

This is a repository copy of *A robust neural index of high face familiarity*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/140129/>

Version: Accepted Version

Article:

Wiese, Holger, Tüttenberg, Simone C., Ingram, Brandon T. et al. (4 more authors) (2019) A robust neural index of high face familiarity. *Psychological Science*. pp. 261-272. ISSN: 1467-9280

<https://doi.org/10.1177/0956797618813572>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Durham Research Online

Deposited in DRO:

11 October 2018

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Wiese, H. and Tüittenberg, S.C. and Ingram, B.T. and Chan, C.Y.X. and Gurbuz, Z. and Burton, A.M. and Young, A.W. (2018) 'A robust neural index of high face familiarity.', *Psychological science*. .

Further information on publisher's website:

<http://journals.sagepub.com/loi/pssa>

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Running Head: A ROBUST NEURAL INDEX OF FACE FAMILIARITY

A robust neural index of high face familiarity

Holger Wiese^{a*}, Simone C. Tüttenberg^a, Brandon T. Ingram^a, Chelsea Y.X. Chan^a, Zehra Gurbuz^a, A. Mike Burton^b, & Andrew W. Young^b

^aDepartment of Psychology, Durham University, Durham, DH1 3LE, UK

^bDepartment of Psychology, University of York, York, YO10 5DD, UK

*Corresponding author:

Dr Holger Wiese

Department of Psychology

Durham University

E007 Wolfson Building, Queen's Campus

Stockton-on-Tees TS17 6BH

United Kingdom

Phone: 0044 191 3340433

Email: holger.wiese@durham.ac.uk

Abstract

Humans are remarkably accurate at recognising familiar faces, while their ability to recognize, or even match, unfamiliar faces is much poorer. However, previous research has failed to identify neural correlates of this striking behavioural difference. Here we show a clear difference in brain potentials elicited by highly familiar faces versus unfamiliar faces. This effect starts 200 ms after stimulus onset and reaches its maximum at 400-600 ms. This Sustained Familiarity Effect is substantially larger than previous candidates for a neural familiarity marker and is detected in almost all participants, representing a reliable index of high familiarity. While its scalp distribution is consistent with a generator in the ventral visual pathway, its modulation by repetition and degree of familiarity suggests an integration of affective and visual information.

Keywords: Face Recognition, Personal Familiarity, EEG, Event-Related Potentials, N250, Sustained Familiarity Effect

A robust neural index of high face familiarity

Humans are remarkably good at recognising familiar faces, which are identified without apparent effort, under a wide variety of environmental and stimulus changes, and from never-before seen and severely degraded images (Burton, Wilson, Cowan, & Bruce, 1999). While recent research has made considerable progress at identifying the perceptual and cognitive processes underlying face recognition (Bruce & Young, 2012), previous experiments have failed to describe a neural marker that is substantial and reliable enough to capture this remarkable ability. There are well-known neural correlates of face processing in general (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Kanwisher, McDermott, & Chun, 1997), but the large behavioural differences between our recognition ability with familiar and unfamiliar faces (Jenkins, White, Van Montfort, & Burton, 2011; Jenkins & Burton, 2011) have not produced reliable corresponding differences in neural markers. This disparity may reflect two factors that have been widely ignored in neuroscientific face recognition research: within-identity image variability (Burton, Kramer, Ritchie, & Jenkins, 2016) and the level of familiarity of the presented stimuli (Ramon & Gobbini, in press).

The hallmark of face recognition is our ability to identify a well-known face from nearly any image (Burton, Jenkins, & Schweinberger, 2011). This ability is less self-evident than might be assumed, as it can be very difficult to recognise that different pictures show the same unfamiliar person (Jenkins et al., 2011). Accordingly, image-independent robust representations are necessary for familiar face recognition (Burton, Jenkins, Hancock, & White, 2005), and accessing such representations should result in a reliable neural response.

Brain imaging studies on image-independent face recognition with fMRI have shown the Fusiform Face Area (FFA) to adapt to an individual identity over a limited range of changes of viewing angle (Ewbank & Andrews, 2008), but stimuli varying on a large number of image and stimulus properties concurrently do not elicit a corresponding effect (Davies-

Thomson, Gouws, & Andrews, 2009). However, image-invariant adaptation has been shown in the medial temporal lobe (Weibert et al., 2016), consistent with findings of differential activation for famous relative to unfamiliar faces in the parahippocampal gyrus (Bar, Aminoff, & Ishai, 2008). Similarly, intracranial recordings have found hippocampal cells selectively responding to widely varying pictures of individual identities (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Moreover, the N250r event-related brain potential (ERP), reflecting facilitated access to perceptual face representations in repetition priming, is observed when different pictures of the same person are used as primes and targets (Schweinberger, Pickering, Jentsch, Burton, & Kaufmann, 2002). Previous studies have also shown that the N250 can differentiate between familiar and unfamiliar faces (Gosling & Eimer, 2011), even when multiple naturally varying images are presented (Andrews, Burton, Schweinberger, & Wiese, 2017). However, these image-independent effects in ERPs are small, which contrasts with the remarkable ease with which familiar faces are recognised in daily life.

Most previous studies presented celebrities rather than personally familiar faces, which are not only richer in their visual representations, but also elicit an affective response (Bauer, 1984; Tranel, Damasio, & Damasio, 1995) unlikely to be shared to the same extent for celebrity faces (Ramon & Gobbini, in press; Schweinberger & Burton, 2003). The few studies that did use personally familiar faces either presented pictures of moderately familiar people (e.g. university lecturers; Herzmann, Schweinberger, Sommer, & Jentsch, 2004) and/or standardised images that were repeatedly presented (Butler, Mattingley, Cunnington, & Suddendorf, 2013; Caharel, Courtay, Bernard, Lalonde, & Rebai, 2005; Gobbini, Leibenluft, Santiago, & Haxby, 2004; Keyes, Brady, Reilly, & Foxe, 2010). Both standardisation (because of the artificial nature of these stimuli) and repetition (because of habituation to specific images) arguably reduce affective responses, and minimise any

contribution from image invariance. Previous studies therefore may have substantially underestimated neural familiarity effects.

The present experiments aimed to identify neural correlates of naturalistic face recognition, reflecting our remarkable ability to deal with within-person variability. We were additionally interested in whether reliable familiarity effects would be detectable in individual participants when the above-described limitations were circumvented. Experiment 1 directly compared highly personally familiar with unfamiliar faces, and yielded a robust and reliable ERP familiarity effect. We then conducted Experiment 2 as a control to estimate the probability of finding comparable results in individual participants by chance. Finally, in Experiment 3 we tested whether *personal* familiarity was critical, and if so, whether the level of familiarity (high versus lower) would further modulate the effect.

Experiment 1

Method

Participants

We tested 20 under-/postgraduate students and staff members at Durham University (19 female, mean age = 20.7 years \pm 4.0 SD). This sample size was determined in a power analysis based on the N250 difference between famous and unfamiliar faces in Andrews et al. (2017; paired-sample t-test; $d_z = 1.03$; Power = 0.99) using G*Power (Faul, Erdfelder, Lang & Buchner, 2007). All participants received course credits or monetary compensation, were right-handed, reported normal or corrected to normal vision, and did not take central-acting medication. Participants gave written informed consent and the study was approved by the ethics committee of Durham University's Department of Psychology.

Stimuli

Participants provided 25 images of each of two personally highly familiar people (relatives, close friends, etc.) known from beyond the university. Accordingly, participants in

Experiment 1 had chosen the original images of familiar faces they viewed in cropped form in the experiment. Consent of the depicted persons was obtained via e-mail. Twenty pictures per identity were then selected by the experimenters, excluding images with low resolution. Eight pictures of butterflies were used as targets. Rectangles around the faces/butterflies were cropped from the original images, re-sized, copied into a frame of 190×285 pixels, and converted to grey-scale (see Figure 1). Images were matched for luminance.



Figure 1. a) Sample stimuli and balancing of face IDs across participants. b) Trial structure of Experiment 1. All images are published with permission of the depicted persons.

Procedure

Participants were seated in an electrically shielded chamber with their heads in a chin rest at 100 cm distance from a monitor. Participants were paired using the same set of images (e.g., ID1, ID2, ID3, ID4), such that familiar faces for Participant 1 (ID1, ID2) were unfamiliar for Participant 2, and familiar faces for Participant 2 (ID3, ID4) were unfamiliar for Participant 1 (see Figure 1). The experiment consisted of four blocks with 88 trials each. In each block, all 40 familiar face images, 40 unfamiliar face images, and 8 butterflies were presented in randomised order at a visual angle of $2.9^\circ \times 4.3^\circ$ for 1,000 ms each, followed by a fixation cross, varying between 1,500 and 2,500 ms (2,000 ms on average). The task was to press a button whenever a butterfly was presented.

EEG recording and analysis

64-channel EEG was recorded from sintered Ag/Ag-Cl electrodes (ANT Neuro, Enschede, Netherlands; DC-120 Hz, 512 Hz sampling frequency), with an electrode on the forehead as ground and Cz as recording reference. Recording sites corresponded to an extended 10-20 system. Blinks were corrected using the algorithm implemented in BESA 6.3 (Gräefeling, Germany). EEG was segmented from -200 to 1000 ms relative to stimulus onset, with the first 200 ms as baseline. Artefact rejection was carried out using an amplitude threshold of 100 μ V and a gradient criterion of 75 μ V. Remaining trials were recalculated to the common average reference and averaged according to experimental conditions. Mean amplitudes of N170 (130 – 180 ms), N250 (200-400 ms), and from 400-600 ms were analysed at occipito-temporal electrodes TP9/TP10 using repeated-measures Analyses of Variance (ANOVAs).

Results

Behavioural results are reported in the supplementary material. Familiar faces elicited substantially more negative amplitudes than unfamiliar faces (Figure 2a). Topographical maps of this familiarity effect (unfamiliar minus familiar; Figure 2b) showed spatially restricted effects at bilateral occipito-temporal channels, with maxima at electrodes TP9 and TP10. Mean difference curves with 95% CIs revealed reliable familiarity effects (with the lower CI limit larger than zero) starting approximately 200 ms after stimulus onset and lasting until the end of the analysis epoch (Figure 2c). Familiarity effects peaked at approximately 400 ms and remained at this level until approximately 600 ms (Sustained Familiarity Effect, SFE). This latter part of the familiarity effect decreased over blocks (see right part of Figure 2d).

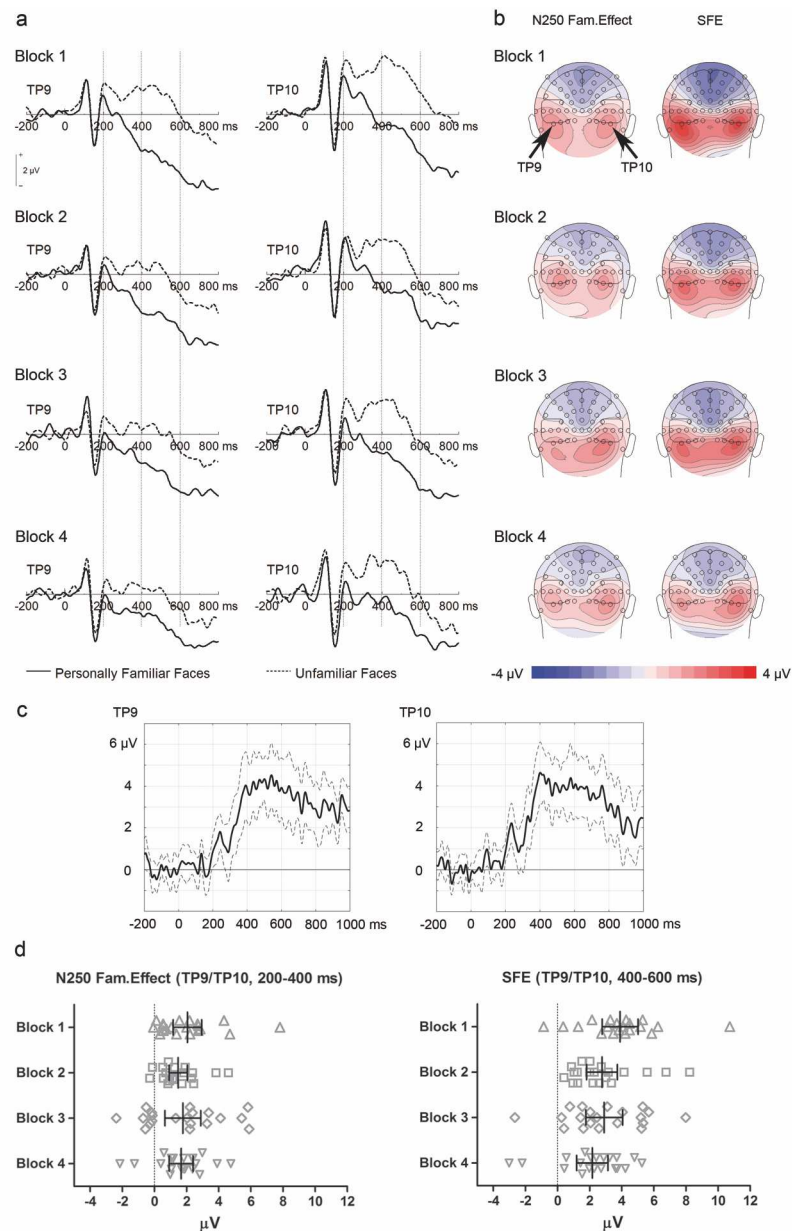


Figure 2. Results of Experiment 1. a) Grand average ERPs (N=20) at left and right occipito-temporal channels TP9 and TP10 for familiar versus unfamiliar faces in blocks 1-4. Dotted lines demark the N250 (200-400 ms) and SFE time windows (400-600 ms). b) Scalp-topographical voltage maps (spherical spline interpolation, 110° equidistant projections) for N250 and SFE (unfamiliar minus familiar faces). c) Mean difference (\pm 95% CI) for unfamiliar minus familiar faces in the first block. d) Individual and mean (\pm 95% CI) familiarity effects for the N250 and SFE time ranges.

Statistical analyses confirmed these observations. While analysis of N170 did not result in significant familiarity effects (see supplementary material), a repeated-measures ANOVA on N250 mean amplitudes with the factors familiarity (familiar, unfamiliar), block (1 - 4) and hemisphere (left, right) yielded main effects of familiarity, $F(1, 19) = 30.95, p < .001, \eta_p^2 = .620, M_{diff.} = 2.05 \mu V, 95\% CI [1.16, 2.94]$, and block, $F(3, 57) = 5.83, p = .002, \eta_p^2 = .235$. The interaction was not significant, $F(3, 57) = 0.64, p = .595, \eta_p^2 = .032$.

Analysis in the SFE time window (400-600 ms) yielded a main effect of familiarity, $F(1, 19) = 57.16, p < .001, \eta_p^2 = .751$, and an interaction of familiarity by block, $F(3, 57) = 3.76, p = .016, \eta_p^2 = .165$. Follow-up comparisons revealed familiarity effects in all blocks, block 1: $F(1, 19) = 53.13, p < .001, \eta_p^2 = .737, M_{diff.} = 3.89 \mu V, 95\% CI [2.78, 5.01]$, block 2: $F(1, 19) = 36.57, p < .001, \eta_p^2 = .658, M_{diff.} = 2.77 \mu V, 95\% CI [1.81, 3.72]$, block 3: $F(1, 19) = 28.04, p < .001, \eta_p^2 = .596, M_{diff.} = 2.90 \mu V, 95\% CI [1.76, 4.05]$, block 4: $F(1, 19) = 21.54, p < .001, \eta_p^2 = .531, M_{diff.} = 2.16 \mu V, 95\% CI [1.19, 3.14]$, with decreasing effect sizes over blocks.

We further tested the reliability of familiarity effects in individual participants in the SFE time window. We applied a bootstrapping approach (Di Nocera & Ferlazzo, 2000), using 10,000 random re-assignments of individual participants' EEG epochs to familiarity conditions. Reliable effects were assumed if the true individual familiarity effect at TP9/TP10 was larger than 95% of random re-samplings. Reliable differences between familiar and unfamiliar faces were detected in 17/20 participants, $P = .85, 95\% CI [.64, .95]$ (Figure 3a).

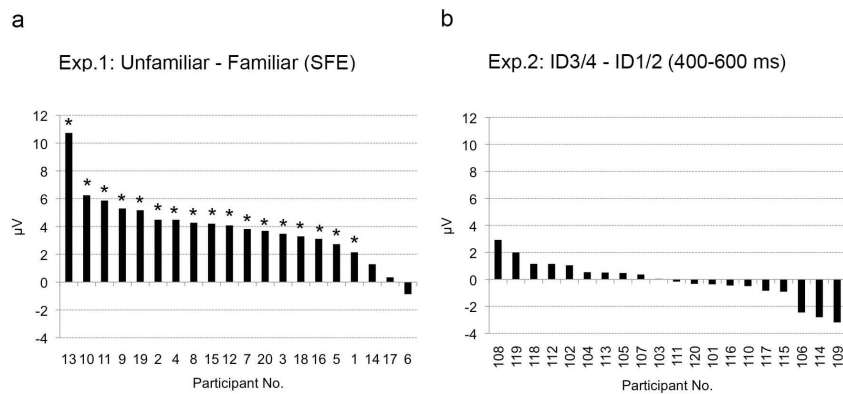


Figure 3. Results of the bootstrapping analysis of Experiments 1 (a) and 2 (b). Individual familiarity effects are averaged across the SFE at TP9/TP10. Asterisks mark reliable effects.

Experiment 2

Experiment 2 was conducted to examine how often an SFE-like effect would emerge in individual participants if all faces were unfamiliar. We therefore presented the same images as in Experiment 1 to participants who were all unfamiliar with the depicted identities and analysed potential differences between these unfamiliar sets.

Method

Participants

We tested 20 under-/postgraduate students at Durham University (14 female, mean age = 21.3 years \pm 2.8 SD). Criteria for participation and compensation were identical to Experiment 1. Participants gave written informed consent and the study was approved by the ethics committee of Durham University's Department of Psychology.

Stimuli

Stimuli were identical to Experiment 1.

Procedure

The procedure was identical to Experiment 1, except that all faces were unfamiliar. Participants were randomly assigned to a specific version of the experiment created for participants in Experiment 1, and data were analysed as if the same identities had been familiar and unfamiliar for the respective participants in Experiment 2.

EEG recording and analysis

EEG recording and analysis was analogous to Experiment 1.

Results

ERP results are depicted in Figure 4. No significant differences between identity sets were observed (see supplementary material). Bootstrapping (Figure 3b) carried out analogously to Experiment 1 revealed reliable differences between the unfamiliar ID sets in 0/20 participants, $P = 0$, 95% CI [0, .16].

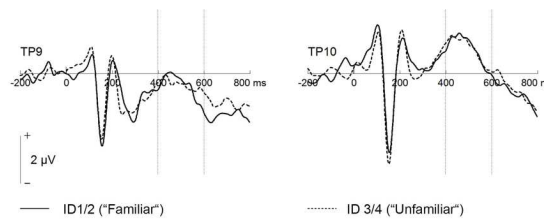


Figure 4. Grand average ERPs (N=20) from Experiment 2 at left and right-hemispheric occipito-temporal channels TP9/TP19.

Experiment 3

Although Experiment 1 generated a strong and highly reliable SFE it remained uncertain whether this was driven by *personal* familiarity, by *highly* familiar faces, or by a combination of these factors. In Experiment 3, we therefore presented participants with pictures of personally highly familiar faces, personally but less familiar faces, familiar celebrity faces (British royals), and unfamiliar faces.

Method

Participants

We tested 18 undergraduate Applied Psychology students at Durham University (11 female, mean age = 19.9 years \pm 1.7 SD). All except for two participants were right-handed, all reported normal or corrected to normal vision, and none took central-acting medication. Compensation was identical to Experiment 1. Participants gave written informed consent and the study was approved by the ethics committee of Durham University's Department of Psychology.

Stimuli

50 pictures of five personally highly familiar people (close friends from university; three female, two male), two lecturers (one female, one male) and two unfamiliar people (friends of the experimenters unknown to the participants; one female, one male), respectively, were collected from photo archives of the depicted people and the authors. Accordingly, in contrast to Experiment 1, none of the images had been selected by any participant. Fifty pictures of Prince William and Kate Middleton were collected from the internet for the celebrity condition. Staged portraits were avoided, and images were chosen to show similar variability relative to the other conditions. Eight pictures of butterflies were used as targets. Editing of the images was analogous to Experiment 1.

Procedure

Each participant was presented with 220 pictures, consisting of 50 images of four identities (personally highly familiar person, lecturer, celebrity, unfamiliar identity) and 20 pictures of butterflies. Eleven participants were tested with a female and seven with a male highly familiar identity. Identities for the other conditions were chosen so that two identities in each experiment were female and male. Stimuli were presented for 1,000 ms, with an average interstimulus interval of 1,500 ms (jittered between 1,000 and 2,000 ms) during which a fixation cross was shown. All other aspects were identical to the previous experiments.

EEG recording and analysis

EEG recording and analysis were analogous to Experiment 1.

Results

We observed strong familiarity effects for personally highly familiar, but only small effects for lecturer and celebrity faces (Figure 5a). N250 and SFE for personally highly familiar faces demonstrated a clear occipito-temporal scalp distribution (Figure 5b), and

reliable effects were observed starting at approximately 200 ms after stimulus onset (Figure 5c). Again, the familiarity effect reached its maximum between 400 and 600 ms (SFE).

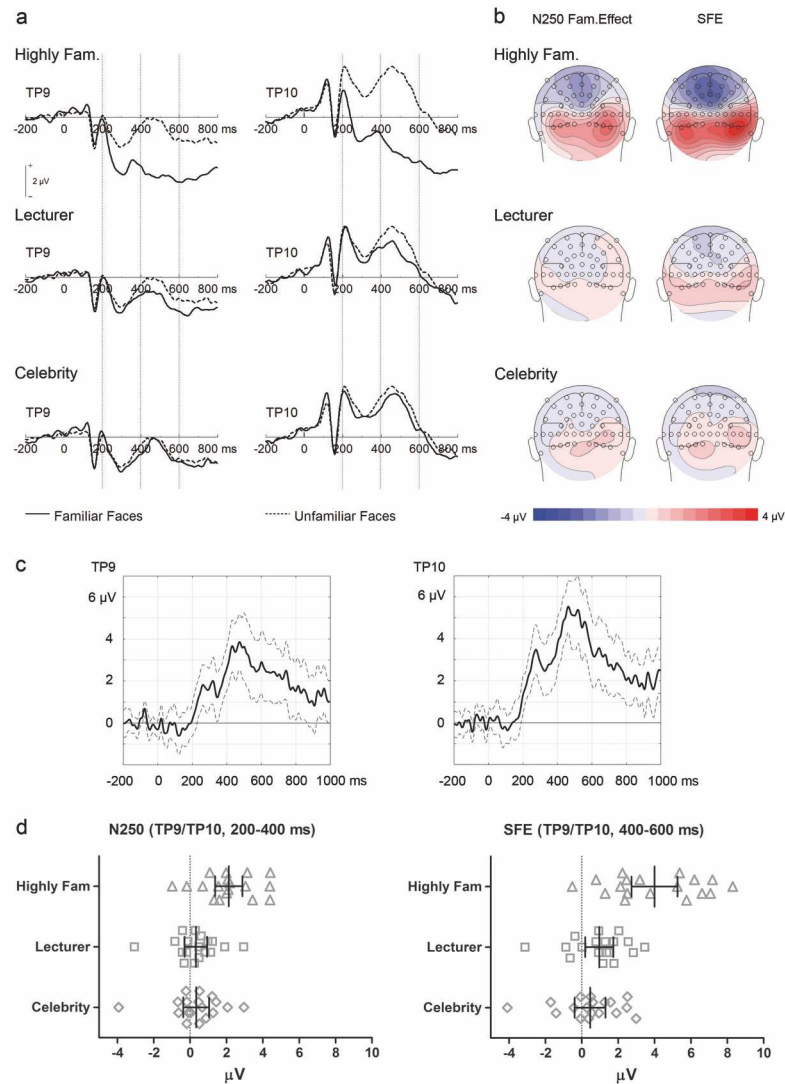


Figure 5. Results of Experiment 3. a) Grand average ERPs (N=18) for familiar faces at different levels of familiarity versus unfamiliar faces at left and right occipito-temporal channels TP9 and TP10. Dotted lines demark the N250 and SFE time windows. b) Scalp-topographical voltage maps (spherical spline interpolation, 110° equidistant projections) for N250 and SFE (unfamiliar minus familiar faces) for different levels of familiarity. c) Mean (\pm 95% CI) difference for unfamiliar minus highly familiar faces. d) Individual and mean (\pm 95% CI) familiarity effects for the N250 and SFE time ranges.

Analysis of N170 did not reveal significant familiarity effects (see supplementary). A repeated-measures ANOVA on N250 yielded a main effect of familiarity, $F(3, 51) = 21.88$, $p < .001$, $\eta_p^2 = .563$. Simple contrasts revealed significant differences between personally highly familiar and unfamiliar faces, $F(1, 17) = 38.39$, $p < .001$, $\eta_p^2 = .693$, but neither between lecturer and unfamiliar nor between celebrity and unfamiliar faces, both $F < 1$. A

significant interaction of familiarity by hemisphere, $F(3, 51) = 3.75, p = .016, \eta_p^2 = .181$, suggested larger familiarity effects for personally highly familiar versus unfamiliar faces over the right relative to the left hemisphere (see Figure 3c).

The ANOVA in the SFE time window revealed main effects of familiarity, $F(3, 51) = 30.47, p < .001, \eta_p^2 = .642$. Analyses of simple contrasts yielded significant differences between personally highly familiar and unfamiliar, $F(1, 17) = 46.32, p < .001, \eta_p^2 = .732$, as well as between lecturer and unfamiliar faces, $F(1, 17) = 6.14, p = .024, \eta_p^2 = .265$. The difference between celebrity and unfamiliar faces was not significant, $F < 1$. An interaction of familiarity by hemisphere, $F(3, 51) = 5.23, p = .003, \eta_p^2 = .235$, reflected larger familiarity effects over the right hemisphere for personally highly familiar faces in particular (Figure 3c).

Bootstrapping in the SFE time window revealed reliable familiarity effects for the personally highly familiar versus unfamiliar comparison in 15/18 participants, $P = .83$, 95% CI [.61, .94]. Comparison of the lecturer versus unfamiliar conditions yielded reliable effects in 4/18 participants, $P = .22$, 95% CI [.09, .45], whereas reliable effects for the celebrity versus unfamiliar condition were detected in 5/18 participants, $P = .28$, 95% CI [.13, .51].

Combining data from block 1 of Experiment 1 and the personally highly familiar condition of Experiment 3 resulted in reliable effects in 32/38 participants, $P = .84$, 95% CI [.70, .93].

General Discussion

The present study examined neural correlates of face recognition. Using multiple ambient images of personally highly familiar faces, we observed two ERP familiarity effects with more negative amplitudes for familiar than unfamiliar faces. Both showed occipito-temporal scalp distributions consistent with ventral visual stream generators. While the first (N250 familiarity effect, 200-400 ms) may reflect a match between the stimulus and a stored perceptual face representation, the second (SFE, 400-600 ms) is presumably related to the activation of stored semantic and affective content. Critically, the SFE represents a

remarkably strong (approximately 4 μ V) and highly reliable (shown by approximately 84% of the participants) effect. Moreover, the SFE (but not the N250) decreased with image repetition, and was substantially reduced for lesser-known personally familiar faces (lecturers) and absent for celebrity faces. These findings have strong implications, not only for theoretical accounts of the neural basis of face recognition but potentially also for applied settings.

First, any interpretation of the SFE has to explain why familiarity effects further increased following the N250. The N250 is typically interpreted as reflecting access to visual representations (Schweinberger & Burton, 2003; Schweinberger & Neumann, 2016). Activity of the underlying generator should therefore be sufficient for perceptual face recognition. The SFE might then reflect the activity of a “familiarity hub” that integrates the N250 stage with additional semantic, episodic and affective information. Such information is critical to social interaction and its integration can be seen as the core process of identifying familiar *persons* via their faces.

The observed modulation of the SFE by image repetition and levels of familiarity might further suggest that affective information plays a particularly important role. Presenting familiar faces elicits an affective response (Tranel et al., 1995; Herzmann et al., 2004), and its importance is impressively demonstrated by cases of Capgras’ delusion, in which patients with deficits in affective processing but relatively intact visual face recognition believe that highly familiar people have been replaced by “impostors” (Ellis, Young, Quayle, & DePauw, 1997; Ellis & Lewis, 2001; Hirstein & Ramachandran, 1997). Both long-known friends and relatives (Experiment 1) and close friends from university (Experiment 3) are likely to generate strong affective responses. At the same time, both lecturers and celebrities were presumably of less personal relevance, which resulted in a reduced (for personally but less familiar lecturers) or even absent SFE response (for celebrities). An alternative explanation solely in terms of feedback from semantic processing stages seems less likely, as more

semantic detail should be available for long-known celebrities relative to lecturers mostly known from large-scale classes. Finally, repeating images of familiar faces over blocks probably resulted in decreasing affective responses due to habituation. A reduced SFE with image repetition is therefore well in line with the above interpretation.

Neural models of face processing typically assume structures such as the amygdala, the insula or the ventral striatum to be involved in affective processing (Gobbini & Haxby, 2007), and activity in these regions is unlikely picked up by scalp EEG. Interestingly, some models assume feedback loops from affective to perceptual brain regions (Schweinberger & Burton, 2003; Gobbini & Haxby, 2007). We suggest that the SFE, presumably generated in the ventral visual pathway (see also supplementary material), reflects the existence of such loops. Specifically, affective processing may “boost” the perceptual mechanisms associated with visual face recognition, and it appears to be this additional activity generated by feedback from deep brain structures that is observed in the SFE. Although additional work is necessary to further substantiate this suggestion, our finding may represent a theoretical advancement to models of face recognition largely focusing on feedforward processing (Bruce & Young, 1986), and is in line with recent empirical (Bobes, Lage Castellanos, Quinones, Garcia, & Valdes-Sosa, 2013; Ewbank, Henson, Rowe, Stoyanova, & Calder, 2013; Johnston, Overell, Kaufmann, Robinson, & Young, 2016) and theoretical (Kramer, Young, & Burton, 2018) developments emphasising the importance of top-down processing.

Our bootstrapping analyses suggest that the SFE reliably occurs in 84% of the participants and that a false positive response is unlikely (0/20 participants in Experiment 2). As the use of multiple ambient images allows analysis of ERPs related to single identities, the SFE can form an objective marker of high personal familiarity with an individual face. Importantly, the effect was observed independent of the participants’ explicit indication that they had recognised the person. It might therefore be possible to develop a technique for testing familiarity even when participants are motivated to conceal their knowledge.

Our approach goes beyond the available P300-based variants of the Guilty Knowledge Test (GKT; Farwell & Donchin, 1991). Such experiments typically present participants with mock crimes, and information learnt during this first phase is then “detected” in the EEG part. Our procedure is substantially different, and both techniques measure different types of information. The GKT will detect information that is perceived as standing out from irrelevant control trials. In the context of face recognition, it will “detect” familiarity with specific images based on only brief encounters or pictures seen in the media. By contrast, a positive result in an SFE-based test would indicate substantial personal familiarity with the target person. Further studies now need to determine the exact level of familiarity necessary to elicit the effect and whether it can be suppressed by intentional use of countermeasures (Rosenfeld, Soskins, Bosh, & Ryan, 2004).

In contrast to previous studies, we did not observe a significant N250 familiarity effect for celebrities in Experiment 3. However, ERPs for celebrities were more negative relative to unfamiliar faces, and previously observed effects were relatively small (e.g., approximately 1 μ V in Gosling & Eimer, 2011). In such cases, sample variation will result in non-significant findings in some experiments. Additionally, the use of multiple ambient images in a butterfly detection task may have further reduced the effect in the present study. It is possible that participants did not recognise the celebrities in some specific pictures, and such unrecognised trials went into the respective averages (see also familiarity ratings in the supplement).

In conclusion, the present study is the first to present a strong and reliable neural correlate of the remarkable human ability to recognise personally highly familiar faces from a wide variety of images, including pictures that were never seen before – the Sustained Familiarity Effect. Its ventral scalp distribution as well as its modulation by image repetition and levels of familiarity indicate that this effect is driven by the integration of perceptual and affective information. We propose that this integrative process is at the very core of identifying a highly familiar person.

Author contributions

H.W., A.M.B and A.W.Y. developed the study concept and wrote the paper. H.W., S.T., B.I., C.C. and Z.G., designed the experiments, performed the data collection and analysed the data.

Acknowledgments

The authors are grateful to Samiha Salam-Khan and Megan Elley for help during EEG recordings.

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, 1620-1632. doi:10.1080/17470218.2016.1195851
- Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex*, 18, 1233-1238.
- Bauer, R.M. (1984). Autonomic Recognition of Names and Faces in Prosopagnosia - a Neuropsychological Application of the Guilty Knowledge Test. *Neuropsychologia*, 22, 457-469. doi:10.1016/0028-3932(84)90040-X
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551-565. doi:10.1162/jocn.1996.8.6.551
- Bobes, M.A., Lage Castellanos, A., Quinones, I., Garcia, L., & Valdes-Sosa, M. (2013). Timing and tuning for familiarity of cortical responses to faces. *PLoS One*, 8, e76100. doi:10.1371/journal.pone.0076100
- Bruce, V., & Young, A. (1986). Understanding Face Recognition. *British Journal of Psychology*, 77, 305-327.
- Bruce, V., & Young, A. (2012). *Face Perception*. Hove: Psychology Press.
- Burton, A.M., Jenkins, R., & Schweinberger, S.R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102, 943-958. doi:10.1111/J.2044-8295.2011.02039.X
- Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10, 243-248. doi:10.1111/1467-9280.00144

- Burton, A.M., Jenkins, R., Hancock, P.J.B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, 256-284.
doi:10.1016/J.Cogpsych.2005.06.003
- Burton, A.M., Kramer, R.S.S., Ritchie, K.L., & Jenkins, R. (2016). Identity From Variation: Representations of Faces Derived From Multiple Instances. *Cognitive Science*, 40, 202-223. doi:10.1111/cogs.12231
- Butler, D.L., Mattingley, J.B., Cunningham, R., & Suddendorf, T. (2013). Different Neural Processes Accompany Self-Recognition in Photographs Across the Lifespan: An ERP Study Using Dizygotic Twins. *Plos One*, 8. doi:10.1371/journal.pone.0072586
- Caharel, S., Courtay, N., Bernard, C., Lalonde, R., & Rebai, M. (2005). Familiarity and emotional expression influence an early stage of face processing: An electrophysiological study. *Brain and Cognition*, 59, 96-100.
doi:10.1016/J.Bandc.2005.05.005
- Davies-Thompson, J., Gouws, A., & Andrews, T.J. (2009). An image-dependent representation of familiar and unfamiliar faces in the human ventral stream. *Neuropsychologia*, 47, 1627-1635. doi:10.1016/j.neuropsychologia.2009.01.017
- Di Nocera, F., & Ferlazzo, F. (2000). Resampling approach to statistical inference: Bootstrapping from event-related potentials data. *Behavior Research Methods Instruments & Computers*, 32, 111-119.
- Ellis, H.D., & Lewis, M.B. (2001). Capgras delusion: a window on face recognition. *Trends in Cognitive Sciences*, 5, 149-156. doi:10.1016/S1364-6613(00)01620-X
- Ellis, H.D., Young, A.W., Quayle, A.H., & DePauw, K.W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society B-Biological Sciences*, 264, 1085-1092.

- Ewbank, M.P., & Andrews, T.J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *Neuroimage*, 40, 1857-1870.
doi:10.1016/j.neuroimage.2008.01.049
- Ewbank, M.P., Henson, R.N., Rowe, J.B., Stoyanova, R.S., & Calder, A.J. (2013). Different neural mechanisms within occipitotemporal cortex underlie repetition suppression across same and different-size faces. *Cerebral Cortex*, 23, 1073-1084.
doi:10.1093/cercor/bhs070
- Farwell, L.A., & Donchin, E. (1991). The Truth Will out - Interrogative Polygraph (Lie Detection) with Event-Related Brain Potentials. *Psychophysiology*, 28, 531-547.
doi:10.1111/j.1469-8986.1991.tb01990.x
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Gobbini, M.I., & Haxby, J.V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45, 32-41. doi:10.1016/J.Neuropsychologia.2006.04.015.
- Gobbini, M.I., Leibenluft, E., Santiago, N., & Haxby, J.V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, 22, 1628-1635.
doi:10.1016/j.neuroimage.2004.03.049
- Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia*, 49, 2736-2745.
doi:10.1016/j.neuropsychologia.2011.05.025
- Herzmann, G., Schweinberger, S.R., Sommer, W., & Jentsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, 41, 688-701.
- Hirstein, W., & Ramachandran, V.S. (1997). Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. *Proc Biol Sci*, 264, 437-444. doi:10.1098/rspb.1997.0062

- Jenkins, R., & Burton, A.M. (2011). Stable face representations. *Philos Trans R Soc Lond B Biol Sci*, 366, 1671-1683. doi:10.1098/rstb.2010.0379
- Jenkins, R., White, D., Van Montfort, X., & Burton, M.A. (2011). Variability in photos of the same face. *Cognition*, 121, 313-323. doi:10.1016/j.cognition.2011.08.001
- Johnston, P., Overell, A., Kaufman, J., Robinson, J., & Young, A.W. (2016). Expectations about person identity modulate the face-sensitive N170. *Cortex*, 85, 54-64. doi:10.1016/j.cortex.2016.10.002
- Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302-4311.
- Keyes, H., Brady, N., Reilly, R.B., & Foxe, J.J. (2010). My face or yours? Event-related potential correlates of self-face processing. *Brain and Cognition*, 72, 244-254. doi:10.1016/j.bandc.2009.09.006
- Kramer, R.S.S., Young, A.W., Burton, A.M. (2018). Understanding face familiarity. *Cognition*, 172, 46-58. doi:10.1016/j.cognition.2017.12.005
- Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102-1107. doi:10.1038/Nature03687
- Ramon, M., & Gobbini, M.I. (in press). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, doi:10.1080/13506285.2017.1405134.
- Rosenfeld, J.P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, 41, 205-219. doi:10.1111/1.1469-8986.2004.00158.x
- Schweinberger, S.R., & Burton, A.M. (2003). Covert recognition and the neural system for face processing. *Cortex*, 39, 9-30. doi:10.1016/S0010-9452(08)70071-6

Schweinberger, S.R., & Neumann, M.F. (2016). Repetition effects in human ERPs to faces.

Cortex, 80, 141-153. doi:10.1016/j.cortex.2015.11.001

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, 398-409.

Tranel, D., Damasio, H., & Damasio, A.R. (1995). Double Dissociation between Overt and

Covert Face Recognition. *Journal of Cognitive Neuroscience*, 7, 425-432.

Weibert, K., Harris, R.J., Mitchell, A., Byrne, H., Young, A.W., Andrews, T.J. (2016). An

image-invariant neural response to familiar faces in the human medial temporal lobe.

Cortex, 84, 34-42.