (4) Asian, hard, Run 2, (5) Caucasian, easy, Run 1, (6) Caucasian, easy, Run 2, (7) Caucasian, hard, Run 1, and (8) Caucasian, hard, Run 2. None of the faces shown in the matrices of the learning phases was

used in any of the other matrices and none of the distractors in the recognition phase was used twice.

2.2.3. Scoring and Data Treatment

The data treatment followed the procedure of Cepulic et al. (2018) and was conducted via a Python script, using Numpy and Pandas packages. For obtaining accuracy data, we excluded fast guesses with reaction times (RT) < 200 ms and calculated the average accuracy for each block. For calculating RT data, we excluded incorrect trials (15.04% of trials in the whole dataset). Then, the data was winsorized (i.e., Tukey correction; Tukey, 1977) at the within-person level, that is, for each block, all RTs longer than the third quartile plus 1.5 times the interquartile range were set to this limit value (7.9% of trials in the whole dataset). Based on the winsorized data, average RTs for each condition in each participant were calculated. We also inspected the averaged data across all conditions for interindividual outliers based on the same criterion; no participants fell outside of this limit. Finally, the mean RTs per condition and participant were transformed by multiplying the reciprocal of the average RT (ms) by 1000 ($1000/\text{RT}$) to normalize the distribution. The transformed RT data reflect the number of correct responses per second, that is, larger values indicate greater speed.

2.3. EEG Session

2.3.1. Materials

The stimuli for the EEG experiment were taken with permission from the face database of Xi'an Jiaotong University (<http://www.aiar.xjtu.edu.cn/info/1015/1639.htm>). The faces were selected based on an initial distinctiveness rating. A total of 35 female face pictures were rated by 21 female students of Zhejiang Normal University, different from the participants of the EEG experiment. The instructions for the distinctiveness ratings were: "Imagine that you are in a dining hall, and this person is in the crowd. How hard is it for you to notice the person?" Ratings were on a 5-point scale from (1) very easy to (5) very hard. In addition, the faces were rated on

attractiveness (“In your opinion, how attractive is this face?”) on a 5-point scale from (1) very attractive to (5) very unattractive, and emotionality of facial expression (“In your opinion, what expression does this face show?”), also on a 5-point scale from (1) very negative to (5) very positive.

On the basis of these ratings, we selected a low-distinctive face as the target face; this was done to increase the difficulty of target face recognition. According to these ratings the target face was hard to notice, $M = 3.90$ ($SD = 0.97$), attractiveness was average, $M = 2.33$ ($SD = 0.89$), and emotionality was slightly positive, $M = 3.14$ ($SD = 0.47$).

From the other faces, we selected a set of 6 easy-to-notice (high-distinctive) faces and 4 hard-to-notice (low-distinctive, typical) faces. High- and low-distinctive faces differed in distinctiveness as intended, $M = 2.80$ ($SD = 0.98$) vs. $M = 3.99$ ($SD = 0.85$), facial expressions were rated as more positive in more distinctive than in more typical faces, $M = 3.48$ ($SD = 0.69$) vs. $M = 2.63$ ($SD = 0.69$), and attractiveness was rated lower in more distinctive than in more typical faces, $M = 2.90$ ($SD = 0.95$) vs. $M = 2.13$ ($SD = 0.87$). According to a permutation test, the differences in all three dimensions were significant, $p = .005$, $.019$, and $.007$, respectively. The confound of distinctiveness with attractiveness was to be expected because faces close to the average of faces are usually considered more attractive and it is also not surprising that more attractive faces are perceived as showing a more positive expression.

2.3.2. Apparatus and EEG Recordings

The EEG experiment was conducted in an electrically-shielded, quiet, and dimly-lit room. Participants were seated about 60 cm away from the computer monitor. The experimental procedure was implemented in E-Prime 3.0 software (Psychology Software Tools, USA). Using conductive gel, the EEG signal was recorded from 64 active electrodes, placed according to the 10-10 system in a BrainProducts actiCap; impedances were kept below 10 k Ω . The electrodes were connected to a BrainProducts actiChamp amplifier; sampling rate was 1 kHz; no online

filtering or active noise canceling was applied. Online reference was the Cz electrode, and bipolar recording of AFz versus the outer canthus of the left eye served as EOG channel.

2.3.3. EEG preprocessing

The EEG preprocessing was based on the procedures described in Kotowski et al. (2023). The code was written in Python 3.9.7 using the MNE 0.24.0 package (Gramfort et al., 2013) and is publicly available at (after acceptance). In short, any manually identified corrupted channel (various individual channels in 17 participants) were interpolated using spherical splines (Perrin et al., 1989). The EEG channels were recalculated to common-average reference and band-pass filtered between 2 and 40 Hz using a third-order zero-phase forward-backward digital Butterworth filter (Gustafsson, 1996). ERP epochs between -100 to 500 ms relative to stimulus onset in correct trials were extracted. Ocular artifacts were corrected by removing independent components (ICs) correlated with eye blinks in EOG signal by more than 2 *SDs* higher than others. We manually removed some additional artifactual ICs for 16 participants (on average 1.4 ICs per participant, as reported in the code provided above) that were not detected by the algorithm due to weak EOG signals or due to artifacts of non-ocular origin. Finally, the ERPs were corrected with respect to the 100 ms pre-stimulus baseline.

2.3.4. ERP Amplitude Analysis

The ERP component of central interest was the N250. In order to quantify this component, ERP epochs were averaged separately for the first and second half of the EEG session and per condition. Conditions were the target face and the non-target faces divided into high- versus low-distinctive ones. Since the own face ERPs were not relevant for the present questions, we will not report them here. The N250 component was measured as in Sommer et al. (2021) but limited to the subset of electrodes available, that is, to the averaged channels TP9, P7, PO7, O1, TP10, P8, PO8, and O2, in which the N250 amplitude was measured between 230 and 320 ms.

2.4. Permutation test for correlation coefficients

The classic approach to multiple hypothesis testing, for example, the Bonferroni correction, is very conservative in some cases, increasing the risk of Type 2 errors. In order to adequately test the six correlations for which we formulated hypotheses, we applied a simultaneous permutation test as described in the following. Let us denote X_1 (N250 targets), X_2 (N250 low-distinct non-targets), X_3 (N250 high-distinct non-targets), X_4 (recognition accuracy), X_5 (recognition speed). We were interested in six correlation coefficients ρ_i $i = 1, \dots, 6$ (between each of the three variables X_1, X_2, X_3 and each of the two variables: X_4, X_5). The problem can be addressed by verifying the hypothesis:

$$H_0: \max_i |\rho_i| = 0$$

$$H_1: \max_i |\rho_i| > 0$$

Hence, we took the maximum from the (absolute) values of the six mentioned correlation coefficients as a test statistic:

$$T = \max_i |r_i| \quad i = 1, \dots, 6$$

where r_i is a sample Pearson correlation coefficient. To perform the permutation test we calculated the value T_0 of statistics for our tested sample from the experiment. Assuming that the null hypothesis is true, we performed $N = 10,000$ random permutations (shuffles) of the data in each of the five variables, calculated the value of the test statistic for each permutation and created the empirical distribution of $T_j, j = 1, 2, \dots, N$.

If the calculated ASL (Achieved Significance Level) (Good, 1994):

$$ASL \approx \frac{\text{card}\{T_i \geq T_0\}}{N}$$

is less than the significance level $p = 0.05$, the null hypothesis is rejected. Please note that apart from being less conservative than Bonferroni correction, the permutation test of statistical significance has the advantage of not assuming normal distribution of the data.

3. Results

3.1. Psychometric Test

For the 39 participants with ERP data, we merged the psychometric results from all blocks and across Asian and Caucasian faces. In the difficult task, accuracy was $M = 74.23\%$ correct ($SD = 7.55$; range 60.00 – 92.50%) and mean inverted RTs (1/s) were $M = 0.90$ ($SD = 0.18$; range: 0.56 – 1.18). In the easy task, accuracy was $M = 95.17\%$ correct ($SD = 4.49$; range 80 – 100%) and mean inverted RTs (1/s) were $M = 1.14$ ($SD = 0.17$; range: 0.68 – 1.46). Paired t -tests revealed that in the easy task accuracy was significantly higher than in the hard task ($t_{(61.89)} = 14.88, p < 0.001$); similarly, inverted RTs showed higher speed in the easy task ($t_{(38)} = 11.17, p < 0.001$). These results demonstrate that the two tasks differ in accuracy and response speed as intended, yielding easy versus difficult versions indeed.

In order to assess the reliability of the psychometric tasks, we calculated the correlations of the performance between Run 1 and Run 2 (after averaging across Asian and Caucasian faces) for each test versions. For the easy task, inverted RTs in Run 1 and Run 2 were correlated $r = 0.820$, corresponding to a split-half reliability of 0.901 after Spearman-Brown correction. For the difficult tasks, accuracy in Run 1 and Run 2 was correlated $r = 0.541$, corresponding to a split-half reliability of 0.702 after Spearman-Brown correction. These results are similar as those of Cepulic et al. (2018; online supplement), where the inverted RTs in the two runs in easy Caucasian and Asian faces correlated $r = 0.71$ and $r = 0.56$, respectively, and for the accuracy in the hard versions the correlations were $r = 0.59$ and $r = 0.58$, respectively. Moreover, the present results indicate excellent reliability for the easy task and sufficient reliability for psychometric purposes for the difficult task.

Table 1

Sample average results of the psychometric tasks

Condition	M accuracy	M inverted RTs
Asian Easy	95.39% $\pm$ 5.38%	1.17 $\pm$ 0.18
Caucasian Easy	94.95% $\pm$ 6.07%	1.10 $\pm$ 0.19
Asian Hard	73.16% $\pm$ 7.64%	0.92 $\pm$ 0.21
Caucasian Hard	75.30% $\pm$ 9.08%	0.88 $\pm$ 0.17
Easy (Average)	95.17% $\pm$ 4.49%	1.14 $\pm$ 0.17
Hard (Average)	74.23% $\pm$ 7.55%	0.90 $\pm$ 0.18

3.2. ERP Test

Accuracy of recognizing the target face was very good with $M = 95.69\%$ correct responses ($SD = 3.72$; range 86 – 100).

3.2.1. Group level analyses

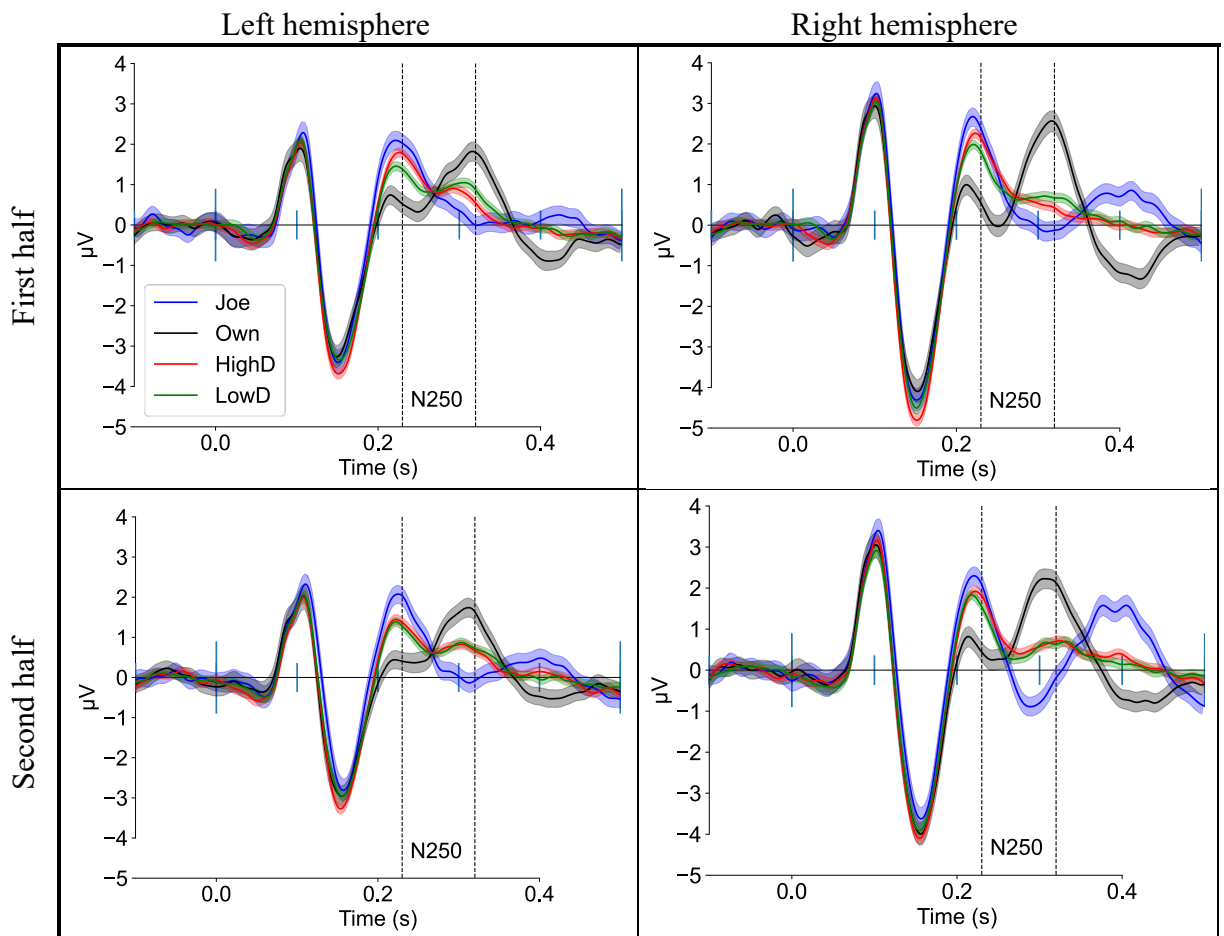
We first aimed to replicate that the mean N250 amplitude to target trials increases from the first to the second part of the experiment as reported by Tanaka et al. (2006) and Sommer et al. (2021). This was indeed the case: N250 amplitude became significantly more negative from the first half of the session, $M = 0.81$, $SD = 1.16 \mu V$, to the second half, $M = 0.38$, $SD = 0.98 \mu V$, as confirmed by a one-sided Wilcoxon signed-rank test for H_1 alternative ($p = 0.0007$). ERP waveforms were obtained by averaging over all electrodes within each ROI, per condition and each half of the experiment (see Fig. 2).

For the N250 components to the non-target faces, we asked whether these responses changed across the experiment and how this depends on face distinctiveness. A repeated measures ANOVA indicated a significant amplitude change towards more negative (less positive) values from the first to the second half of the experiment, $M = 1.03 \mu V$, $SD = 1.03$ vs. $M = 0.77 \mu V$, $SD = 0.71$, $F(1,38) = 7.52$, $p < .05$, $\eta^2 = .17$. In contrast to the P200, there

was no main effect of distinctiveness in the N250, $F(1,38) = 1.70$, $p > .05$, $\eta^2 = .043$, nor did it interact with experiment half, $F(1,38) = .774$, $p > .05$, $\eta^2 = .020$.

Figure 2

ERPs in the N250 ROI separated for hemisphere and the first and second half of the experiment. Superimposed are the responses to the target face, the high- and low-distinctive non-target faces, and the participant's own face.



To estimate the reliability of the N250, we correlated the amplitudes of the first and second half of the experiment. For the target face-elicited N250 this correlation was $r = 0.748$, yielding good reliability after Spearman-Brown correction (0.855). For high-distinctive non-target faces the correlation was $r = 0.876$ and for low-distinctive non-target faces it was $r =$

0.864, yielding excellent Spearman-Brown corrected reliabilities of 0.934 and 0.927, respectively.

3.2.2. Brain-behavior relationships

In the next step, we correlated N250 amplitude differences between the first and second part of the experiment with the performance in the easy versus difficult face memory tests. As common in psychometrics, we used accuracy as a memory measure in the difficult task and the inverted RTs as a performance indicator in the easy tasks.

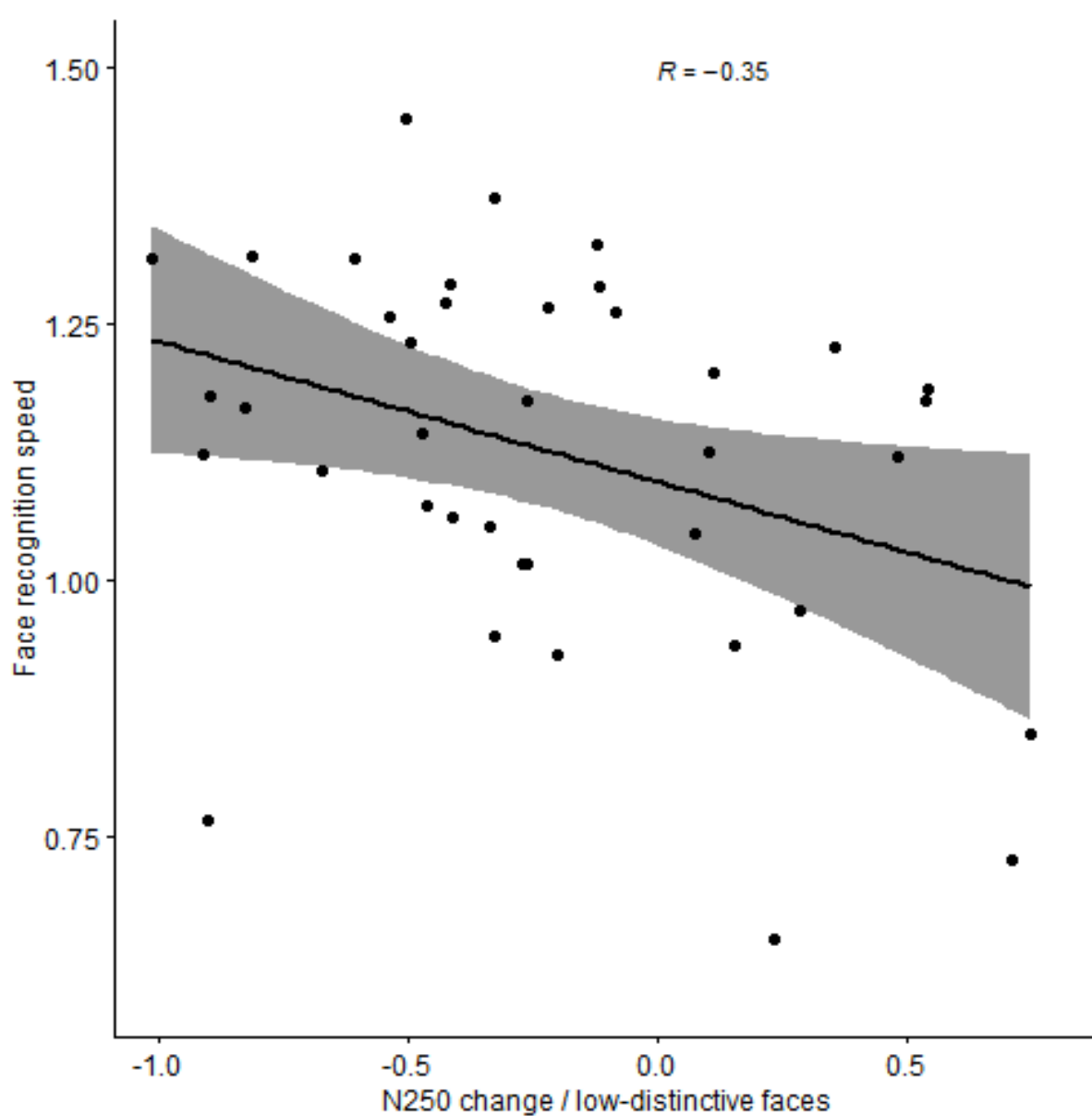
For the N250 amplitude differences to target faces the correlations were small for both the easy versus the difficult memory tasks, $r = -0.035$ and $r = -0.070$, respectively. For the N250 in response to non-target faces, we calculated amplitude differences between experimental halves separately for high- and low-distinctive faces. The correlations with recognition performance of high-distinctive faces with easy versus difficult task performance were $r = -0.201$ and $r = -0.181$, respectively. The corresponding correlations for low-distinctive faces were $r = -0.355$ and $r = -0.163$, respectively.

According to the permutation test described in Section 2.4, we can reject the null hypothesis with a probability of 5% that all six considered pairs of variables are independent. Hence, for at least one pair of variables (i.e., the largest N250 low-distinct non-targets X2 vs. memory speed X5) there is a statistically significant linear relationship. The same permutation test was used to test the H1 that the second largest correlation coefficient (for the modulus) is greater than zero. The corresponding null hypothesis could not be rejected even at a level of 0.1. Thus, there is only one significant correlation, which is between the change in N250 amplitude to low-distinctive non-target faces and the psychometric face recognition speed ($r = -0.355$). The scatter plot of corresponding data of this distribution is shown in Figure 3. Of note, there are at least four data points where performance in the face recognition test was rather slow. Although, upon closer inspection these participants did not show any unusual behavior,

we excluded these four participants to assess the stability of the association. As a consequence, this particular correlation dropped to $r = -.25$ and the permutation test of all 6 correlations of interest failed significance.

Figure 3

Association of the change in N250 amplitude to low-distinctive non-target faces from the first to the second half of the session and the psychometric face recognition speed.



4. Discussion

The main aim of the present study was to investigate the relationship between the N250 component and face recognition performance on the individual differences level by applying the paradigm of Tanaka et al. (2006), where participants had to identify a single target face among several non-target faces. We tested this relationship separately for target faces and non-target faces, while distinguishing – within non-target faces – between high and low distinctive faces. For target faces, we replicated the increase of N250 amplitude across the first and second half of the experiment, but did not find a relationship to recognition performance. The N250 to non-target faces showed – in contrast to previous reports – an increase across experiment parts. Importantly, the increase of the N250 to low-distinctive non-target faces was correlated with memory performance in the easy task version.

4.1. Psychometric measurements

We should first point out that the psychometric task applied to Chinese participants successfully yielded an easy and a difficult version, markedly differing in the percentage of correct responses (95 vs. 74% correct). Therefore, the test, albeit a single task in each version fulfills the criterion of a speed test, where the variance is mainly in RTs, and a difficult version, where individuals differ in the proportion of correct responses. The relative independence of the speed and accuracy tasks is reflected in their low correlation ($r = 0.36$). Importantly, both task versions had shown sufficient to excellent reliability.

4.2. ERPs

Extending reports of Tanaka et al. (2006) and Sommer et al. (2021), the N250 to target faces was larger in the second than in the first part of the experiment. The N250 has been related to facial representations at a certain level of abstraction (Wiese et al., 2021) and the increase across the experiment likely reflects the increasing build-up of these representations. Since our stimulus materials consisted of Asian faces and our participants were Chinese students, the results indicate the culture independence of this increasing face representations.

Interestingly, also the non-target elicited N250 increased from the first to the second half of the experiment. The non-target faces did not have to be explicitly identified at an individual level but only had to be rejected as not being the target face. These results go beyond findings of both the studies of Tanaka et al. (2006) and Sommer et al. (2021), which did not find such effects in the N250 amplitude to non-target faces. We presume that a crucial factor in bringing out the N250 effects to non-targets is the nature of these faces. In contrast to the previous studies, we had controlled the stimulus materials for distinctiveness and can therefore claim that the target face was of low-distinctiveness. Low-distinctive target faces are likely harder to recognize than high-distinctive target faces (e.g., Light et al., 1979). Therefore, it is possible that in order to detect the target face, our participants were forced to scrutinize the non-target faces more than the participants of the previous studies and therefore encoded them incidentally. Attention to a stimulus is a crucial factor in memory encoding.

4.3. Brain-Behavior Relationships

Contrary to our expectations, at the inter-individual level the amount of change of the target-elicited N250 from the first to the second experiment half did not correlate with memory performance. Since these correlations were essentially zero, we suspect that this null result is not due to a lack of power or a low signal-to-noise ratio, as indicated by the high split-half reliabilities of the target-elicited N250 and the good to excellent reliability of the face memory test results. Rather it seems that the robust increase in N250 amplitude across the experiment, although possibly reflecting increasing familiarity with the designated target face, is not related to individual differences in face recognition. One conceivable reason is the ease of recognizing the target face, leading to a ceiling effect.

The relationship of the across-session changes of the N250 to non-target faces and memory performance was significant only between N250 to low-distinctive faces and performance in the easy but not in the difficult memory task. Wilhelm et al. (2010) and

Hildebrandt et al. (2010) have demonstrated in large samples that face cognition in easy and difficult tasks must be distinguished, similar to the need for such a distinction when measuring general cognitive abilities. Therefore, it is an interesting finding that also the brain-behavior relationship seems to be specific for the kind of memory task – the association with the easy task was double in effect size than with the hard task. The simplest explanation for the unverifiable association between the across-session change of the N250 and performance in the difficult memory test is the lower reliability of the accuracy than of the recognition speed test. However, since our sample size was only modest and memory task performance was measured on the manifest level, it remains important to replicate these findings with a larger sample and to investigate a set of indicators for testing these associations at the latent level.

Interestingly, the brain-behavior relations of the easy task performance were found for across-session N250-changes in ERPs to low-distinctive rather than to high-distinctive faces. Intuitively, one might expect a stronger relationship between the neural signal of building-up a memory trace of high-distinctive faces and face recognition performance than for the harder to process low-distinctive faces. However, the stronger relationship found for across-session increases in low-distinctive face-elicited N250 may become plausible when considering that the ERP task required the discrimination of the low-distinctive target face from the likewise low-distinctive non-target faces. Individuals with better face recognition ability may have solved this task by a faster build-up of a representation of these non-target faces, as manifested in a larger N250 amplitude increase across the session. In contrast, memory representations for high-distinctive non-target faces may not have been necessary to make the distinction from the non-distinctive target face.

To the best of our knowledge this is the first study that shows a brain-behavior relationship between the N250 to faces and face memory. Hence, the present results extend previous findings from the N250r (Herzmann et al., 2010; Kaltwasser et al. 2014) to a condition

where the N250 amplitude evolves across multiple presentations. Since the difference of the N250 across experimental halves can also be considered a kind of repetition effect, it would be of interest, whether the difference between experimental halves in the N250 is related to the N250r. This would be of particular interest, considering the close conceptual relations of these components.

4.4. Limitations

In the present study, following the original design of Tanaka et al. (2006) we have used the same image for all individuals serving as non-target faces and also the target face image was always the same. Hence, the present study may be criticized as tapping into (face) image recognition rather than face recognition, which is characterized by high variability of the images associated by the same person (Burton et al. 2016). Although a recent publication by Olderbak et al. (2022) has demonstrated that the psychometric structure of face cognition does not depend on the usage of the same or different images, one of the future research tasks should be a replication with different images of the same individuals. In this research, larger samples, more elaborated test batteries and explicit tasks for ERP recordings might be used to follow up the present findings.

Further limitations are the small number of participants, which make the association observed susceptible to outliers (see Fig. 3), and the restriction to the manifest levels of single tasks. Future studies should extend the present approach to a larger sample and to the latent level by including multiple indicators on the brain and behavioral sides.

4.5. Conclusions

In conclusion, the present study showed that the within-experiment increase of the N250 component of the ERP to incidentally encoded low-distinct non-target faces is related to the speed of face processing at an individual differences level. This finding – although in need of further elaboration – extends the present knowledge about brain-behavior relationships for

faces. Specifically, it indicates that individuals with better abilities in the speed of face processing in memory tasks have an advantage in familiarization (memory trace formation) for low-distinctive, hard to recognize faces.

Author Contributions

Werner Sommer: Conceptualization, Methodology, Writing. **Krzysztof Kotowski:** Data Analysis, Visualization, Software. **Yang Shi:** Experimental set up, Data Collection, Data Analysis. **Adam Switonski:** Data Analysis. **Andrea Hildebrandt:** Methodology, Editing. **Katarzyna Stapor:** Methodology, Data Analysis, Writing

Acknowledgements

We thank Xiaohua Cao for generously providing his lab for conducting this research at a time when Polish and German labs were unavailable to the Covid 19 pandemic.

References

- Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *Quarterly Journal of Experimental Psychology*, 70(8), 1620-1632. <https://doi.org/10.1080/17470218.2016.1195851>
- Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition*, 12, 219-228.
- Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences*, 9(4), 180-187. <https://doi.org/10.1016/j.tics.2005.02.011>
- Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, 40(1), 202-223. <https://doi.org/10.1111/cogs.12231>
- Cepulic, D. B., Wilhelm, O., Sommer, W., & Hildebrandt, A. (2018). All categories are equal, but some categories are more equal than others: The psychometric structure of object and face cognition. *Journal of Experimental Psychology-Learning Memory and Cognition*, 44(8), 1254-1268. <https://doi.org/10.1037/xlm0000511>
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576-585. <https://doi.org/10.1016/j.neuropsychologia.2005.07.001>
- Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419-430. <https://doi.org/10.1080/0264329070138049>

- Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform Gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience*, 23(7), 1723-1740. <https://doi.org/10.1162/jocn.2010.21545>
- Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1), 149-161. <https://doi.org/10.1109/tsmca.2007.909557>
- Good P. I. (1994). *Permutation tests: A practical guide for testing hypotheses*. Springer, New York
- Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia*, 49(9), 2736-2745. <https://doi.org/10.1016/j.neuropsychologia.2011.05.025>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267. <https://doi.org/10.3389/fnins.2013.00267>
- Gustafsson, F. (1996). Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing* 44, 988–992. <https://doi.org/10.1109/78.492552>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Herzmann, G., Kunina, O., Sommer, W., & Wilhelm, O. (2010). Individual Differences in Face Cognition: Brain-Behavior Relationships. *Journal of Cognitive Neuroscience*, 22(3), 571-589. <https://doi.org/10.1162/jocn.2009.21249>

- Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, 25(4), 794-810. <https://doi.org/10.1037/a0019774>
- Huang, L., Song, Y., Li, J., Zhen, Z., Yang, Z., & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Frontiers in Human Neuroscience*, 8, 483. <https://doi.org/10.3389/fnhum.2014.00483>
- Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: A replication and extension. *Cognitive, Affective, & Behavioral Neuroscience*, 14, 861-878. <https://doi.org/10.3758/s13415-013-0234-y>
- Kaufmann, J. M., Schweinberger, S. R., & Burton, A. M. (2009). N250 ERP Correlates of the Acquisition of Face Representations across Different Images. *Journal of Cognitive Neuroscience*, 21(4), 625-641. <https://doi.org/10.1162/jocn.2009.21080s>
- Kotowski, K., Ochab, J., Stapor, K., & Sommer, W. (2023). The importance of ocular artifact removal in single-trial ERP analysis: The case of the N250 in face learning. *Biomedical Signal Processing and Control*, 79, 104115. <https://doi.org/https://doi.org/10.1016/j.bspc.2022.104115>
- Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience*, 4(1), 89-94. <https://doi.org/10.1038/82947>
- Li, J., Huang, L., Song, Y., & Liu, J. (2017). Dissociated neural basis of two behavioral hallmarks of holistic face processing: The whole-part effect and composite-face effect. *Neuropsychologia*, 102, 52-60. <https://doi.org/10.1016/j.neuropsychologia.2017.05.026>

- Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 212-228.
- Nowparast Rostami, H., Sommer, W., Zhou, C., Wilhelm, O., & Hildebrandt, A. (2017). Structural encoding processes contribute to individual differences in face and object cognition: Inferences from psychometric test performance and event-related brain potentials. *Cortex*, 95, 192-210.
<https://doi.org/http://dx.doi.org/10.1016/j.cortex.2017.08.017>
- Olderbak, S., Pollak, K., Sanchez, J. F., & Hildebrandt, A. (2022). Assessing individual differences in face recognition: No impact of task design features on the structure of this ability. *Psychological Assessment*, 34, 1172-1179.
<https://doi.org/10.1037/pas0001173>
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology* 72, 184–187. [https://doi.org/10.1016/0013-4694\(89\)90180-6](https://doi.org/10.1016/0013-4694(89)90180-6)
- Ramot, M., Walsh, C., & Martin, A. (2019). Multifaceted integration: Memory for faces is subserved by widespread connections between visual, memory, auditory, and social networks. *Journal of Neuroscience*, 39(25), 4976-4985.
<https://doi.org/10.1523/JNEUROSCI.0217-19.2019>
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252-257. <https://doi.org/10.3758/PBR.16.2.252>
- Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex*, 39, 9-30. [https://doi.org/10.1016/S0010-9452\(08\)70071-6](https://doi.org/10.1016/S0010-9452(08)70071-6)

- Schweinberger, S. R., Pfütz, E. M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology-Learning Memory and Cognition*, 21, 722-736.
<https://doi.org/10.1037//0278-7393.21.3.722>
- Schweinberger, S. R., Kaufmann, J. M., Moratti, S., Keil, A., & Burton, A. M. (2007). Brain responses to repetitions of human and animal faces, inverted faces, and objects -- An MEG study. *Brain Research*, 1184, 226-233.
- Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, 14(3), 398-409.
[https://doi.org/10.1016/S0926-6410\(02\)00142-8](https://doi.org/10.1016/S0926-6410(02)00142-8)
- Sommer, W., Heinz, A., Leuthold, H., & Matt, J. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, 23, 1-11. <https://doi.org/10.3758/BF03210552>
- Sommer, W., Stapor, K., Kończak, G., Kotowski, K., Fabian, P., Ochab, J., Bereś, A., Ślusarczyk, G. (2021). The N250 event-related potential as an index of face familiarity: a replication study. *Royal Society Open Science*, 8, 202356.
<https://doi.org/10.1098/rsos.202356>
- Tanaka, J.W., Curran, T., Porterfield, A.L., Collins, D. (2006). Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience* 18, 1488–1497.
<https://doi.org/10.1162/jocn.2006.18.9.1488>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43A, 161-204.
<https://doi.org/10.1080/14640749108400966>

- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology*, 44A, 671-703.
- Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, 69(10), 1996-2019. <https://doi.org/10.1080/17470218.2014.990392>
- Wiese, H., Anderson, D., Beierholm, U., Tüttenberg, S.C., Young, A.W., Burton, A.M. (2021). Detecting a viewer's familiarity with a face: Evidence from event-related brain potentials and classifier analyses. *Psychophysiology*, 59, e13950. <https://doi.org/10.1111/psyp.13950>
- Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces: One element of social cognition. *Journal of Personality and Social Psychology*, 99(3), 530-548. <https://doi.org/10.1037/a0019972>