

Author's Accepted Manuscript

Enhancement of face-sensitive ERPs in older adults induced by face recognition training

Katharina Limbach, Jürgen M. Kaufmann, Holger Wiese, Otto W. Witte, Stefan R. Schweinberger



PII: S0028-3932(18)30484-6
DOI: <https://doi.org/10.1016/j.neuropsychologia.2018.08.010>
Reference: NSY6883

To appear in: *Neuropsychologia*

Received date: 7 January 2018
Revised date: 1 August 2018
Accepted date: 11 August 2018

Cite this article as: Katharina Limbach, Jürgen M. Kaufmann, Holger Wiese, Otto W. Witte and Stefan R. Schweinberger, Enhancement of face-sensitive ERPs in older adults induced by face recognition training, *Neuropsychologia*, <https://doi.org/10.1016/j.neuropsychologia.2018.08.010>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Enhancement of face-sensitive ERPs in older adults induced by face recognition training

Katharina Limbach^{1,*}, Jürgen M. Kaufmann^{1,4}, Holger Wiese^{2,4}, Otto W. Witte³, & Stefan R.
Schweinberger^{1,4,*}

¹*Department of General Psychology, Friedrich Schiller University, Jena, Germany*

²*Department of Psychology, Durham University, UK*

³*Jena University Hospital of Neurology*

⁴*DFG Research Unit Person Perception, Jena, Germany*

E-Mail: katharina.limbach@rub.de

E-Mail: stefan.schweinberger@uni-jena.de

*Correspondence concerning this article should be addressed to Katharina Limbach, Developmental Neuropsychology, Ruhr Universität Bochum, Universitätsstraße 150, 44801 Bochum

*Correspondence Stefan R. Schweinberger, Department of General Psychology, Friedrich Schiller University of Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany

Abstract

A common cognitive problem reported by older people is compromised face recognition, which is often paralleled by age-related changes in face-sensitive and memory-related components in event-related brain potentials (ERPs). We developed a new training using photorealistic caricatures based on evidence that caricatures are beneficial for people with compromised face processing. Twenty-four older participants (62-75 yrs, 13 female) completed 12 training sessions (3 per week, 60 min each) and 24 older participants (61-76 yrs, 12 female) acted as controls. Before and after training

ACCEPTED MANUSCRIPT
(or waiting), participants took part in a diagnostic test battery for face processing abilities, and in ERP experiments on face learning and recognition. Although performance improvements during the training provided little evidence for generalization to other face processing tasks, ERPs showed substantial training-related enhancements of face-sensitive ERPs. Specifically, we observed marked increases of the N170, P200 and N250 components, which may indicate training-induced enhancement of face detection and activation of identity-specific representations. Thus, neuronal correlates of face processing are plastic in older age, and can be modulated by caricature training.

Keywords: Age, face processing, caricaturing, event-related brain potentials, training

1. Introduction

Face processing, the ability to perceive, recognize and remember faces, is not only important for everyday social communication and interaction, but is also of general interest in brain and cognitive sciences as it requires successful integration of processes such as attention, perception, learning and memory (Başar et al., 2007). Face processing abilities show a developmental trajectory with a relatively late maturation and highest performance usually in the early 3rd decade of life (Germine et al., 2011). These abilities then stay at high levels up unto the 6th or 7th decade, when marked declines are usually found (Crook and Larrabee, 1992; Hildebrandt et al., 2011; Megreya and Bindemann, 2015; Savaskan et al., 2007). Of note, the decline in face processing cannot be explained by general age-related impairments, for example in memory or vision including acuity or contrast sensitivity (Hildebrandt et al., 2011). Specific aspects such as face memory or speed of face processing may decline somewhat earlier (Hildebrandt et al., 2010). Declines in the ability to retrieve a person's name are commonly described as a particularly disturbing and restrictive aspect of aging (Chaby and Narme, 2009; Leirer et al., 1990). As face recognition is often a necessary prerequisite for name retrieval (Burton and Bruce, 1993; Cohen and Burke, 1993), experienced difficulties in face naming may be partially due to age-related changes in face recognition.

Despite these consistent findings of a decline in face processing in older age, which are paralleled by age-related changes in electrophysiological markers of face processing (Chaby et al., 2003; Pfützte et al., 2002; Wiese et al., 2008; Wolff et al., 2012), there is an ongoing debate about its nature and process-specificity. For instance, age-related deficits in face memory could arise at

perceptual, representational or semantic stages. Diagnostic and chronometric approaches like EEG can help to distinguish between those possibilities (Hildebrandt et al., 2011; Pfützte et al., 2002; Wiese et al., 2017). Whereas expertise-related modulation of face-sensitive ERP components by same- and other-race faces (Komes et al., 2014) and access to semantic representations in memory seem to be unaffected by age (Wiese et al., 2017), there are conflicting results as to whether perceptual aspects of holistic face processing are compromised (Wiese et al., 2013) or spared (Boutet and Faubert, 2006; Meinhardt-Injac et al., 2014). However, access to domain-specific and domain-general representations of familiar persons that are crucial for successful face recognition seems to work less efficiently. Accordingly, compromised face processing may be especially due to changes in early perceptual and/or representational processing stages. This would be in agreement with theories of cognitive aging that suggest declines in perceptual processes and spared semantic knowledge (Craik and Bialystok, 2006; Ofen and Shing, 2013; Wiese et al., 2017).

We considered that promising approaches of training face recognition in older adults should thus focus on optimizing perceptual and representational processing. However, face processing is generally very difficult to improve, and training effects are often limited to the studied material, with little generalization. Both specific instructions on strategies to encode a face and (single-trial) feedback yield little benefit compared to spontaneous performance (Sporer, 1991) and often do not transfer to new faces (Alenezi and Bindemann, 2013; Hussain et al., 2009; White et al., 2014). Although Dolzycka and colleagues (2014) were able to increase the speed of face processing in middle-aged adults via an extensive 29 days training, this increase was not specific, but generalized to other complex visual objects. A further training on face-memory did not improve performance.

Approaches to enhance perceptual face processing via training in people with prosopagnosia (or face-blindness, a condition in which even personally familiar faces cannot be recognized; for a recent review see DeGutis et al., 2014a) have yielded more promising results. Two recent studies focusing on an enhancement of holistic processing reported marked improvements in everyday life and increased performance up to the level of normal control participants in face perception tasks at least for front-views (DeGutis et al., 2014a, b). As this training was based on the idea that individual second-order spatial configurations of faces are crucial for individuation, one limitation could concern

the recognition of familiar or famous faces. There is now substantial evidence suggesting that spatial information only plays a minor role at best, both for the recognition of famous (for an overview, see Burton et al., 2015) and learned faces (Itz et al., 2014).

Other evidence for successful modification of face processing was obtained by Tanaka and Pierce (2009), who showed that training at a subordinate level (i.e., learning to name individual face stimuli) increased recognition of other-race faces. Moreover, the amplitude of the recognition-related ERP component N250 to other-race faces increased after 5 days of 45 min training, indicating training-induced plasticity in electrophysiological correlates of face recognition. Of note, there is consistent evidence that the combined presentation of faces with names and other personal semantic information may boost perceptual representations, thus enhancing face learning (Kaufmann et al., 2009; Schwartz and Yovel, 2016) and categorical perception of unfamiliar faces (Kikutani et al., 2008).

A promising new approach to enhance perceptual representations that are diagnostic for individual recognition is digital caricaturing. Caricaturing can be used to enhance the idiosyncratic spatial and/or texture characteristics of a face relative to an average face, thus making it more distinctive from a hypothetical prototype (Irons et al., 2014; Kaufmann and Schweinberger, 2012; Kaufmann et al., 2013; Rhodes, 1996; Schulz et al., 2012). Distinctiveness has been shown to facilitate encoding and recognition of faces (Dewhurst et al., 2005; Going and Read, 1974; Schulz et al., 2012; Sporer, 1991; Vokey and Read, 1992). Spatial caricaturing, which enhances the idiosyncratic shape of features as well as their spatial configuration, has been shown to improve face recognition in individuals with poor face recognition skills (Kaufmann et al., 2013) and in participants with computer-simulated macular degeneration (Irons et al., 2014). Texture caricaturing enhances idiosyncratic color, pigmentation and luminance information, and may be even more efficient in improving face recognition for known faces (Itz et al., 2014). Importantly, a recent study showed transfer from caricature learning, in that faces learned as caricatures rather than veridicals (i.e. unaltered) were recognized better even when presented as veridical images at test ('caricature generalization benefit', Itz et al., 2017).

In the current study, we report on a newly-designed training that combines an enhancement of perceptual and representational face processing with learning tasks that include personal semantics and names. To maximize potential effects of perceptual training, we used faces that were caricatured in both shape and texture. We assessed training effects both at the behavioral and the electrophysiological level to specifically investigate the potential of older adults to benefit from intensive training using photorealistic caricatures. Of relevance, the latencies and amplitudes of several face-sensitive ERP components were previously shown to exhibit age-related changes (e.g. Pfütze et al., 2002; Wiese et al., 2008). The occipito-temporal N170 (maximal around 150-200 ms) is related to early structural encoding of faces or face detection at the categorical level (Bentin and Deouell, 2000; Eimer, 2000b, 2011). Previous research has furthermore shown increases in the N170 or the N1 (a component also shown for other object classes) after massed exposure with images from categories of animals, visual objects, or unfamiliar script (Gauthier et al., 2003; Maurer et al., 2010; Scott et al., 2006; 2008). The occipito-temporal P200 (or P2), a peak at about 200 to 250 ms after stimulus onset, is smaller for other-race compared to own-race faces. In individuals with substantial expertise with other-race faces, the effect was reduced (Stahl et al., 2008, 2010). The P200 is also smaller for caricatured compared to veridical faces (Kaufmann and Schweinberger, 2012; Schulz et al., 2012). This reduced P200 may reflect deviation of a face from a current prototype (see also Halit et al., 2000; Latinus and Taylor, 2006; Lucas et al., 2011). Finally, recent results suggest a selective sensitivity of the P200 to *spatial* prototypicality (Kloth et al., 2017; Schweinberger and Neumann, 2016).

The third family of occipito-temporal components related to face processing includes the N250, which has been associated with identity processing of familiar faces (Kaufmann et al., 2009; Schweinberger, 2011; Tanaka et al., 2006), and the N250r, and which is larger for repeated compared to non-repeated familiar faces (Schweinberger et al., 2002). The N250 is seen as a correlate of the acquisition of individual face representations during learning, as well as the activation of acquired face representations (for a review, see Schweinberger and Neumann, 2016). Finally, a central-parietal late-positive component (LPC) is typically enhanced for previously encountered items compared to novel ones ('old/new effect') and likely reflects episodic memory (Rugg et al., 1996) in the context of newly

ACCEPTED MANUSCRIPT
learned faces (Wiese et al., 2008). For familiar faces, this old/new effect might represent activation of person identity nodes (Burton et al. 1990) and facilitated activation of semantic information (Kaufmann et al., 2009; Schweinberger and Burton, 2003; Schweinberger et al., 1995).

Our main hypotheses were that the present training would (a) enhance identity processing of faces, (b) transfer to increase performance when learning new facial identities, and in parallel (c) increase the occipito-temporal N250 component, and the N250 familiarity effect in particular. Alternatively, to the extent that training would enhance earlier processes of face detection or encoding of spatial typicality due to massed exposure to faces, training effects should show up in earlier occipito-temporal N170 and P200 components. Finally, we hypothesized that training effects on later memory processes would be reflected in a central-parietal LPC.

2. Material and Methods

2.1 Participants

Forty-eight older White Caucasian participants (61-76 yrs, 25 females) were recruited for the experiment and contributed data. All lived independently in the city of Jena in Germany, were right-handed and reported normal or corrected-to-normal vision. The Freiburg Visual Acuity Test (FrACT, Version 3.5.5, Bach, 1996) at 90 cm viewing distance showed mean visual acuity of logMar = .20, range: 0.05 – 0.59, and mean Michelson Contrast score of 1.81, range: 0.71 – 4.67. Participants neither reported neurological/psychiatric conditions nor substance abuse, nor received central-acting medication. For balanced age and gender distributions, participants were pseudo-randomly assigned to the training (24 participants; 13 female) and no-treatment control group (24 participants; 12 female) before the first session. All participants gave written informed consent. Experimental procedures were approved by the University of Jena Ethics board and conformed to the Declaration of Helsinki. Participants who completed the training received 140 € and participants in the control group received 45 € as a reimbursement for their time.

2.2 General Procedure

The study encompassed an extensive multi-session/multi-measure training program (see Fig. 1 for an overview of the procedure). Here we focus on the EEG experiments conducted in the pre-, post- and follow-up sessions. The training comprised twelve sessions (three one-hour sessions per week),

without EEG recording. The first training session started two days after the pre-session, and the last training session was followed by the post-session the day after. The follow-up was conducted four weeks after the post-session. Participants in the control group only conducted the pre- and post-sessions, scheduled four weeks apart to match the duration of the training.

Except for more extensive diagnostic testing in the pre-session, the same procedure was used in the pre-, post- and follow-up sessions. Each session was scheduled at roughly the same time of day for each participant (usually between 9 am and 12.30 pm). The order of the diagnostic test battery and EEG experiments was counterbalanced across participants, but constant within participants. These sessions usually lasted between 2.5 and 3.5 h.

2.3 Experimental procedures (EEG)

The experimental procedure for the *Learning Experiment*, which was always conducted before the *Famous Faces Experiment*, is depicted in Figure 2. Participants were instructed to memorize each individual face during the study phase (each face was presented three times). In the immediately following test phase, all 25 ‘old/learned’ faces were presented randomly intermixed with 25 ‘new/novel’ faces. Participants decided as quickly and accurately as possible whether a given face had been presented in the directly preceding study phase. This procedure was repeated in a second learning and test block using different faces, so that in total 50 faces were learned.

To exclude the use of simple image cues for recognition, different pictures of the learned faces were presented at test (e.g. slightly different facial expression or viewing angle, Fig. 2). If participants did not respond within the 1500 ms presentation of the picture, the trial was discarded and they received written feedback during the 1200 ms inter-stimulus interval that they should respond faster. The allocation of response hands (left/right index finger for old/new faces) was counterbalanced across participants, but was constant across all sessions within participants. In an initial practice block with different stimuli, participants could practice the task with single-trial feedback until they felt competent.

In the *Famous Faces Experiment*, participants were presented with photographs of 50 familiar (national and international celebrities chosen to fit our participant group) and 50 unfamiliar faces in random order and had to decide whether or not a familiar person was depicted. The Famous Faces

Experiment was used to determine the degree to which any training effects would generalize to the processing of other untrained faces, which were well represented already before the experiment.

Timing, procedure and allocation of response hands (left/right for familiar/unfamiliar) were as in the test phase of the Learning Experiment. Again, participants practiced the task with different stimuli.

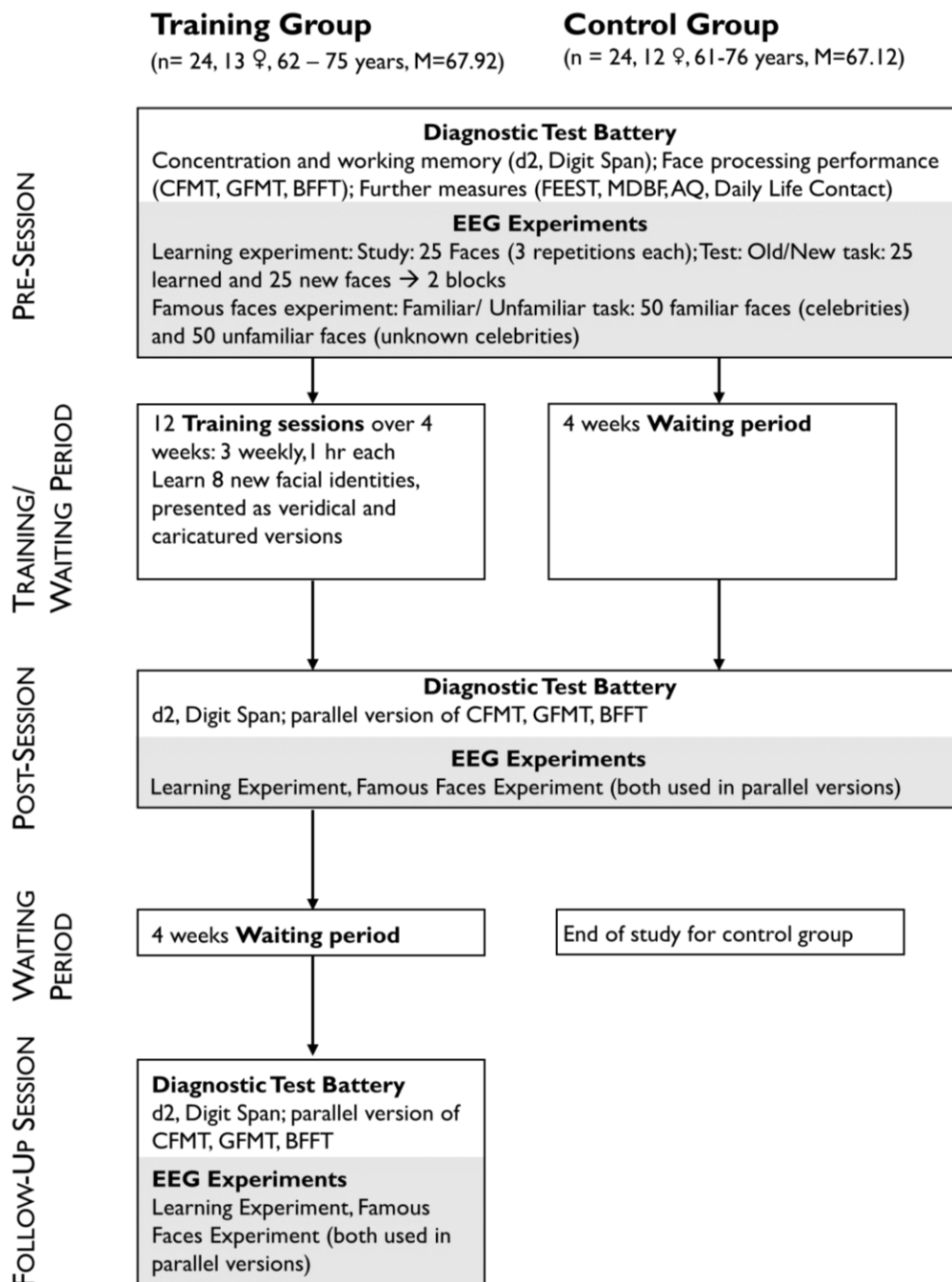


Figure 1. Study Design. Analysis focuses on gray shaded EEG experiments. Diagnostic test battery is described in the supplements. Note: CFMT - Cambridge Face Memory Test; GFMT - Glasgow Face Matching Test; BFFT - Bielefelder Famous Faces Test; FEEST - Facial Expressions of Emotion Stimulus Test; MDBF – Multi-dimensional Mood Questionnaire (‘Multidimensionaler Befindlichkeitsfragebogen’); AQ – Autism Quotient.

identities were used in each session, and stimulus sets were counterbalanced across conditions and sessions. At the end of each EEG session, participants rated their familiarity with the celebrities in the Famous Faces Experiment on a paper-and-pencil questionnaire. For this purpose, the name and at least one semantic information (e.g. Boris Becker, former professional tennis player) was provided and participants rated familiarity on a 3-point scale (2: “familiar with the person, and have seen a picture before”; 1: “only familiar with the name”; 0: “not familiar with the person”). Celebrities that were rated as unknown or only known by their name were excluded from further analysis.

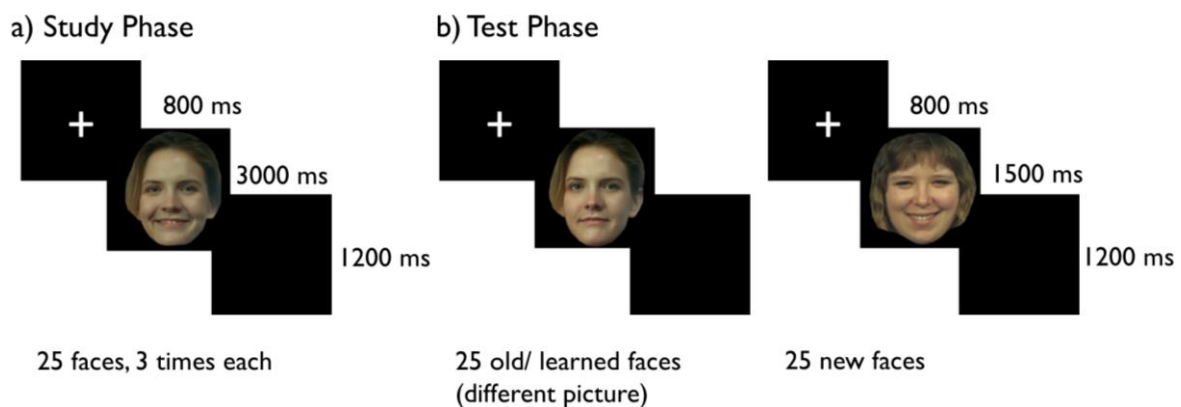


Figure 2. Trial sequences of the Learning Experiment. a) Study phase b) Test phase with ‘old/new task’. Participants were asked to respond whether the face was ‘old’ during its presentation. The experiment consisted of two study and test phase cycles, resulting in 50 old and 50 new faces overall. Figure not drawn to scale.

2.3.1 Stimuli. Stimuli for the *Learning Experiment* consisted of 308 unfamiliar Caucasian faces (50% female), each available in two different versions and taken from the GUFID and FERET databases (Burton et al., 2010; Phillips et al., 2000; Phillips et al., 1998). For the *Famous Faces Experiment* a total of 155 familiar and 155 unfamiliar Caucasian faces were compiled from freely available pictures on the internet. To account for possible differences in image quality or attractiveness, pictures of foreign celebrities that are not famous in Germany were used as unfamiliar faces. All images were edited using Adobe PhotoshopTM (CS4, version 11.0) to remove jewellery and any information from the neck downwards, and were presented against black background. Image size was 400×530 pixel. Size of each face was adjusted towards equivalent interocular distances.

2.4 Training tasks

During each training session eight new facial identities were learned. The order of the twelve sessions was counterbalanced across participants. Typically, two participants were trained at the same

time working on two equivalent computers separated by room dividers. Each training session consisted of a learning, a matching and a forced-choice recognition task (cf. Fig. 3 and supplements). We selected 96 full color Caucasian faces with neutral expression from the Jena 3D Face Database (J3DFD). Faces were presented without external features in six different viewpoints and in both the veridical (unaltered) and caricatured version (shape and texture were simultaneously enhanced by 50%, for a similar procedure see Itz et al., 2014).

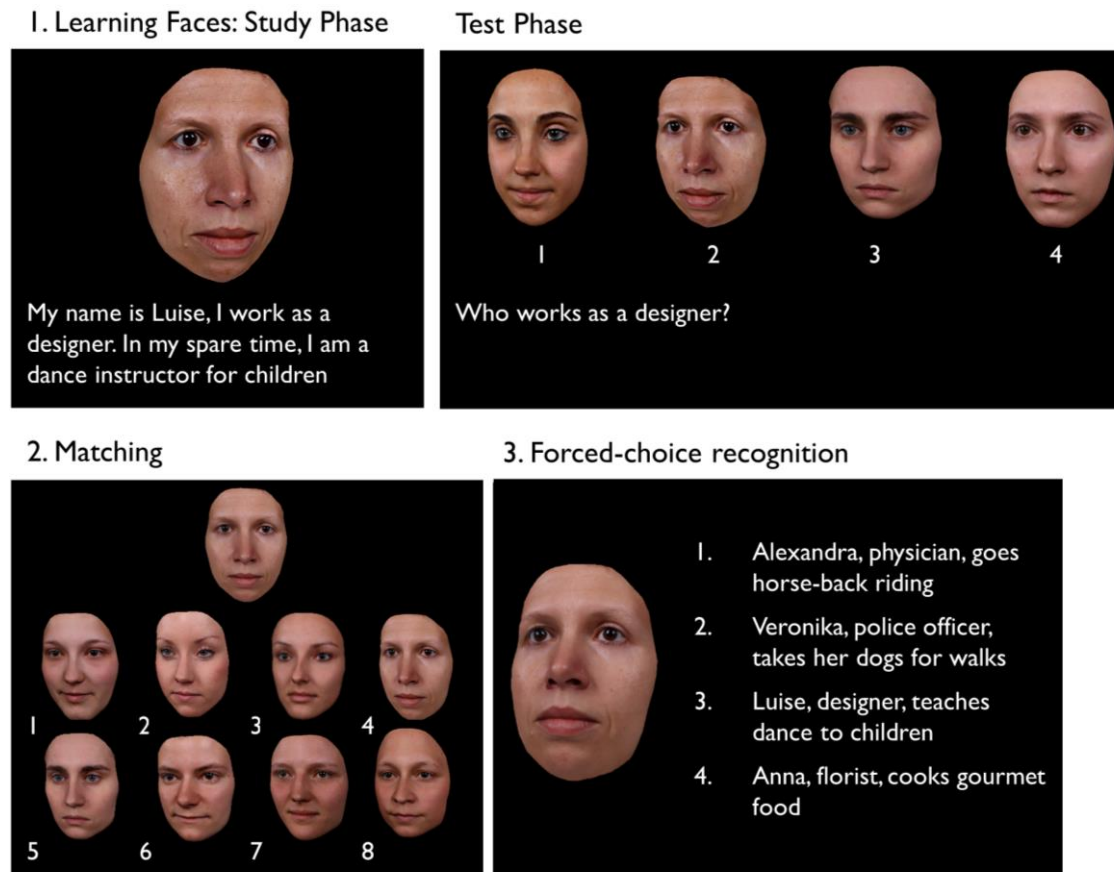


Figure 3. Training tasks. Each training session focused on eight identities. 1. Each identity was presented continuously for 16 s, but in three different views (in order, view 1 without semantic information, followed by view 1, view 2, and view 3 each with semantic information, for 4 s each). Across successive study/test cycles, each identity was shown as veridical and as caricature. Note that within each cycle, the study phase was repeated three times, and the corresponding test phase included individual trials for each category of semantic information (i.e., name, occupation, hobby). The study/test cycle was repeated at least once and up to four times, depending on performance (criterion: 87.5% correct; cf. supplement). 2. The matching task was used in a simultaneous version as depicted (target displayed on top) and in a delayed-matching to sample version (isolated target shown for 3 s before the array was presented). 3. Forced-choice recognition task. The target face in this example was shown as caricature in the Learning task and as veridical in the other two tasks. Note that both veridical and caricatured versions were used for all identities in 2. and 3. Feedback was given after each trial.

2.5 Electrophysiological recording and analysis

During EEG experiments, participants were seated in front of a computer screen in a dimly lit, electrically shielded and noise-attenuated chamber with their heads in a chin rest to ensure a constant distance of 90 cm to the monitor. EEG was recorded continuously with a 512-Hz sampling rate from

DC to 120 Hz, using a 32-channel BioSemi Active II system (BioSemi, Amsterdam, Netherlands).

Active sintered Ag/AgCl-electrodes were mounted in an elastic cap with recording sites at Fz, Cz, Pz, Iz, FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, F9, F10, FT9, FT10, TP9, TP10, P9, P10, PO9, PO10, I1 and I2. Note that BioSemi systems work with a 'zero-Ref' setup with ground and reference electrodes replaced by a CMS/DRL circuit (<http://www.biosemi.com/faq/cms&drl.htm> for further information). Offline, data were analyzed using the EEGLAB toolbox (Delorme and Makeig, 2004) and custom-written scripts for MATLAB (The Math Works, Inc., Natick, MA). Pre-processing routines were identical for both EEG experiments.

Continuous EEG was first digitally high-passed filtered at 0.1 Hz using a finite impulse response filter and then segmented into 3.2 s epochs starting 1 s before stimulus onset. In a first step, channels that contained excessive noise were discarded and data were re-referenced to average reference. Then segments were baseline corrected to -200 to 0 ms. Epochs containing excessive noise ($\pm 500 \mu\text{V}$) were automatically rejected before submitting data to an independent component analysis (ICA). Components related to eye blinks were excluded (1-2 components per participant). Subsequently, segments were visually inspected for artifacts (i.e. atypical artifacts and muscle activity) and previously rejected channels were interpolated (using spherical-spline interpolation). Only trials from the test phase were analyzed for the Learning Experiment; for both experiments, only trials with correct responses were further analyzed. Average trial numbers (SEM) across sessions in the Learning Experiment were 34.3 (1.3) for hits, 34.2 (1.1) for correct rejections (CR) in the *training group*, and 35.3 (1.3) for hits, 35.2 (1.2) for correct rejections (CR) in the *control group*. For the Famous Faces Experiment, corresponding numbers were 27.2 (1.2) for hits, and 35.9 (1.3) for CR in the *training group* and 30.4 (1.8) for hits, and 34.2 (1.5) for CR in the *control group*. Trial numbers did not differ between experimental groups or sessions (all $p \geq .15$). Only in the Famous Faces Experiment there were significantly more trials for CR than hits (in both groups and each session; $-4.658 > t < -3.369$; $p \leq .001$), potentially reflecting a tendency towards conservative responses and/or the exclusion of unknown celebrities ($M = 10$ excluded per session and participant). Two participants were excluded due to low trial numbers (< 10), for the Learning Experiment only.

ms), N170 (160–200 ms), P200 (250–360 ms), N250 (360–480 ms), and late positive components LPCa (400–600 ms) and LPCb (600–800 ms). Time intervals for the components and electrodes of interest were chosen based on distinct peaks identified in the grand mean averages across all conditions and on previous research while considering somewhat prolonged latencies in older adults (e.g. Wiese et al., 2008; 2012; 2013; Wolff et al. 2012). Accordingly, P100 was quantified at O1/O2, N170, P200 and N250 at P9, P10, PO9, and PO10, and, LPCa and LPCb at P3, Pz, P4, C3, Cz, C4, F3, Fz and F4.

3. Results

Behavioral and EEG data were submitted to analyses of variance (ANOVA) and t-tests. Where appropriate, Epsilon corrections for heterogeneity of covariances (Huynh and Feldt, 1976) were performed throughout. Effect sizes are reported as unbiased Cohen's d_{av} for dependent t-tests and as η_p^2 for ANOVAs. In addition to the omnibus ANOVAs we report planned comparisons for the most relevant EEG effects, i.e., the differences between the pre- and post-session on both the N170 and the N250, as well as the difference in the familiarity effect between the pre- and post-session on the N250. These comparisons were based both on our hypotheses and the results of the omnibus ANOVAs.

3.1. Training performance

In general, participants showed higher accuracy and shorter RTs for caricatured compared to veridical faces in all tasks and across sessions (please see supplements for a detailed analysis and depiction of caricaturing effects). Most importantly, there were also significant performance changes across sessions. To quantify those performance changes across training sessions, we compared mean performance levels across sessions 1-6 with performance across sessions 7-12, at a task-specific level (accuracy in the learning task for the 2nd cycle, i.e. after presentation of training faces both as veridicals and caricatures; accuracy and RT for the matching and recognition tasks. Performance was averaged across trials with veridical and caricatured faces). Accuracy significantly increased in the learning (S1-6 vs. S7-12: 0.647 vs. 0.690; $t(23) = -2.778$, $p_{corr} = .0427$, $d_{av} = .28$) and the recognition task (0.794 vs. 0.831; $t(23) = -3.051$, $p_{corr} = .0227$, $d_{av} = .32$). No significant change in accuracy was shown for the simultaneous matching task (0.935 vs. 0.938, $t(23) = -0.668$, $p_{corr} > .9$), whereas

accuracy decreased for the delayed matching task (0.914 vs. 0.895; $t(23) = 2.997, p_{corr} = .0257, d_{av} = .32$). In turn, RTs significantly decreased in both matching tasks (simultaneous: 6936 ms vs. 6355 ms; $t(23) = 4.658, p_{corr} = .0033, d_{av} = .40$; delayed: 4763.1 ms vs. 4590 ms; $t(23) = 2.745, p_{corr} = .0346, d_{av} = .21$). No significant change in reaction time was shown for the recognition task (6545 ms vs. 6233 ms, $t(23) = 1.562, p_{corr} > .4$). As new sets of faces were learned in each session, improvements reflect transfer effects across training sessions.

3.2 Diagnostic test battery

Participants in the training and control group did not differ in age, visual acuity or performance in any measure in the pre-session (cf. supplements). Across tests, face processing performance was about 10% lower than performance of norm groups of young adults, in line with other findings (e.g. Germine et al., 2011; Hildebrandt et al., 2010). Overall, we did not observe prominent training-induced improvements in the diagnostic test battery (see Table S2).

3.3 Learning Experiment

3.3.1 Behavioral performance. We analyzed signal detection parameters d' and response bias (c) with repeated measures on session (pre vs. post), and the between-subjects factor group (training vs. control). Hit and false alarm rates equal to zero or one were adjusted to 0.999 or 0.001 respectively (Macmillan and Creelman, 2005). ANOVAs for accuracies and correct reaction times (RTs) were performed analogously, but with the additional factor familiarity (learned/old vs. new).

Table 1: Performance in the Learning Experiment. Note: Table only contains effects involving session or group (left), and comparisons of both pre- and post-sessions with the follow-up session in the training group (right).

	Pre		Post		Factor, Test statistic	Follow Up	Factor, Test statistic
	Training	Control	Training	Control		Training	
d'	1.31 (0.54)	1.66 (0.66)	1.39 (0.53)	1.49 (0.60)	Session $F(1,44) < 1, p > .5$; Session \times Group $F(1,44) = 3.227,$ $p = .079$	1.34 (0.36)	pre vs. follow-up $t(21) = -.336, p > .7$; post vs. follow-up $t(21) = -.481, p > .6$
c	0.02 (0.41)	- 0.02 (0.49)	0.05 (0.31)	- 0.02 (0.35)	Session $F(1,44) < 1, p > .8$; Session \times Group $F(1,44) < 1, p > .8$	-0.01 (0.39)	pre vs. follow-up $t(21) = -.377, p > .7$; post vs. follow-up $t(21) = -1.280, p > .2$
Acc					Session		pre vs. follow-up
Old	0.709 (0.165)	0.766 (0.144)	0.728 (0.103)	0.755 (0.122)	$F(1,44) < 1, p > .8$; Session \times Group	0.740 (0.117)	Session $F(1,21) < 1, p > .4$;
New	0.732 (0.115)	0.763 (0.131)	0.751 (0.133)	0.745 (0.129)	$F(1,44) = 2.717, p > .1$;	0.722 (0.128)	post vs. follow-up Session $F(1,21) < 1, p > .5$;

	ACCEPTED MANUSCRIPT				Session	pre vs. follow-up
RT						
Hits	979.1 (100.2)	952.7 (102.0)	1020.1 (124.3)	934.0 (107.4)	$F(1,44) = 3.196,$ $p = 0.081;$	979.1 (103.48)
CR	1110.1 (148.4)	1063.9 (119.3)	1119.0 (137.3)	1107.2 (148.4)	Session \times Group $F(1,44) < 1, p > .5$	1110.1 (109.79)
						post vs. follow-up Session $F(1,21) = 1.895,$ $p > .1;$

There were no main effects for group (all $F_s < 2.3$, all $p_s > .14$). As can be seen in Table 1, despite a tendency for an interaction between session and group in d' , participants in neither group improved their performance between sessions. Across groups, a main effect of familiarity in RTs, $F(1,44) = 55.771, p < .001, \eta_p^2 = .559$, indicated faster responses to learned compared to new faces.

3.3.2 EEG. ANOVAs were performed with familiarity (hits vs. CR) and session (pre vs. post), and the between-subjects factor group (training vs. control). Analysis of P100 included an additional topographical factor hemisphere. Face-sensitive components N170, P200 and N250 were analysed using additional topographical factors of site (anterior vs. posterior) and hemisphere. The LPCa and LPCb were analyzed using topographical factors of site (frontal/central/parietal) and laterality (left/midline/right). Only effects involving the factors of interest, namely session, familiarity and group, are reported below. No main effects for group were observed (all $F_s < 1.4$, all $p_s > .2$).

P100. No significant effects of interest were found.

N170. We found an effect of session, $F(1,44) = 6.910, p = .012, \eta_p^2 = .136$ that was qualified by an interaction of Site \times Session \times Group, $F(1,44) = 10.487, p = .002, \eta_p^2 = .192$. Follow-up ANOVAs were performed separately for each group. The analysis of the *training group* yielded an effect of session, $F(1,21) = 7.300, p = .013, \eta_p^2 = .258$ that was further qualified by an interaction of Site \times Session, $F(1,21) = 8.768, p = .007, \eta_p^2 = .295$. Further analysis revealed no effect for more anterior electrodes P9 and P10, $F(1,21) < 1, p > .6$, but an effect of session at PO9 and PO10, $F(1,21) = 14.095, p = .001, \eta_p^2 = .402$. Here the N170 was substantially larger in the post- compared to the pre-session (see Fig. 4 and 5). No effects involving session were found for the *control group*, all $F_s < 2$; all $p_s > .1$.

P200. We obtained significant interactions of Site \times Session \times Group, $F(1,44) = 8.882, p = .005, \eta_p^2 = .168$, as well as of Site \times Hemisphere \times Session \times Group, $F(1,44) = 5.911, p = .019, \eta_p^2 = .118$. Separate ANOVAs per group revealed significant interactions of Site \times Session, $F(1,21) = 6.915,$

$p = .016$, $\eta_p^2 = .24$, and Site \times Hemisphere \times Session, $F(1,21) = 5.325$, $p = .031$, $\eta_p^2 = .202$, for the *training group* only. Separate ANOVAs per electrode revealed larger amplitudes in the post-, compared to the pre-session at P9, $F(1,21) = 6.089$, $p = .022$, $\eta_p^2 = .225$. No effect of session was found at the other electrodes, all $F_s < 1.8$, all $p_s > .2$. Furthermore, no significant effects involving session were found for the *control group*, all $F_s < 2.5$; all $p_s > .1$ (Fig. 4 & 5).

N250. We observed an interaction of Hemisphere \times Familiarity, $F(1,44) = 6.599$, $p = .014$, $\eta_p^2 = .130$, indicating larger negativity for CR compared to hits over the right hemisphere, $F(1,44) = 4.209$, $p = .046$, $\eta_p^2 = .087$, but not over the left hemisphere, $F(1,44) < 1$, $p > .35$. We also found a significant interaction of Site \times Session \times Group, $F(1,44) = 7.057$, $p = .011$, $\eta_p^2 = .138$ that was further qualified by an interaction of Site \times Hemisphere \times Session \times Group, $F(1,44) = 5.045$, $p = .03$, $\eta_p^2 = .103$. Separate ANOVAs per group again did not reveal any significant effects involving session for the *control group*, all $F_s < 2.7$; all $p_s > .23$. However, for the *training group* we found a significant interaction of Site \times Session, $F(1,21) = 7.567$, $p = .012$, $\eta_p^2 = .265$, indicating greater negativity in the post-, compared to the pre-session at the more posterior electrodes PO9 and PO10, $F(1,21) = 5.472$, $p = .029$, $\eta_p^2 = .207$. No effect of session was found for P9 and P10, $F(1,21) < 1$, $p > .5$.

Taken together, we found significant effects of training in face-sensitive components N170, P200 and N250. The negative components N170 and the N250 increased at PO9 and PO10 in the post- compared to the pre-session only in the *training group*. Please see Table 2 and Figure 6 for further analysis (using planned comparisons and reporting effect sizes and their 95% CI) and depiction of these results. At the same time, the positive component P200 increased at P9 in the post compared to the pre-session in the *training group*.

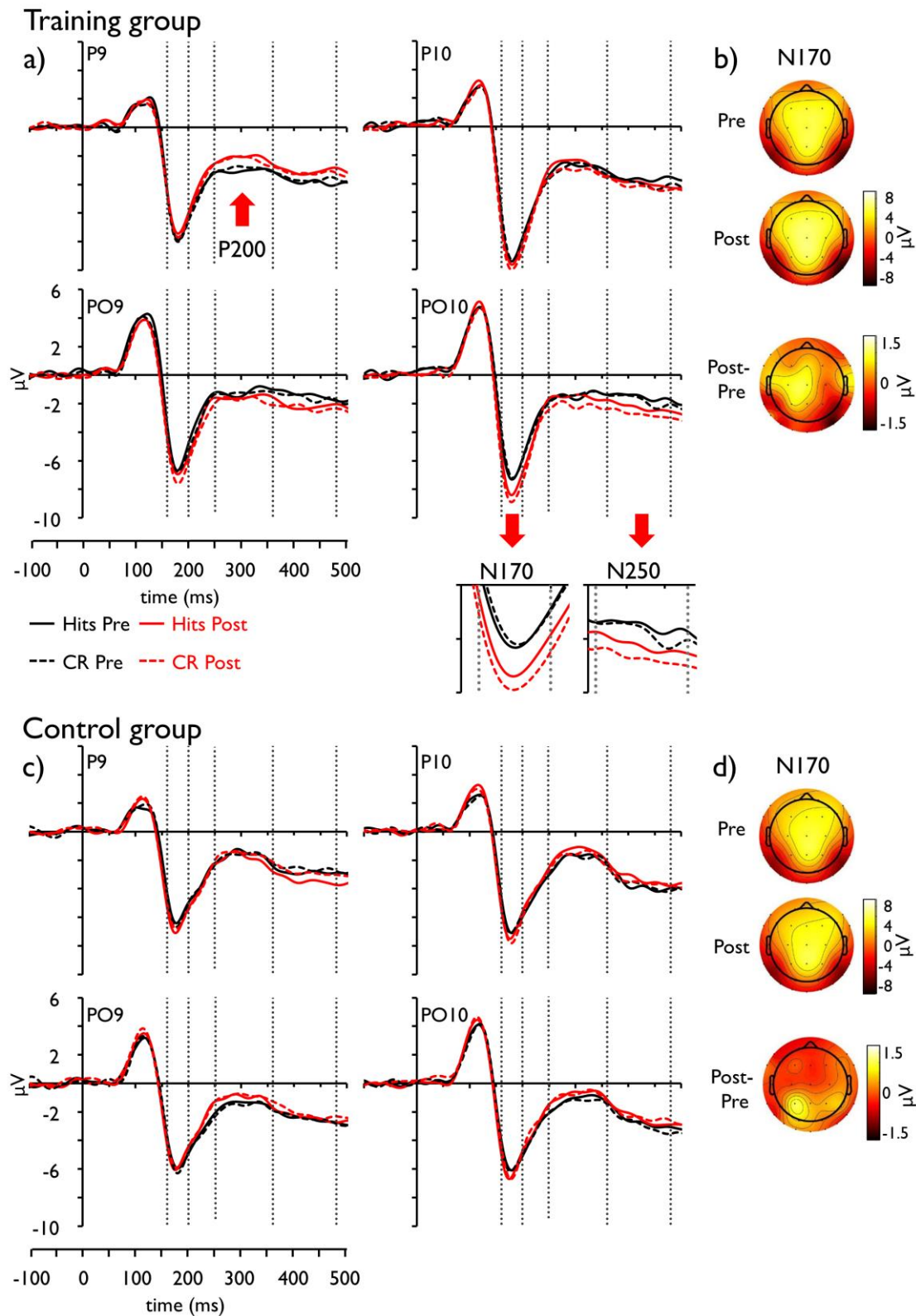


Figure 4: ERPs in the Learning Experiment. a + c) Test phase ERPs at occipito-temporal electrodes for training and control group, separately for hits and correct rejections in the pre- and post- session. b + d) Topographical voltage maps for the N170 time window (160-200 ms) in the pre- and post-session, averaged across hits and CR. Maps of differences between post and pre-session illustrate the increased right posterior N170 in the training group.

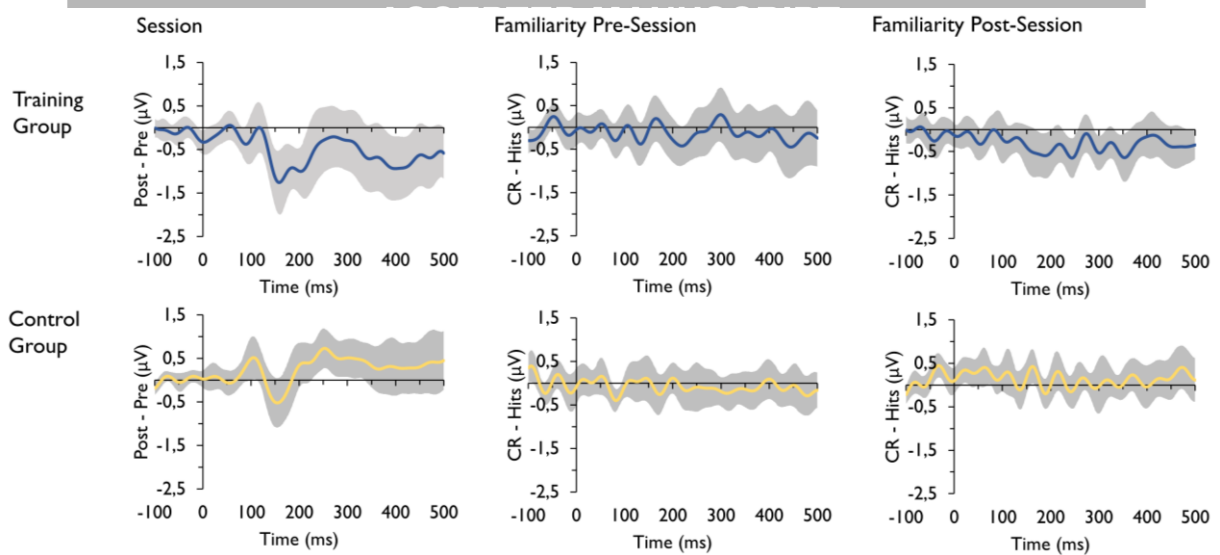


Figure 5: ERP Difference waves in the Learning Experiment. Results are shown for averages across PO9 and PO10. Difference waves surrounded by 95% confidence intervals for main effects of session and familiarity. Training group is shown in blue, control group is shown in yellow. Effects of session are calculated by subtracting the average waveforms (hits and CR) from the pre-session from those of the post-session. Familiarity effects are shown per session and are calculated as (CR – hits).

Table 2: Planned comparisons for most relevant effects in the Learning Experiment. Results are shown averaged across PO9 and PO10. Please refer to Figure 6 for a graphical depiction of these results.

	Training Group					Control Group				
	M_{diff} (S_{diff})	t (df)	p	Cohen's d_{av}	95% CI	M_{diff} (S_{diff})	t (df)	p	Cohen's d_{av}	95% CI
N170 Session	-1.03 (1.33)	-3.63 (21)	.002**	-0.306	-0.510 -0.118	-0.11 (1.34)	-0.41 (23)	.685	-0.035	-0.208 0.136
N250 Session	-0.78 (1.64)	-2.24 (21)	.036*	-0.295	-0.587 -0.020	0.31 (1.45)	1.036 (23)	.311	0.131	-0.125 0.392
N250 Session × Familiarity	-0.08 (1.57)	-0.24 (21)	.814	-0.064	-0.610 0.479	0.26 (1.32)	0.960 (23)	.347	0.271	-0.301 0.856

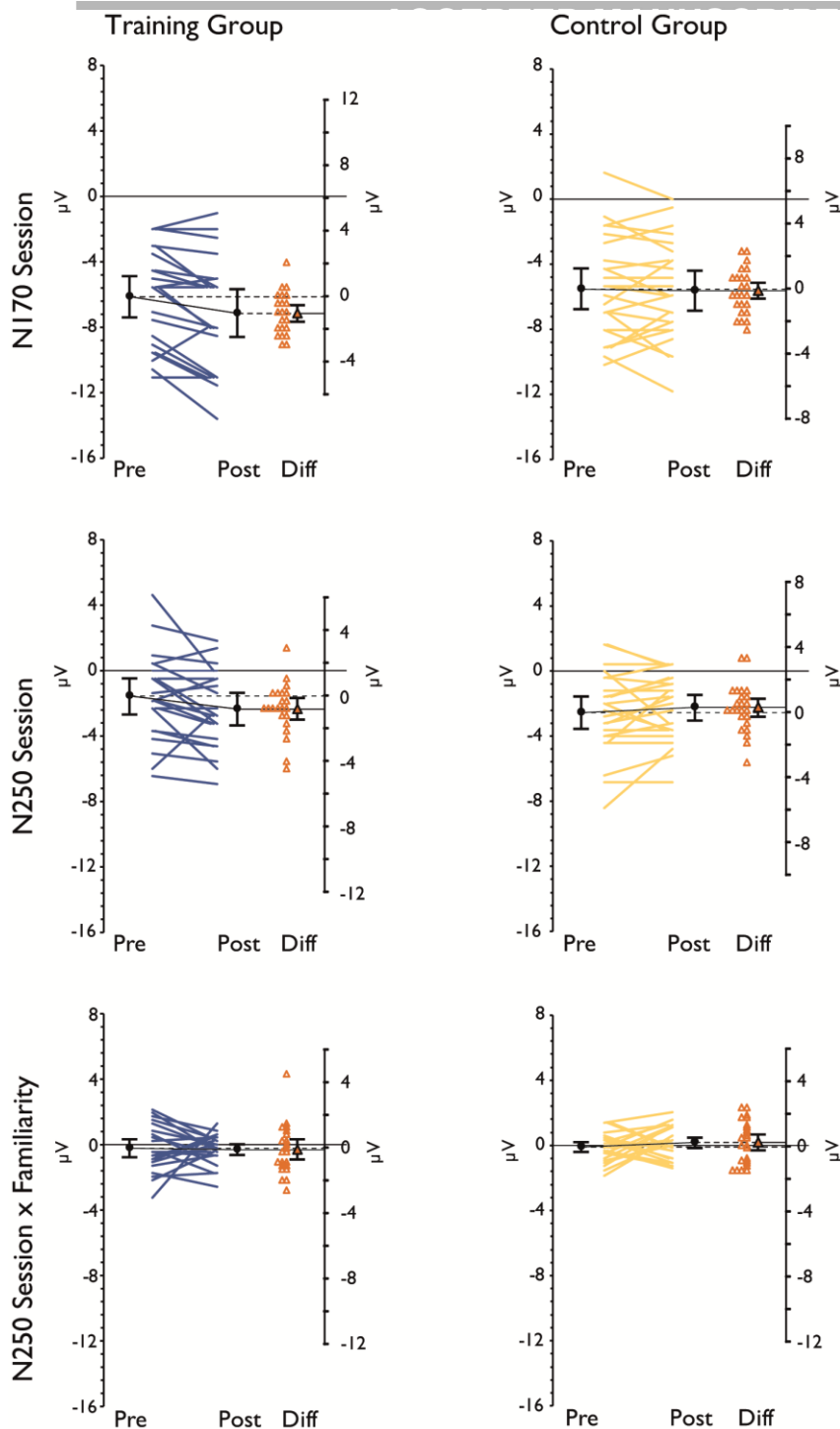


Figure 6: ERP effects of session in the Learning Experiment. Results are shown for averages across PO9 and PO10, the training group is depicted on the left in blue and the control group on the right in yellow. Lines represent single participant data (pre- and post-session.) Orange triangles show differences between sessions. Filled circles show mean per session and filled triangles show mean differences (please note separate scale to the right). Error bars depict 95% confidence intervals. Significant differences between sessions were found for the N170 and N250 in the training group only. Figures were made with ESCI intro software (Cumming & Calin-Jageman, 2017).

LPCa. An interaction of Laterality \times Familiarity, $F(2,88) = 7.507$, $p = .002$, $\eta_p^2 = .146$, $\epsilon = .842$, was followed up by separate ANOVAs per laterality. These revealed an effect of familiarity, with larger positivity for CR than hits only at left-hemispheric electrodes, $F(1,44) = 6.755$, $p = .013$,

$\eta_p^2 = .133$. No effects for familiarity were found over the midline, $F(1,44) < 1$, $p > .5$, or right hemispheric electrodes, $F(1,44) = 3.717$, $p = .06$. We also observed an interaction of Site \times Session \times Group, $F(2, 88) = 3.956$, $p = .031$, $\eta_p^2 = .082$, $\epsilon = .817$. However, when followed up for each group separately, no significant effects involving session emerged, all $F_s < 2.6$, all $p_s > .1$.

LPCb. An effect of familiarity, $F(1,44) = 5.086$, $p = .029$, $\eta_p^2 = .104$, and interactions of Laterality \times Session, $F(2,88) = 4.973$, $p = .009$, $\eta_p^2 = .102$, and of Laterality \times Familiarity, $F(2,88) = 8.407$, $p < .001$, $\eta_p^2 = .160$, were followed up by separate ANOVAs for each laterality level. These revealed significant effects for session, $F(1, 44) = 4.285$, $p = .044$, $\eta_p^2 = .089$ and familiarity, $F(1, 44) = 13.396$, $p = .001$, $\eta_p^2 = .233$, at the left electrodes, as well as an effect of familiarity at the midline electrodes, $F(1, 44) = 6.489$, $p = .014$, $\eta_p^2 = .129$. The effects of familiarity generally reflected greater positivity for CR compared to hits (i.e., a “reversed” old/new effect that has been frequently reported in older participants, Duarte et al., 2006; Wiese et al., 2012). The effects of session over the left hemisphere reflected more positivity in the post- compared to the pre-session. No effects involving session or familiarity were found at right hemispheric electrodes, all $F < 2$, all $p_s > .1$. Furthermore, there were two three-way interactions: Site \times Session \times Group, $F(2,88) = 4.315$, $p = .028$, $\eta_p^2 = .089$, $\epsilon = .727$, and Session \times Familiarity \times Group, $F(1,44) = 6.086$, $p = .018$, $\eta_p^2 = .122$, which were followed up by separate ANOVAs per group. For the *training group*, main effects for session, $F(1,21) = 4.741$, $p = .041$, $\eta_p^2 = .184$, and familiarity, $F(1,21) = 4.727$, $p = .041$, $\eta_p^2 = .184$, indicated more positivity for CR, and in the post-session. While the Session \times Familiarity interaction did not reach significance, $F(1,21) = 3.408$, $p = .079$, visual inspection suggests a larger reversed old/new effect in the post compared to the pre-session (cf. Fig. S3). Although there was a significant Site \times Session interaction, $F(2, 46) = 4.289$, $p = .034$, $\eta_p^2 = .157$, $\epsilon = .711$, for the *control group*, neither site showed a significant effect of session when tested separately (all $F < 2.5$, all $p_s > .1$).

Follow-up session. For the *training group* only, we compared the follow-up session both to the pre- and to the post-session in separate ANOVAs. Results involving session are briefly discussed below (for more details and a Fig., cf. supplements).

P100. There was an effect of session, showing less positivity in the follow-up compared to the post-session, $F(1,21) = 5.040$, $p = .037$, $\eta_p^2 = .194$.

N170. Comparison of the pre- vs. follow-up session revealed a trend for an interaction of Site

× Session, $F(1,21) = 3.469$, $p = .077$, $\eta_p^2 = .142$. As the analogous effect was significant in the pre- vs. post-session comparison, we pursued it with separate analyses per site, which revealed greater N170s in the follow-up session at PO9 and PO10, $F(1,21) = 4.457$, $p = .047$, $\eta_p^2 = .175$, but not at P9 and P10, $F < 1$, all $p > .6$. No differences were found between the post- and the follow-up sessions, all $F_s < 2.6$, all $p_s > .1$, suggesting a sustained increase of the N170 four weeks after training.

P200. Effects for session showed a larger P200 in the follow-up session, when compared to both the pre-session, $F(1,21) = 6.775$, $p = .017$, $\eta_p^2 = .244$, and, remarkably, the post-session, $F(1,21) = 6.493$, $p = .019$, $\eta_p^2 = .236$. Figures 4 and S2 suggests that this P200 increase was topographically limited in the post-session, and even more widespread in the follow-up session. In fact, the increase in P200 in the follow-up vs. the post-session did not interact with topographical factors, all $F_s < 2.5$, all $p_s > .1$, whereas the increase in P200 in the follow-up vs. the pre-session was more prominent at P9 and P10, as indicated by an interaction of Site × Session, $F(1,21) = 4.733$, $p = .041$, $\eta_p^2 = .184$, and an effect of session only at P9 and P10, $F(1,21) = 11.596$, $p = .003$, $\eta_p^2 = .356$.

N250. We found an interaction of site and session, $F(1,21) = 7.706$, $p = .011$, $\eta_p^2 = .268$ for the pre- vs. follow-up session. Separate ANOVAs only showed a trend for less negativity in the follow-up session at P9 and P10, $F(1,21) = 3.619$, $p = .071$, $\eta_p^2 = .147$, but not at PO9 and PO10, $F < 1.2$, $p > .2$. There were no other significant effects involving session for the N250, all $F_s < 1.3$, all $p_s > .2$.

LPCa. There was less positivity in the follow-up compared to the post-session $F(1,21) = 4.628$, $p = .043$, $\eta_p^2 = .181$, suggesting that the increase directly after the training was not sustained over the 4 week waiting period. No additional differences between the follow-up and the pre- or post-session were found for LPCa or LPCb components, all $F_s < 3$, all $p_s > .1$.

3.4 Famous Faces Experiment

Analyses of this experiment were analogous to those of the Learning Experiment.

3.4.1 Behavioral data. Performance (d' , c and accuracy) did not change significantly across sessions (Table 3). For RTs, main effects of familiarity and session reflected faster responses for familiar (hits) than unfamiliar faces (CR), and slower responses in the post- than the pre-session.

Table 3: Performance in the Famous Faces Experiment. Note: Only effects involving session or group (left), and comparisons of both pre- and post-sessions with the follow-up session in the training group (right) are shown.

	Pre		Post		Factor, Test statistic	Follow Up Training	Factor, Test statistic
	Training	Control	Training	Control			
d'	1.45 (0.45)	1.56 (0.63)	1.45 (0.51)	1.56 (0.75)	Session $F(1,46) < 1, p > .9$ Session \times Group $F(1,46) < 1, p > .9$ Group $F(1,46) < 1, p > .4$	1.49 (0.66)	pre vs. follow-up $t(23) = -.292, p > .7$ post vs. follow-up $t(23) = -.329, p > .7$
c	0.04 (0.40)	-0.14 (0.35)	0.14 (0.44)	-0.10 (0.35)	Session $F(1,46) = 2.289,$ $p > .1;$ Session \times Group $F(1,46) < 1, p > .5$ Group $F(1,46) = 4.270,$ $p = .044; \eta_p^2 = .085$	0.08 (0.54)	pre vs. follow-up $t(23) = -.426, p > .6$ post vs. follow-up $t(21) = .672, p > .5$
Acc					Session		pre vs. follow-up
Familiar	0.740 (0.121)	0.802 (0.109)	0.705 (0.148)	0.789 (0.124)	Session $F(1,46) < 1, p > .7;$ Session \times Group	0.720 (0.147)	Session $F(1,23) < 1, p > .4$
Un-familiar	0.751 (0.147)	0.715 (0.161)	0.773 (0.140)	0.729 (0.158)	Group $F(1,46) < 1, p > .4$	0.751 (0.157)	post vs. follow-up Session $F(1,23) < 1, p > .7$
RT (ms)					Session,		pre vs. follow-up
Hits	995.3 (94.4)	938.0 (118.0)	1019.0 (119.6)	938.5 (109.2)	$F(1,46) = 4.820,$ $p = .033, \eta_p^2 = .095;$ Session \times Group	1011.9 (99.9)	Session $F(1,23) = 1.701,$ $p > .2$
CR	1188.6 (129.7)	1159.8 (121.9)	1226.9 (153.4)	1192.3 (138.6)	Group $F(1,46) < 1, p > .5$ $F(1,46) = 3.276,$ $p = .077; \eta_p^2 = .066$	1216.4 (137.5)	post vs. follow-up Session $F(1,23) < 1, p > .6$

3.4.2 EEG. No ERP component showed a main effect of group (all F s < 2.3 , all p s $> .1$), with the only exception of a more positive LPCb component for the *control* compared to the *training* group, $F(1, 46) = 4.780, p = .034, \eta_p^2 = .094$ (Fig. S5).

P100. We found an interaction of Session \times Group, $F(1,46) = 4.368, p = .042, \eta_p^2 = .087$. Separate ANOVAs per group did not reveal an effect of session for the *training* group, $F(1,23) = 1.224, p = .280$. For the *control* group there was a trend for session, $F(1,23) = 4.191, p = .052, \eta_p^2 = .152$, with slightly smaller amplitudes in the post- compared to the pre-session.

N170. An effect of familiarity, $F(1,46) = 6.237, p = .016, \eta_p^2 = .119$, indicated larger N170 amplitudes for hits compared to CR. The main effect of session, $F(1,46) = 4.491, p = .039, \eta_p^2 = .089$, was further qualified by an interaction with group, $F(1,46) = 6.699, p = .013, \eta_p^2 = .127$. Separate ANOVAs revealed an effect of session, with larger N170s in the post- than the pre-session in the *training* group only, $F(1,23) = 9.321, p = .006, \eta_p^2 = .288$. No effects involving session were found in the *control* group, $F(1,23) < 1, p > .7$ (Fig. 7 & 8).

P200. A main effect of familiarity, $F(1,46) = 20.324, p < .001, \eta_p^2 = .306$, was qualified by

two interactions: Site \times Familiarity, $F(1,46) = 4.237, p = .045, \eta_p^2 = .084$, and Hemisphere \times Familiarity \times Group, $F(1,46) = 7.793, p = .008, \eta_p^2 = .145$. There was also an interaction of Site \times Session \times Group, $F(1,46) = 4.912, p = .032, \eta_p^2 = .096$. Separate ANOVAs revealed significant main effects of familiarity for both groups, with more positivity for CR than hits (*training group*: $F(1,23) = 4.842, p = .038, \eta_p^2 = .174$; *control group*: $F(1,23) = 17.201, p < .001, \eta_p^2 = .428$). For the *training group*, there was also an interaction of Site \times Session, $F(1,23) = 4.758, p = .040, \eta_p^2 = .171$, but separate analyses per site did not yield significant effects for session, all $F_s < 2.6$, all $p_s > .1$. To follow up on the Site \times Familiarity interaction, separate ANOVAs were performed per site. Although familiarity effects appeared to be more prominent at the more anterior sites (cf. Fig. 6), larger positivity for CR vs. hits was seen both at P9 and P10, $F(1,46) = 33.190, p < .001, \eta_p^2 = .419$, and at PO9 and PO10, $F(1,46) = 6.425, p = .015, \eta_p^2 = .123$.

N250. A main effect of familiarity, $F(1,44) = 13.301, p = .001, \eta_p^2 = .224$, was qualified by interactions of Familiarity \times Group, $F(1,46) = 5.071, p = .029, \eta_p^2 = .099$, and Site \times Familiarity, $F(1,46) = 4.814, p = .033, \eta_p^2 = .095$. There was also an interaction of Site \times Session \times Group, $F(1,46) = 6.911, p = .012, \eta_p^2 = .131$. Separate ANOVAs per group showed a Site \times Session interaction, $F(1,23) = 6.006, p = .022, \eta_p^2 = .207$, for the *training group*. Effects for session were absent at P9 and P10, $F(1,23) < 1, p > .8$, but emerged at PO9 and PO10, $F(1,23) = 4.611, p = .043, \eta_p^2 = .167$, with larger negativity in the post- vs. the pre-session. For the *control group* we found a main effect of familiarity, with larger negativity for hits vs. CR, $F(1,23) = 18.484, p < .001, \eta_p^2 = .446$, but no effects involving session, all $F_s < 1.6$, all $p_s > .2$. Across groups, separate ANOVAs for each site revealed larger negativity for hits vs. CR at P9 and P10, $F(1,46) = 22.665, p < .001, \eta_p^2 = .325$, and a trend at PO9 and PO10, $F(1,46) = 3.044, p = .088, \eta_p^2 = .061$.

Taken together, training-induced enhancements in both the N170 and the N250 also generalized to the Famous Faces Experiment, please see Figure 9 and Table 4 for further analyses. In contrast to the Learning Experiment, we observed (1) the usual pattern of larger N250 for familiar vs. unfamiliar faces, and (2) no prominent training-induced changes in the P200.

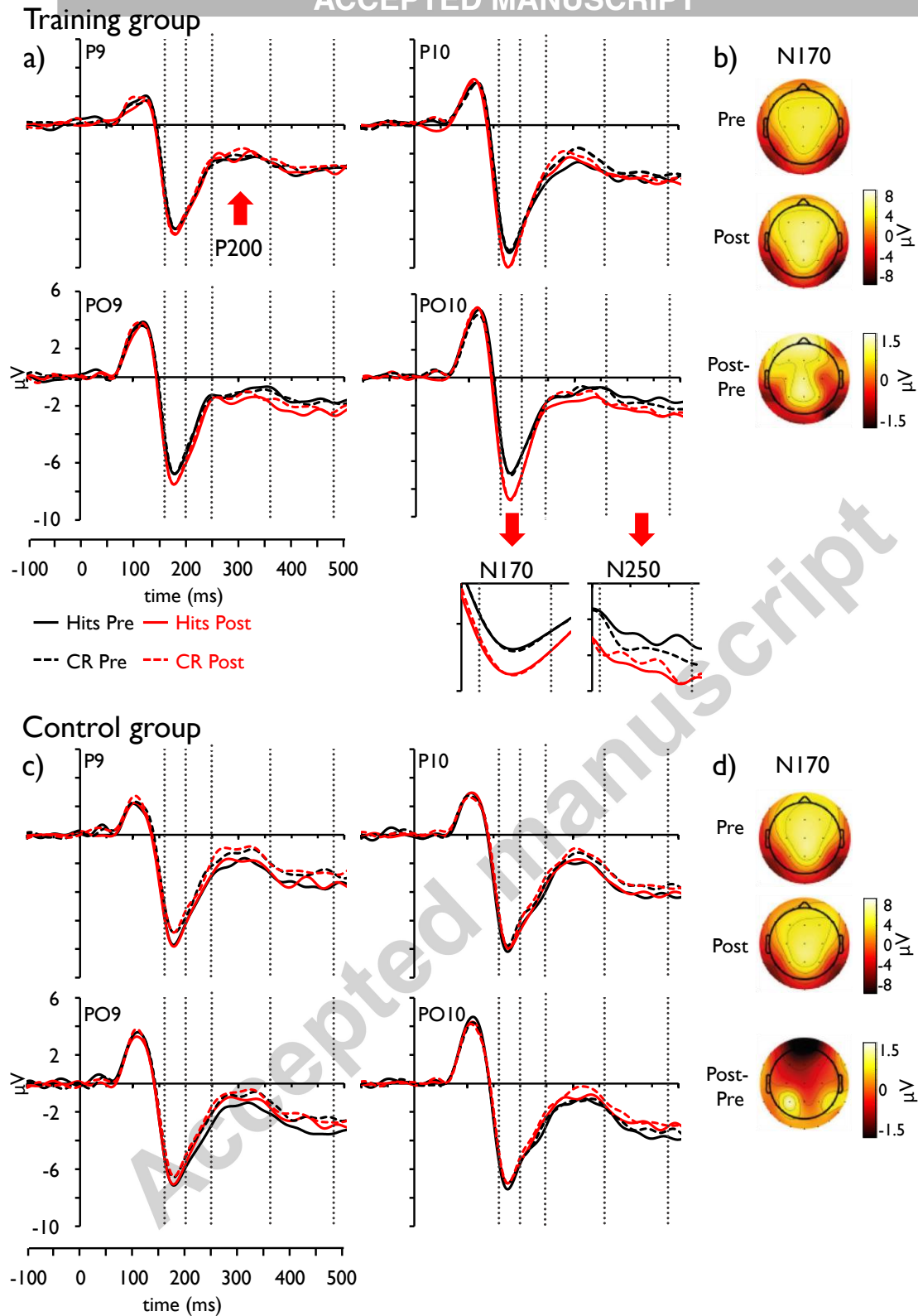


Figure 7: ERPs in the Famous Faces Experiment. a + c) ERPs at occipito-temporal electrodes for training and control group separately for hits and correct rejections in each session. b + d) Topographical voltage maps of for the N170 time window (160-200 ms) in each session, averaged across hits and CR. Maps of differences between post and pre-session illustrate the increased right posterior N170 in the training group.

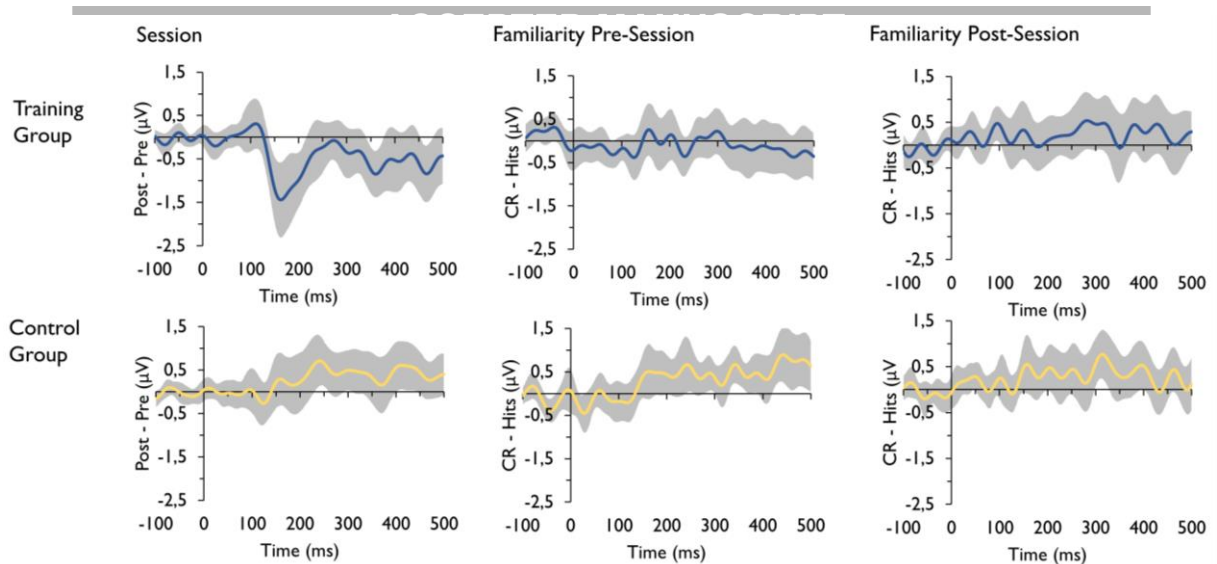


Figure 8. ERP Difference waves in the Famous Faces Experiment. Results are shown for averages across PO9 and PO10. Difference waves surrounded by 95% confidence intervals for main effects of session and familiarity. Training group is shown in blue, control group is shown in yellow. Effects of session are calculated by subtracting the average waveforms (hits and CR) from the pre-session from those from the post-session. Familiarity effects are shown separately per session and calculated as (CR – hits).

Table 4: Planned comparisons for most relevant effects in the Famous Faces experiment. Results are shown averaged across PO9 and PO10. Please refer to Figure 9 for a graphical depiction of these results.

	Training Group					Control Group				
	M_{diff} (S_{diff})	t (df)	p	Cohen's d_{av}	95% CI	M_{diff} (S_{diff})	t (df)	p	Cohen's d_{av}	95% CI
N170 Session	-1.24 (1.87)	-3.25 (23)	.004**	-0.336	-0.577 -0.113	0.20 (1.46)	0.67 (23)	.512	0.058	-0.119 0.238
N250 Session	-0.61 (1.57)	-1.90 (23)	.07(*)	-0.238	-0.507 0.020	0.41 (1.25)	1.606 (23)	.122	0.153	-0.042 0.354
N250 Session × Familiarity	0.480 (1.84)	1.28 (23)	.213	0.358	-0.204 0.913	-0.30 (1.32)	-1.10 (23)	.281	-0.276	-0.794 0.230

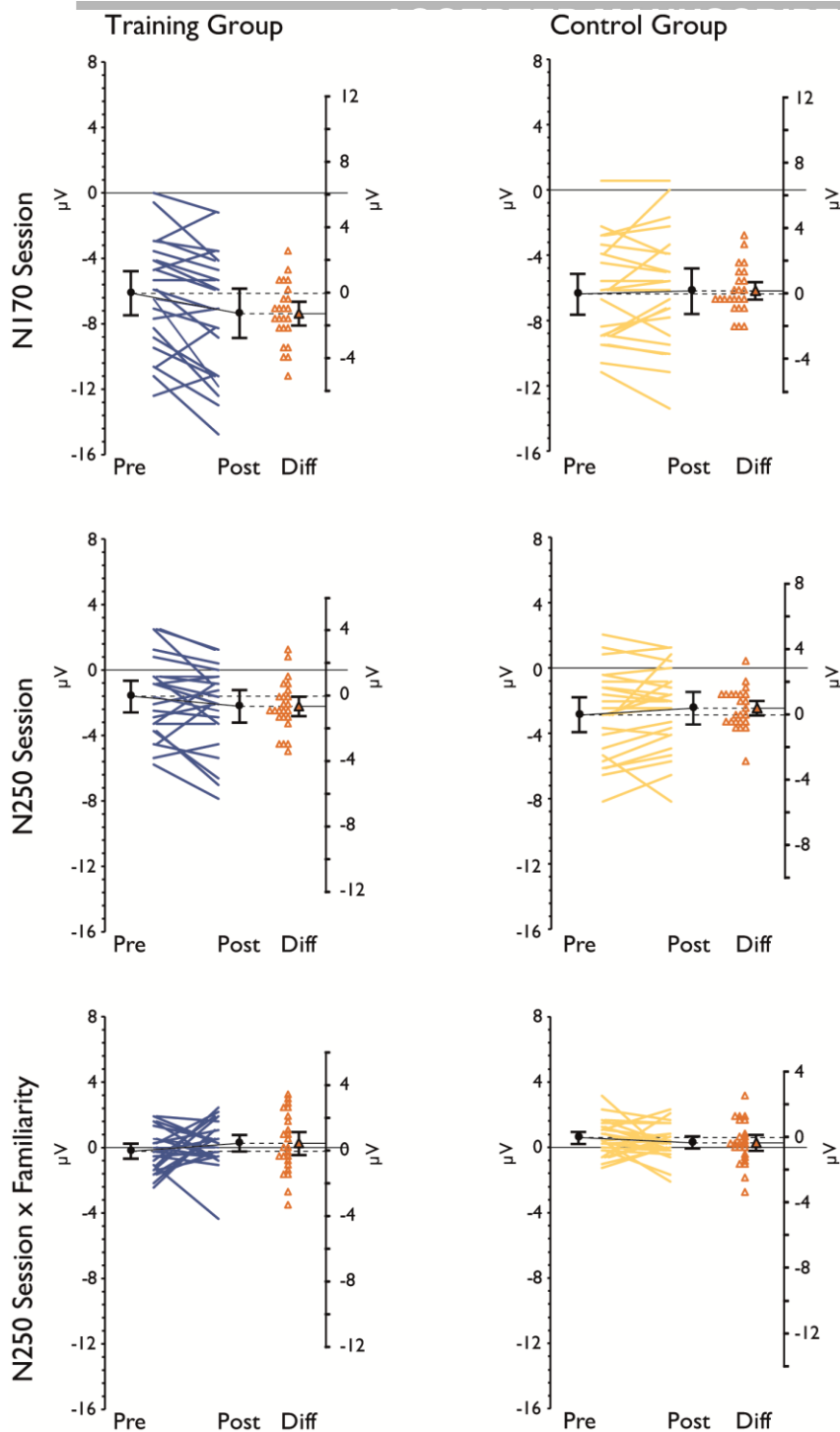


Figure 9. ERP effects of session in the Famous Faces Experiment. Results are shown for averages across PO9 and PO10, the training group is depicted on the left in blue and the control group on the right in yellow. Lines represent single participant data (pre- and post-session). Orange triangles show differences between sessions (please note separate scale to the right). Filled circles show mean per session and filled triangles show mean differences. Error bars depict 95% confidence intervals. Significant differences between sessions were found for the N170 in the training group only (and a trend for the N250 in the training group). Figures were made with ESCI intro software (Cumming & Calin-Jageman, 2017).

LPCa. We found an interaction of Site \times Familiarity, $F(2,92) = 8.118, p = .001, \eta_p^2 = .150$, and several interactions involving group: Session \times Group, $F(1,46) = 8.726, p = .005, \eta_p^2 = .159$, Familiarity \times Group, $F(1,46) = 6.240, p = .016, \eta_p^2 = .119$, Laterality \times Session \times Group, $F(2,92) =$

4.189, $p = .018$, $\eta_p^2 = .0839$, and Session \times Familiarity \times Group, $F(1,46) = 6.939$, $p = .011$, $\eta_p^2 = .131$. Separate analyses per group revealed an effect of session in the *training group*, $F(1,23) = 5.507$, $p = .028$, $\eta_p^2 = .193$, with larger positivity in the post-session. The *control group* exhibited a main effect of familiarity, $F(1,23) = 7.755$, $p = .011$, $\eta_p^2 = .252$, with larger positivity for hits vs. CR, and an interaction of Laterality \times Session, $F(2,46) = 4.276$, $p = .020$, $\eta_p^2 = .157$. Follow-up analysis per laterality revealed an effect of session, with smaller positivity in the post-session at midline electrodes only, $F(1,23) = 7.151$, $p = .014$, $\eta_p^2 = .237$, but not at the left or right electrodes, all $F_s < 1$, all $p_s > .3$. Across groups, separate ANOVAs per site revealed effects for familiarity only at frontal sites, $F(1,46) = 16.215$, $p < .001$, $\eta_p^2 = .261$, with larger positivity for hits than CR, but not at parietal or central sites, all $F_s < 1$, all $p_s > .5$.

LPCb. The analysis yielded interactions of Session \times Group, $F(1,46) = 5.191$, $p = .027$, $\eta_p^2 = .101$, Site \times Session, $F(1.597, 73.469) = 6.297$, $p = .006$, $\eta_p^2 = .120$, and Site \times Familiarity, $F(2,92) = 4.083$, $p = .020$, $\eta_p^2 = .082$. Separate ANOVAs revealed a trend for more positivity in the post-session for the *training group*, $F(1,23) = 3.888$, $p = .061$, $\eta_p^2 = .145$, but not in the *control group*, $F(1,23) = 1.520$, $p = .230$. Across groups, separate ANOVAs per site yielded an effect of session at parietal sites only, $F(1, 46) = 9.743$, $p = .003$, $\eta_p^2 = .175$, with larger positivity in the post- compared to the pre-session. While no effects involving the factors session or familiarity were found at central sites, all $F_s < 1$, all $p_s > .9$, a significant effect of familiarity, with larger positivity for hits than CR, $F(1, 46) = 6.507$, $p = .014$, $\eta_p^2 = .124$, was found at frontal sites (cf. Fig. S5).

Follow-up session. N170. We found an increase in the follow-up compared to the pre-session, $F(1,23) = 6.808$, $p = .016$, $\eta_p^2 = .228$, which was qualified by an interaction of Site \times Hemisphere \times Session, $F(1,23) = 7.274$, $p = .013$, $\eta_p^2 = .240$. Separate ANOVAs per electrode revealed significantly larger N170 amplitudes in the follow-up session at PO10, $F(1,23) = 7.943$, $p = .010$, $\eta_p^2 = .257$, and a trend in the same direction at P9, $F(1,23) = 3.466$, $p = .075$, $\eta_p^2 = .131$. No effects of session were found at PO9 or P10, all $F_s < 2.6$, all $p_s > .1$. We did not observe an effect of session when comparing the post- to the follow-up session, $F(1,23) < 1$, $p > .5$, again suggesting a sustained increase in N170 amplitude four weeks after training (Fig. S4).

especially for CR over the left hemisphere. In fact, there were interactions of Hemisphere \times Session \times Familiarity, $F(1,23) = 7.039, p = .014, \eta_p^2 = .234$, and of Site \times Hemisphere \times Session \times Familiarity, $F(1,23) = 6.731, p = .016, \eta_p^2 = .226$. Separate ANOVAs per electrode revealed an effect of Familiarity, $F(1,23) = 10.848, p = .003, \eta_p^2 = .320$, and an interaction of Session \times Familiarity, $F(1,23) = 6.808, p = .016, \eta_p^2 = .228$, at P9. Separate ANOVAs suggested a trend for smaller negativity in the follow-up session for CR, $F(1,23) = 3.828, p = .063, \eta_p^2 = .143$, but not for hits, $F(1,23) < 1, p > .8$. At P10 we observed trends for session, $F(1,23) = 3.572, p = .071, \eta_p^2 = .134$, familiarity, $F(1,23) = 3.400, p = .078, \eta_p^2 = .129$, as well as for an interaction of Session \times Familiarity $F(1,23) = 3.063, p = .093, \eta_p^2 = .118$, suggesting smaller negativity for CR and in the follow-up session. No effects involving session were found at PO9 or PO10, all F s < 1 , all p s $> .4$. An analysis including data from of the post- and the follow-up session only revealed trends for interactions of Hemisphere \times Site \times Session, $F(1,23) = 3.156, p = .089, \eta_p^2 = .121$, and of Hemisphere \times Session \times Response, $F(1,23) = 3.261, p = .084, \eta_p^2 = .124$.

N250. We found a significant interaction of Site \times Hemisphere \times Session \times Familiarity, $F(1,23) = 7.004, p = .014, \eta_p^2 = .233$, for the follow-up compared to the pre-session. Separate ANOVAs revealed a significant interaction of Session \times Familiarity, $F(1,23) = 7.390, p = .012, \eta_p^2 = .243$, at P9, with larger negativity for hits compared to CR in the follow-up session only, pre: $t(23) = -.619, p_{corr} > .9$; follow-up: $t(23) = -5.434, p_{corr} < .001, d_{av} = .44$. No corresponding effects were found at the other electrodes, all F s < 1.3 , all p s > 2 . No difference was found between the post- and the follow-up session.

For the LPC measures, there were not significant effects involving the factor session, all F s < 3.5 , all p s $> .075$.¹

¹ Due to the number of statistical tests reported in the previous sections, a potential concern is that inflated Type I error rates could have influenced some of the present findings. However, it should be noted that we tested several specific hypotheses derived from previous research. More crucially, the observed patterns of ERP training effects in the Face Learning and Famous Faces Experiments were remarkably similar, even though the two experiments represent independent data sets, making it highly unlikely that such similar and specific patterns across two experiments simply reflect Type I error. We furthermore report planned comparisons for the most relevant effects, please see Figures 6 and 9 and Tables 2 and 4. In response to a reviewer comment, we nevertheless performed an additional data-driven analysis using the LIMO EEG toolbox (Pernet et al., 2011) which compared training effects for the entire set of electrodes and time-points while controlling for multiple

We compared face processing performance and its electrophysiological correlates before and after extensive training (12 1-h sessions in 4 weeks) using photorealistic digital face caricatures in older adults. Although our results showed no improvement in performance beyond training, training-induced changes in face-sensitive ERP components indicated plasticity of face processing mechanisms even in older adults. Moreover, training induced increases of the N170, P200 and N250 components over occipito-temporal cortex were sustained in a follow-up session four weeks after training.

The N170 is related to the categorization of stimuli as faces, and their structural encoding (Bentin and Deouell, 2000; Eimer, 2011). We observed increased N170 responses after caricature training (but not in a no-treatment control group) not only in a Face Learning, but also in a Famous Faces Recognition experiment. Former studies suggested that gaining visual expertise with particular categories of stimuli (e.g. unfamiliar script, animals, objects or faces) can induce enhanced occipito-temporal N170 or N1 responses (Gauthier et al., 2003; Maurer et al., 2010; Scott et al., 2006; 2008; Tanaka and Pierce, 2009; Wang and Bingo, 2010). Because these effects were observed in basic, subordinate and exposure training, Scott et al. (2006; 2008) suggested that this N170 increase was merely due to increased exposure with stimuli from one category per se. However, the effect reported by Scott et al. vanished after one week, whereas the present caricature training-induced N170 effects were sustained for at least four weeks after training. Note also that the present N170 effects were highly consistent across, and independent of, the specific faces used.

These N170 increases could indicate a sustained enhancement of face detection mechanisms after training. Of note, enhanced N170 selectivity to faces was shown also after extensive holistic training in participants with developmental prosopagnosia (DeGutis et al., 2014). Accordingly, tasks focusing on face detection might have revealed concomitant performance improvements that were not evident in our face learning and recognition tasks. Alternatively, the increased N170 could indicate (compensatory) recruitment of additional object-selective neurons. Larger N170 responses are

comparisons. We note that this analysis did not reveal significant effects, suggesting that under the specific conditions of the present study, training effects may have been too small and specific to emerge in this purely data-driven analysis. However, acceptance of the null hypothesis from this data-driven analysis approach would appear overly conservative and inappropriate in the present context of hypotheses (directed at effects with specific topographies at specific time segments) and two experiments with highly similar results.

prevalent for inverted faces (Eimer, 2000a; Rossion et al., 2000; Sagiv and Bentin, 2001) as well as in older adults (Daniel and Bentin, 2012; Gao et al., 2009; Rousselet et al., 2009). One explanation for the N170 inversion effect is that inverted faces might recruit eye-sensitive (Itier and Batty, 2009) or object-sensitive neurons (Rossion et al., 2000) in addition to face-sensitive cells recruited by upright faces. Supporting this idea, both fMRI (Yovel and Kanwisher, 2005) and intracranial recordings (Rosburg et al., 2010) suggest increased activation of object-selective areas of the lateral occipital complex by inverted faces. Decreased differentiation and neural selectivity has also been shown for perceptual areas in older adults (Goh et al., 2010 ; Park et al., 2004), with evidence for compensatory activity in left inferior occipital gyrus, frontal, and parietal regions in high performing older adults during face processing (Lee et al., 2011). Accordingly, such additional recruitment could explain both general N170 enhancements in older compared to younger participants, and specific training-induced enhancements.

We also found enhancements of the P200, especially over the left hemisphere and in the Learning Experiment after the training. Compared to the post-session, P200 enhancements became even more widespread and pronounced at follow-up. Increases in left P200 responses after categorization training could potentially reflect enhanced facial feature integration (Tanaka et al., 2009). Alternatively, P200 amplitude may be sensitive to the perceived typicality of faces, as shown for own- vs. other-race faces (Stahl et al., 2008, 2010), for young vs. old faces in young participants (Wiese et al., 2008), for veridical vs. caricatured versions of a face (Kaufmann and Schweinberger, 2012; Schulz et al., 2012), or for perceptual renormalization of distorted faces during adaptation (Kloth et al., 2017). As our training included both veridical faces and caricatures, participants also may have learned to better discriminate between these (more or less typical) categories of faces. A larger P200 after training could then indicate stronger perception of typicality of the (exclusively veridical) faces in the EEG experiments. The same mechanism might explain why prominent training-induced P200 effects were found in the Learning Experiment only: Faces both in the Learning Experiment and during training depicted young adults, whereas the Famous Faces Experiment included a greater age range. The P200 effects might thus reflect enhanced typicality of specific (here: young adult) face categories.

ACCEPTED MANUSCRIPT

In the Famous Faces Experiment, we observed an effect of familiarity in the P200 across both groups, with smaller positivity (or larger negativity) for hits than correct rejections. Similar recognition effects in the P200 were reported before (e.g. Wiese et al., 2008). This suggests that identity processing of familiar faces can already start at this early stage (Schweinberger and Neumann, 2016), although it is more consistently indexed by the N250 familiarity effect (e.g. Kaufmann et al., 2009; Schweinberger, 2011; Tanaka et al., 2006), in which increased negativity for familiar faces is indicative of recognition. In the present study, we found this pattern in the Famous Faces Experiment only, with a reversed effect in the Learning Experiment. Reversed or absent N250 familiarity effects in older participants were reported before (e.g. Wiese et al., 2008, 2012) and might reflect difficulties in accessing perceptual representations of newly acquired faces. Our original hypothesis that face training might increase N250 familiarity effects was not confirmed (except for the N250 at P9 in the Famous Faces Experiment which exhibited a significant difference between familiar and unfamiliar faces in the follow-up session only, with no corresponding familiarity effect in the pre-session). Combined with null results in performance and the training-induced enhancements of N170 and P200 responses, these findings suggest that while the present training facilitated face categorization and expertise for (specific groups of) faces, it did not modulate processes of individual recognition, neither for newly learned nor for pre-experimentally familiar faces.

Although training did not induce enhanced N250 familiarity effects, it did induce substantial overall increases of N250 amplitudes across sessions in the training group only – mirroring the previously described N170 enhancements. This pattern is reminiscent of findings by Scott et al. (2006; 2008) and Tanaka and Pierce (2009): For objects or faces, their participants learned to classify different exemplars at either a basic (e.g., wading bird, owl) or a subordinate level (e.g., egret, snowy owl). Compared to pre-training, both types of training induced enhanced N170s, whereas only subordinate-level training induced enhanced N250s. Our data might thus suggest that the present training increased both categorical and individual (subordinate) face processing mechanisms.

Cognitive abilities can be differentially affected by age-related declines (Baltes, 1987; Cattell, 1971; Craik and Bialystok, 2006). Older adults may show impaired episodic memory, but preserved semantic memory (Ofen and Shing, 2013, see also Wiese et al., 2017). Intriguingly, our effects in the

ACCEPTED MANUSCRIPT

LPC differed between the Learning Experiment (which involved newly formed perceptual representations and strong episodic contributions), and the Famous Faces Experiment (which presumably involved relatively more semantic contributions): Whereas a typical pattern of a larger LPC for familiar vs. unfamiliar items was found in the Famous Faces Experiment, a reversed old/new effect emerged in the Learning Experiment. Reversed old/new effects were previously reported for older adults (Duarte et al., 2006), especially when recognizing newly learned faces (Wiese et al., 2012), and when source memory was important (Duarte et al., 2006; Li et al., 2004; Swick et al., 2006). Reversed old/new effects might signal attempts to compensate for memory deficits, thereby recruiting different retrieval processes in older compared to young adults (Friedman et al., 2010; Nessler et al., 2007). For the training group, the LPC was enhanced in the post-session, and there was also a trend for an enhanced reversed old/new effect in the post- vs. the pre-session. This finding is not only consistent with the hypothesis of enhanced compensatory processes, but could also reflect training-induced contact to specific groups of faces. A previous study found the reversal of the old/new effect to be particularly prominent for older adults with high levels of daily contact to other older people (Wiese et al., 2012). We therefore hypothesize that the present training-induced LPC enhancement in the Learning Experiment is related to acquired expertise with specific groups of (i.e. young adult) faces – an idea that receives support from the finding that similar training-induced LPC modulations were absent in the Famous Faces Experiment. The LPC in the Famous Faces Experiment only exhibited a typical familiarity effect across sessions, with larger LPC for familiar than unfamiliar faces.

Despite consistent training-induced effects in neural correlates of face processing, concomitant changes in performance were neither observed in the diagnostic tests nor in the EEG experiments. While this could mean that the present behavioral tests were not sufficiently sensitive, another interpretation is that our current tests focused on aspects (of the recognition of individual identities) that were not improved by training. Instead, the training-induced ERP changes suggest enhancements in face detection and expertise for faces. Of relevance, a recent MEG study reports that extensive category training for simple artificial face groups induced both enhancements in behavioral categorization performance and stronger early occipito-temporal activity (Kietzmann et al., 2016).

ACCEPTED MANUSCRIPT

Future research should reveal whether behavioral tasks that focus on these aspects may reveal training-induced performance enhancements. Alternatively, it remains possible that face recognition cannot be (much) improved by training, or that our training was insufficiently intense.

Nevertheless, our training-induced changes in face-sensitive ERPs challenge recent behavioral approaches to face training, which suggest a potential for training only for individual face identities (Dowsett et al., 2016). Dowsett et al. focused on face matching, and despite using a wide array of stimuli, they only investigated transfer effects between three identities. More intense trainings as used here may be necessary to evoke effects that transfer to individual recognition of untrained faces, as we show that it is possible to modify generic face processing mechanisms via training. Our results concur with a recent review by De Gutsis et al. (2014), stressing potentials of perceptual training to improve face processing in individuals with prosopagnosia, and showing that training effects are transferrable to novel face stimuli. Compared to their training, which focused on second-order spatial configurations, our approach of caricaturing may be more relevant for familiar face processing. Although spatial information plays a role for the initial encoding of unfamiliar faces (Itz et al., 2014), good face recognizers are often *less* sensitive to second-order configurations (Itz et al., 2018). Crucially, and in contrast to relatively widespread belief, spatial configurations are remarkably unimportant for the recognition of familiar or famous faces (for an overview, see Burton et al., 2015).

Although we did not include an active control group, it seems unlikely that the present findings reflect general effects such as motivation or attention. We observed no prominent effects for the P100, an early and attention-sensitive visual response (Luck, 2005). Our main findings concern face-sensitive components that were argued to be relatively insensitive to attention (Cauquil et al., 2000; Neumann et al., 2011; Neumann and Schweinberger, 2009). Having established training-induced enhancements in face-sensitive neural responses, we expect that future research will further refine these findings. For instance, the specific role of facial caricature training may be further assessed including active control groups conducting memory training for non-face stimuli, and the specific nature of the present training effects for face as opposed to object processing (e.g. Dolzycka et al., 2014) could be assessed by additional tests. As we only used faces as stimuli, we cannot fully

exclude the possibility that our training effects may generalize to other stimulus categories as well.

However, this seems unlikely when considering that training did not elicit any effects in the P100.

In conclusion, intensive training with photorealistic caricatures focusing on matching and identification of previously unfamiliar faces elicits changes in cortical processing in older adults. Importantly, these changes transfer to other faces. Although we found no improvements in face recognition performance, the substantial and consistent training-induced changes in cortical face processing suggest that early detection of, and expertise for faces can be enhanced by training.

Acknowledgements:

We gratefully acknowledge the help of Kathrin Rauscher, Bettina Kamchen, Anja Schinkel and Franziska Martin during data collection and of Marlena L. Itz during planning. This research was supported by a grant from the Bundesministerium für Bildung und Forschung (BMBF), IRESTRA project (FKZ: 16SV7210) to SRS.

Author Note

This research was supported by a grant from the Bundesministerium für Bildung und Forschung (BMBF), IRESTRA project (FKZ: 16SV7210) to SRS. The funding body had no influence in study design, collection, analysis and interpretation of data, writing of the report or decision to submit the article for publication. Declarations of interests: none.

- Alenezi, H.M., Bindemann, M., 2013. The Effect of Feedback on Face-Matching Accuracy. *Applied Cognitive Psychology* 27, 735-753.
- Bach, M., 1996. The Freiburg Visual Acuity test--automatic measurement of visual acuity. *Optometry and Vision Science* 73, 49-53.
- Baltes, P.B., 1987. Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology* 23, 611-626.
- Başar, E., Özgören, M., Öniz, A., Schmiedt, C., Başar-Eroğlu, C., 2007. Brain oscillations differentiate the picture of one's own grandmother. *International Journal of Psychophysiology* 64, 81-90.
- Bentin, S., Deouell, L.Y., 2000. Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cognitive Neuropsychology* 17, 35-55.
- Boutet, I., Faubert, J., 2006. Recognition of faces and complex objects in younger and older adults. *Memory & Cognition* 34, 854-864.
- Burton, A.M., Bruce, V., 1993. Naming faces and naming names: Exploring an interactive activation model of person recognition. *Memory* 1, 457-480.
- Burton, A.M., Schweinberger, S.R., Jenkins, R., Kaufmann, J.M., 2015. Arguments Against a Configural Processing Account of Familiar Face Recognition. *Perspectives on Psychological Science* 10, 482-496.
- Burton, A.M., White, D., McNeill, A., 2010. The Glasgow Face Matching Test. *Behavior Research Methods* 42, 286-291.
- Cattell, R.B., 1971. *Abilities: their structure, growth, and action*. Houghton Mifflin, Boston.
- Cauquil, A.S., Edmonds, G.E., Taylor, M.J., 2000. Is the face-sensitive N170 the only ERP not affected by selective attention? *Neuroreport* 11, 2167-2171.
- Chaby, L., George, N., Renault, B., Fiori, N., 2003. Age-related changes in brain responses to personally known faces: an event-related potential (ERP) study in humans. *Neuroscience letters* 349, 125-129.

- Chaby, L., Narme, P., 2009. Processing facial identity and emotional expression in normal aging and neurodegenerative diseases. *Psychol Neuropsychiatr Vieil* 7, 31-42.
- Cumming, G. & Calin-Jageman R.J. (2017). *Introduction to the New Statistics: Estimation, Open Science, and Beyond*. New York: Routledge.
- Cohen, G., Burke, D.M., 1993. Memory for proper names: A review. *Memory* 1, 249-263.
- Craik, F.I.M., Bialystok, E., 2006. Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences* 10, 131-138.
- Crook, T.H., 3rd, Larrabee, G.J., 1992. Changes in facial recognition memory across the adult life span. *J Gerontol* 47, 138-141.
- Daniel, S., Bentin, S., 2012. Age-related changes in processing faces from detection to identification: ERP evidence. *Neurobiology of Aging* 33, 206.e201-206.e228.
- DeGutis, J.M., Chiu, C., Grosso, M.E., Cohan, S., 2014a. Face processing improvements in prosopagnosia: successes and failures over the last 50 years. *Frontiers in Human Neuroscience* 8.
- DeGutis, J.M., Cohan, S., Nakayama, K., 2014b. Holistic face training enhances face processing in developmental prosopagnosia. *Brain* 137, 1781-1798.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134, 9-21.
- Dewhurst, S.A., Hay, D.C., Wickham, L.H.V., 2005. Distinctiveness, typicality, and recollective experience in face recognition: A principal components analysis. *Psychonomic Bulletin & Review* 12, 1032-1037.
- Dolzycka, D., Herzmann, G., Sommer, W., Wilhelm, O., 2014. Can Training Enhance Face Cognition Abilities in Middle-Aged Adults? *Plos One* 9, e90249.
- Dowsett, A.J., Sandford, A., Burton, A.M., 2016. Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly journal of experimental psychology* 69, 1-10.
- Duarte, A., Ranganath, C., Trujillo, C., Knight, R.T., 2006. Intact Recollection Memory in High-performing Older Adults: ERP and Behavioral Evidence. *Journal of Cognitive Neuroscience* 18, 33-47.
- Eimer, M., 2000a. Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain research. Cognitive brain research* 10, 145-158.
- Eimer, M., 2000b. Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology* 111, 694-705.
- Eimer, M., 2011. The face-sensitive N170 component of the event-related brain potential. In: Calder, A.J., Rhodes, G., Johnson, M.H., Haxby, J.V. (Eds.), *The Oxford Handbook of Face perception*. University Press, Oxford, pp. 329-344.
- Friedman, D., de Chastelaine, M., Nessler, D., Malcolm, B., 2010. Changes in familiarity and recollection across the lifespan: An ERP perspective. *Brain Research* 1310, 124-141.
- Gao, L., Xu, J., Zhang, B., Zhao, L., Harel, A., Bentin, S., 2009. Aging effects on early-stage face perception: an ERP study. *Psychophysiology* 46, 970-983.
- Gauthier, I., Curran, T., Curby, K.M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience* 6, 428-432.
- Germine, L.T., Duchaine, B., Nakayama, K., 2011. Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* 118, 201-210.
- Goh, J.O., Suzuki, A., Park, D.C., 2010. Reduced neural selectivity increases fMRI adaptation with age during face discrimination. *Neuroimage* 51, 336-344.
- Going, M., Read, J.D., 1974. Effects of Uniqueness, Sex of Subject, and Sex of Photograph on Facial Recognition. *Perceptual and motor skills* 39, 109-110.
- Halit, H., de Haan, M., Johnson, M.H., 2000. Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport* 11, 1871-1875.
- Hildebrandt, A., Sommer, W., Herzmann, G., Wilhelm, O., 2010. Structural invariance and age-related performance differences in face cognition. *Psychol Aging* 25, 794-810.
- Hildebrandt, A., Wilhelm, O., Schmiedek, F., Herzmann, G., Sommer, W., 2011. On the specificity of face cognition compared with general cognitive functioning across adult age. *Psychol Aging* 26, 701-715.
- Hussain, Z., Sekuler, A.B., Bennett, P.J., 2009. Perceptual learning modifies inversion effects for faces and textures. *Vision Research* 49, 2273-2284.
- Irons, J., McKone, E., Dumbleton, R., Barnes, N., He, X., Provis, J., Ivanovici, C., Kwa, A., 2014. A new theoretical approach to improving face recognition in disorders of central vision: face caricaturing. *Journal of Vision* 14.
- Itier, R.J., Batty, M., 2009. Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews* 33, 843-863.
- Itz, M.L., Schweinberger, S.R., Kaufmann, J.M., 2017. Caricature generalization benefits for faces learned with enhanced idiosyncratic shape or texture. *Cogn Affect Behav Neurosci*, 185-197.
- Itz, M.L., Schweinberger, S.R., Kaufmann, J.M., 2018. Familiar face priming: The role of second-order configuration and individual face recognition abilities. *Perception* 47(2), 185-196. doi: 10.1177/0301006617742069.

- Itz, M.L., Schweinberger, S.R., Schulz, C., Kaufmann, J.M., 2014. Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. *Neuroimage* 102 Pt 2, 736-747.
- Kaufmann, J.M., Schulz, C., Schweinberger, S.R., 2013. High and low performers differ in the use of shape information for face recognition. *Neuropsychologia* 51, 1310-1319.
- Kaufmann, J.M., Schweinberger, S.R., 2012. The faces you remember: Caricaturing shape facilitates brain processes reflecting the acquisition of new face representations. *Biological Psychology* 89, 21-33.
- 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, 625-641.
- Kietzmann, T.C., Ehinger, B.V., Porada, D., Engel, A.K., König, P., 2016. Extensive training leads to temporal and spatial shifts of cortical activity underlying visual category selectivity. *Neuroimage* 134, 22-34.
- Kikutani, M., Roberson, D., Hanley, J.R., 2008. What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review* 15, 787-794.
- Kloth, N., Rhodes, G., Schweinberger, S.R., 2017. Watching the brain recalibrate: Neural correlates of renormalization during face adaptation. *Neuroimage* 155, 1-9.
- Komes, J., Schweinberger, S.R., Wiese, H., 2014. Preserved fine-tuning of face perception and memory: evidence from the own-race bias in high- and low-performing older adults. *Front Aging Neurosci* 6, 60.
- Latinus, M., Taylor, M.J., 2006. Face processing stages: Impact of difficulty and the separation of effects. *Brain Research* 1123, 179-187.
- Lee, Y., Grady, C.L., Habak, C., Wilson, H.R., Moscovitch, M., 2011. Face processing changes in normal aging revealed by fMRI adaptation. *J Cogn Neurosci* 23, 3433-3447.
- Leirer, V.O., Morrow, D.G., Sheikh, J.I., Pariante, G.M., 1990. Memory skills elders want to improve. *Experimental Aging Research* 16, 155-158.
- Li, J., Morcom, A.M., Rugg, M.D., 2004. The effects of age on the neural correlates of successful episodic retrieval: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience* 4, 279-293.
- Lucas, H.D., Chiao, J.Y., Paller, K.A., 2011. Why Some Faces won't be Remembered: Brain Potentials Illuminate Successful Versus Unsuccessful Encoding for Same-Race and Other-Race Faces. *Frontiers in Human Neuroscience* 5, 20.
- Luck, S.J., 2005. An introduction to event-related potentials and their neural origins. An Introduction to the event-related potential technique. MIT Press, Cambridge, Massachusetts, pp. 31-38.
- Macmillan, N.A., Creelman, C.D., 2005. Detection theory: A user's guide, 2nd ed. Psychology Press, New York, London.
- Maurer, U., Blau, V.C., Yoncheva, Y.N., McCandliss, B.D., 2010. Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. *Dev Neuropsychol* 35, 404-422.
- Megreya, A.M., Bindemann, M., 2015. Developmental Improvement and Age-Related Decline in Unfamiliar Face Matching. *Perception* 44, 5-22.
- Meinhardt-Injac, B., Persike, M., Meinhardt, G., 2014. Holistic processing and reliance on global viewing strategies in older adults' face perception. *Acta psychologica* 151, 155-163.
- Nessler, D., Friedman, D., Johnson, R., Jr., Bersick, M., 2007. Does repetition engender the same retrieval processes in young and older adults? *Neuroreport* 18, 1837-1840.
- Neumann, M.F., Mohamed, T.N., Schweinberger, S.R., 2011. Face and object encoding under perceptual load: ERP evidence. *Neuroimage* 54, 3021-3027.
- Neumann, M.F., Schweinberger, S.R., 2009. N250r ERP repetition effects from distractor faces when attending to another face under load: Evidence for a face attention resource. *Brain Research* 1270, 64-77.
- Ofen, N., Shing, Y.L., 2013. From perception to memory: Changes in memory systems across the lifespan. *Neuroscience & Biobehavioral Reviews* 37, 2258-2267.
- Park, D.C., Polk, T.A., Park, R., Minear, M., Savage, A., Smith, M.R., 2004. Aging reduces neural specialization in ventral visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13091-13095.
- Pernet, C. R., Chauveau, N., Gaspar, C., and Rousset, G.A., 2011. LIMO EEG: A Toolbox for Hierarchical Linear Modeling of Electroencephalographic Data. *Computational Intelligence and Neuroscience*, . 2011. <https://doi.org/10.1155/2011/831409>
- Pfütze, E.M., Sommer, W., Schweinberger, S.R., 2002. Age-related slowing in face and name recognition: evidence from event-related brain potentials. *Psychol Aging* 17, 140-160.
- Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., 2000. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1090-1104.
- Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J., 1998. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing* 16, 295-306.
- Rhodes, G., 1996. Superportraits. Caricatures and Recognition. Psychology Press, Hove (UK).
- Rosburg, T., Ludwig, E., Dimpelmann, M., Alba-Ferrara, L., Urbach, H., Elger, C.E., 2010. The effect of face inversion on intracranial and scalp recordings of event-related potentials. *Psychophysiology* 47, 147-157.

- Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linoite, S., Crommelinck, M., 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport* 11, 69-74.
- Rousselet, G.A., Husk, J.S., Pernet, C.R., Gaspar, C.M., Bennett, P.J., Sekuler, A.B., 2009. Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. *BMC Neurosci* 10, 114.
- Rugg, M.D., Schloerscheidt, A.M., Doyle, M.C., Cox, C.J.C., Patching, G.R., 1996. Event-related potentials and the recollection of associative information. *Cognitive Brain Research* 4, 297-304.
- Sagiv, N., Bentin, S., 2001. Structural encoding of human and schematic faces: holistic and part-based processes. *J Cogn Neurosci* 13, 937-951.
- Savaskan, E., Müller, S.E., Böhringer, A., Philippsen, C., Müller-Spahn, F., Schächinger, H., 2007. Age determines memory for face identity and expression. *Psychogeriatrics* 7, 49-57.
- Schulz, C., Kaufmann, J.M., Kurt, A., Schweinberger, S.R., 2012. Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces. *Neuroimage* 63, 491-500.
- Schwartz, L., Yovel, G., 2016. The roles of perceptual and conceptual information in face recognition. *J Exp Psychol Gen* 145, 1493-1511.
- Schweinberger, S.R., 2011. Neurophysiological correlates of face recognition. In: Calder, A.J., Rhodes, G., Johnson, M.H., Haxby, J.V. (Eds.), *Oxford handbook of face perception*. Oxford University Press, Oxford New York, pp. 345-366.
- Schweinberger, S.R., Burton, A.M., 2003. Covert Recognition and the Neural System for Face Processing. *Cortex* 39, 9-30.
- Schweinberger, S.R., Neumann, M.F., 2016. Repetition effects in human ERPs to faces. *Cortex* 80, 141-153.
- 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.
- Schweinberger, S.R., Pickering, E.C., Jentsch, 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.
- Scott, L.S., Tanaka, J.W., Sheinberg, D.L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. *J Cogn Neurosci* 18, 1453-1465.
- Scott, L.S., Tanaka, J.W., Sheinberg, D.L., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. *Brain Research* 1210, 204-215.
- Sporer, S.L., 1991. Deep-deeper-deepest? Encoding strategies and the recognition of human faces. *J Exp Psychol Learn Mem Cogn* 17, 323-333.
- Stahl, J., Wiese, H., Schweinberger, S.R., 2008. Expertise and own-race bias in face processing: an event-related potential study. *Neuroreport* 19, 583-587.
- Stahl, J., Wiese, H., Schweinberger, S.R., 2010. Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia* 48, 2027-2040.
- Swick, D., Senkfor, A.J., Van Petten, C., 2006. Source memory retrieval is affected by aging and prefrontal lesions: Behavioral and ERP evidence. *Brain Research* 1107, 161-176.
- 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.
- Tanaka, J.W., Pierce, L.J., 2009. The neural plasticity of other-race face recognition. *Cogn Affect Behav Neurosci* 9, 122-131.
- Vokey, J.R., Read, J.D., 1992. Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition* 20, 291-302.
- Wang, G., Bingo, J., 2010. Three-dimensional object recognition learning alters an early ERP component of N1. *Clinical Neurophysiology* 121, 1473-1480.
- White, D., Kemp, R.I., Jenkins, R., Burton, A.M., 2014. Feedback training for facial image comparison. *Psychonomic Bulletin & Review* 21, 100-106.
- Wiese, H., Kachel, U., Schweinberger, S.R., 2013. Holistic face processing of own- and other-age faces in young and older adults: ERP evidence from the composite face task. *Neuroimage* 74, 306-317.
- Wiese, H., Komes, J., Schweinberger, S.R., 2012. Daily-life contact affects the own-age bias and neural correlates of face memory in elderly participants. *Neuropsychologia* 50, 3496-3508.
- Wiese, H., Komes, J., Tüttenberg, S., Leidinger, J., Schweinberger, S.R., 2017. Age-Related Differences in Face Recognition: Neural Correlates of Repetition and Semantic Priming in Young and Older Adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, 1254-1273.
- Wiese, H., Schweinberger, S.R., Hansen, K., 2008. The age of the beholder: ERP evidence of an own-age bias in face memory. *Neuropsychologia* 46, 2973-2985.

ACCEPTED MANUSCRIPT

Wolff, N., Wiese, H., Schweinberger, S.R., 2012. Face recognition memory across the adult life span: event-related potential evidence from the own-age bias. *Psychol Aging* 27, 1066-1081.

Yovel, G., Kanwisher, N., 2005. The neural basis of the behavioral face-inversion effect. *Current Biology* 15, 2256-2262.

Highlights:

- Novel face-recognition training for older adults, using photorealistic caricatures
- We assessed a large number of diagnostic tests and recorded EEG
- We observed performance improvements *during* training, but little generalization
- Extensive training (12 sessions) enhanced face-sensitive N170, P200 and N250
- Long-lasting ERP changes show plasticity of face processing in older age

Accepted manuscript