Differences between high and low performers in face recognition in electrophysiological correlates of face familiarity and distance-to-norm

Anna Schroeger[1,2,3*], Linda Ficco[1,4*], Stella J. Wuttke[1,5], Jürgen M. Kaufmann[1], Stefan R. Schweinberger[1,4]

[1] Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University, Jena, Germany

[2] Department for the Psychology of Human Movement and Sport, Friedrich Schiller University, Jena, Germany

[3] Department of Experimental Psychology, Justus Liebig University Giessen, Germany

[4] Department of Linguistics and Cultural Evolution, Max-Planck Institute for the Science of Human History, Jena, Germany

[5] Infinite Potential Institute, Santa Barbara, CA, United States of America

* These two authors contributed equally to the manuscript and therefore share first authorship.

Correspondence should be addressed to:

anna.schroeger@psychol.uni-giessen.de

and

linda.ficco@uni-jena.de.

Abstract

Valentine's influential norm-based multidimensional face-space model (nMDFS) predicts that perceived distinctiveness increases with distance to the norm. Occipito-temporal event-related potentials (ERPs) have been recently shown to respond selectively to variations in distance-to-norm (P200) or familiarity (N250, late negativity), respectively (Wuttke & Schweinberger, 2019). Despite growing evidence on interindividual differences in face perception skills at the behavioral level, little research has focused on their electrophysiological correlates. To reveal potential interindividual differences in face spaces, we contrasted high and low performers in face recognition in regards to distance-to-norm (P200) and familiarity (N250). We replicated both the P200 distance-to-norm and the N250 familiarity effect. Importantly, we observed: i) reduced responses in low compared to high performers of face recognition, especially in terms of smaller distance-to-norm effects in the P200, possibly indicating less 'expanded' face spaces in low compared to high performers; ii) increased N250 responses to familiar original faces in high performers, suggesting more robust face identity representations. In summary, these findings suggest the contribution of both early norm-based face coding and robust face representations to individual face recognition skills, and indicate that ERPs can offer a promising route to understand individual differences in face perception and their neurocognitive correlates.

Introduction

How are Known Faces Mentally Encoded? Valentine's Norm-based Multidimensional Face-space Model

Similar to other animal species (Hotta et al., 2017; Kendrick et al., 1996; Sheehan & Tibbetts, 2011; Tibbetts et al., 2019), humans heavily rely on faces for identifying other individuals. Familiar faces can be recognized within a few hundred milliseconds, among thousands of known faces (Jenkins et al., 2018). Successful recognition occurs, on one hand, when different instances of a face are related to a certain identity, across changes in external conditions such as luminance, viewing angle, context, or changes in the face itself due to emotions, make-up, hairstyle or momentary health (Burton et al., 2011; Jenkins et al., 2011). On the other hand, faces belonging to different identities need to be correctly told apart. Both processes (telling faces together and apart) depend on the acquisition of an image-independent, stable mental representation of that particular face identity (Burton et al., 2011). After more than six decades of research, prominent theories of how face recognition works have been formulated (Bruce & Young, 1986; Haxby et al., 2000; Young et al., 2020), and researchers have identified reliable neural correlates of this process (Fox et al., 2009; Freiwald, 2020; Müller et al., 2018; Schweinberger & Neumann, 2016).

According to representational models of face perception, known faces are stored in a mental "face space" (Valentine, 1991; Valentine et al., 2016), based on their similarity. In norm-based multidimensional face space (nMDFS), a face is encoded as a point in an n-dimensional space in which the individual dimensions are defined by orthogonal statistical variation and thus, typically, do *not* necessarily correspond to labels/features or descriptions such as eye distance or mouth size. Since each dimension is assumed to be normally distributed, and the mean values of all dimensions lie in the area of the center of the space, the center represents an average or *norm* face. The distance between each exemplar face and the norm (distance-to-norm, DTN) is defined by the vector from that norm to that exemplar face, with faces closer or more distant to the center being perceived as more typical or distinctive, respectively. For completeness, note that an exemplar-based version (eMDFS), in which a norm or central tendency is irrelevant was also discussed by Valentine (1991). Although eMDFS and nMDFS models can make similar predictions and the debate is not fully resolved (Ross et al., 2014), norm-based coding of facial identity has been advocated by studies using contrastive adaptation of face identity (Leopold et al., 2001, 2005;

Little, 2012; Rhodes & Jeffery, 2006) and computational neuroimaging (Carlin & Kriegeskorte, 2017). Across versions of MDFS, the space will be densely clustered by similar faces, and sparsely populated by dissimilar ones. Recognition is more difficult in contrast to similar distractor faces which entail a higher risk of confusion, which explains why more distinctive faces (which by definition have fewer similar competitors) are recognized more easily (Dawel et al., 2019; Rhodes et al., 1997; Schulz et al., 2012a).

The effects of typicality can be studied by morphing individual faces with an average face (which represents a morph combining a large number of real faces). By varying the contribution of the individual and the average face when morphing (e.g., 0.33 individual; 0.66 average), one can manipulate the relative DTN level of the morphed face (e.g., 0.33 DTN relative to the individual face, see **Figure 1**). The characteristics on which a face differs from the average can even be extrapolated beyond that face, thus resulting in caricatures of that face. In contrast, extrapolations in the opposite direction beyond the average face on the trajectory result in so-called "anti-faces", which differ from the average face in opposite ways relative to the individual face. Studies with such manipulations have uncovered differences in behavioral responses and neural processing dependent on DTN (Ewing et al., 2013; Itz et al., 2014; Jeffery et al., 2013, 2018; Leopold et al., 2001, 2005; Limbach et al., 2022; Little, 2012; Rhodes et al., 2011, 2014; Ross et al., 2014; Schulz et al. 2012b; Wuttke & Schweinberger, 2019).

Research on event-related potentials (ERPs) elicited by faces has identified a number of occipito-temporal components which can be considered as markers of face-processing stages (Schweinberger & Neumann, 2016). In particular, the N170 may represent a marker of face detection (Schweinberger et al., 2002; Itier & Taylor, 2002; 2004). The subsequent P200 component appears to be more specifically related to DTN variations: Larger P200 responses were reported for more vs. less typical faces (Halit et al., 2000), and the P200 also was smaller both for digital caricatures and naturally distinctive faces, compared to their more typical counterparts (Kaufmann & Schweinberger, 2012; Schulz et al., 2012a; Schulz et al., 2012b). In line with these findings, the P200 amplitude has been interpreted as a neural correlate of the perceived "prototypicality of a face's individual second-order spatial configuration" (Schweinberger & Neumann, 2016, p. 146) independent of its individual familiarity. Specific predictions from this hypothesis were confirmed in experiments that investigated adaptation to spatially distorted faces (Kloth et al., 2017) or DTN variations (Wuttke & Schweinberger, 2019). For time ranges following the P200, the occipitotemporal N250 and a sustained later temporal negativity (LN) have both been identified as neural correlates of the activation of identity representations in response to familiar faces: The N250 has been related to familiar face recognition (Eimer et al., 2012; Kaufmann et al., 2009; Tanaka et al., 2006; Zheng et al., 2012), and this component has also been related to the development of stable face representations via learning new faces from natural variability (Andrews et al., 2017). More recent research has additionally drawn attention to the fact that a sustained late negativity (LN), typically measured at latencies from approximately 400 ms, can be related to the activation of visual and identity-specific representations when recognizing familiar or learned faces (Popova & Wiese, 2022; Wiese et al., 2019a, 2019b, 2022a, 2022b; Wuttke & Schweinberger, 2019). However, it remains largely unclear whether the P200, N250 or LN ERP components might serve as neural markers that differ reliably between people with different levels of face recognition skills.

Individual Differences in Face Recognition and Differences in the MDFS

Individuals strongly vary in their ability to recognize faces. Face recognition skills are assumed to be normally distributed in the population, ranging from complete face blindness (Duchaine & Nakayama, 2006; Kennerknecht et al., 2006) to extraordinary face recognition skills (Ramon et al., 2019; Russell et al., 2009). There is now consensus that differences in face recognition skills are related at least to some extent to genetic factors, and are relatively resistant to change or training (DeGutis et al., 2014; Dolzycka et al., 2014). Assessment of face recognition skills often is conducted with a whole battery of tests targeting different sub-processes (e.g., simultaneous identity matching, recognition memory), with only a moderate degree of consistency across different tests (Fysh et al., 2020; Noyes et al., 2018).

Identifying factors determining why individuals perform consistently well or poorly in these tests has been challenging. Some studies suggest that differences in low-level visual processes could impact on face recognition skills. Candidates include altered sensitivity to spatial frequencies (Nador et al., 2021) or frequency-dependent contrast sensitivity in low-performing face recognizers (Dougherty et al., 2009; Fisher et al., 2016; Newcombe, 1989) and old individuals (Boutet et al., 2015; Norton et al., 2009; Owsley & Sloane, 1987), or different use of texture and shape information (Itz et al., 2017; Kaufmann et al., 2013) which might characterize different performance profiles. Other studies report positive relationships between face recognition performance and traits such as extraversion (Lander & Poyarekar, 2015; Li et al., 2010) as well as social network size (Wang et al., 2021), and negative relationships with autistic and social anxiety traits (Davis et al., 2011; Halliday et al., 2014; Klin et al., 1999; Weigelt et al., 2012). A small but significant gender difference, with women performing better than males, also tends to be consistently seen in studies with large samples, but this may be modulated by socio-cultural factors (Mishra et al., 2019; Sunday et al., 2019). Of note, both general intelligence (Gignac et al., 2016; Hildebrandt et al., 2011; Wilhelm et al., 2010) and recognition skills for objects (Dennett et al., 2012b; Wilhelm et al., 2010; Wilmer et al., 2010, 2012), seem to be unrelated to face recognition.

One under-investigated hypothesis is that differences in norm-based encoding in individual nMDFS could be associated with individual differences in face recognition performance. In *Superportraits: Caricatures and Recognition*, G. Rhodes hypothesized that “perhaps our memory representations are themselves caricatured” (Rhodes, 1997, p.115; but see also Kaufmann & Schweinberger, 2008), which could promote the idea that good performers in face recognition could encode faces at higher DTN, as compared to low performers.

Studies investigating face identity and shape aftereffects provide some support for this idea. Engfors et al. (2017) measured shifts in perception in a recognition task where target identities were preceded by adaptors that were either on the same or a different identity trajectory through an average face (norm). Face recognition skills correlated positively with the size of identity aftereffects, a finding which is in line with earlier works (Dennett et al., 2012a; Rhodes et al., 2014). Across the three studies, the correlation was specific to faces and did not generalize to other object categories, suggesting that skill-related differences in nMDFS affect the recognition of faces only. Despite these interesting findings, these studies leave some crucial gaps open. First, none of these utilized a battery of face recognition tests in order to quantify individual differences in face recognition reliably. Second, all of the presented correlations were calculated on the scores from the short version of the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), which cannot well capture the whole spectrum of face recognition skills due to ceiling effects (Russell et al., 2009). Third, these studies offer no insight about how differences in the nMDFS relate to those in face recognition at the neural level.

The Present Study

To understand the extent to which individual differences in the nMDFS and in face recognition skills are related at the behavioral and at the neural level, we adopted the experimental paradigm from Wuttke and Schweinberger (2019). Briefly, participants were instructed to rate the typicality of familiar and unfamiliar faces, while EEG was recorded. Face stimuli could be either famous persons (original familiar faces) or unknown people (original unfamiliar faces), and their respective anti-faces (same DTN, but deviating in opposite direction from the norm than original faces on the same trajectory, and therefore all appearing unfamiliar to participants, see **Figure 1**). DTN was manipulated for all faces by morphing faces with a norm face (average of all presented faces of the same gender). Differently from Wuttke and Schweinberger (2019), we were interested in testing individual differences in face recognition, and thus recruited a larger pre-screened sample containing two extreme performance groups of high and low performers in face recognition.

Note that we expected to replicate the main pattern of findings by Wuttke and Schweinberger (2019) for the entire sample: In particular, we expected that the P200 amplitude would decrease with increasing DTN, across all face types. We also expected that for unfamiliar faces, the original and anti-faces would evoke similar ERPs at each level of DTN, and that the N250 and LN would be larger for (recognizable) familiar original faces, as compared to all the other faces presented.

Our overall hypothesis was that carefully pre-selected high and low performers in face recognition would demonstrate systematic differences in these face-elicited ERP components. Our main theory-driven hypothesis is that, to the extent that high performers exhibit more efficient coding of faces in nMDFS, we would find larger differences in the P200 between DTN levels in high as compared to low performers. Such a finding would substantially extend and refine earlier findings regarding differences between high and low performers in famous face recognition (Kaufmann et al., 2013). Additionally, to the extent that high performers exhibit more efficient activation of stored representations of familiar faces, we predicted to find larger N250 (Herzmann et al., 2010; Kaltwasser et al., 2014; Sommer et al., 2021) and LN responses to familiar original faces (but not to the other face types) in high than low performers. We also explored group differences in other components (P100, N170, LPC), without any strong hypothesis on the direction and strength of such differences.

Methods

General Procedure

The study included two phases: a screening phase and a testing phase. Both these took place in the lab, on two distinct occasions, mostly days or weeks apart from each other. During the screening phase participants performed several tests assessing face learning, memory, matching and recognition, tests of visual acuity and contrast sensitivity, one object recognition test and a questionnaire assessing autistic traits. Based on the screening results, two performance subgroups were selected for the subsequent testing phase. In the testing phase participants performed the same experimental paradigm as in Wuttke and Schweinberger (2019), while undergoing EEG recordings. Neither the study procedures nor the analyses were pre-registered.

Screening Phase

Screening Phase: Participants

During the screening phase 129 participants (46 males), with an average age of 23.9 years old (SD = 3.6) performed a battery of tests. Part of this group was then selected for the testing phase depending on their performance. All participants had normal or corrected to normal vision, provided their informed consent, and received either 20 euros or University credits for their participation. The Ethical Commission of the Faculty of Social and Behavioral Sciences at the Friedrich-Schiller-University of Jena approved the study (reference: FSV 15/15).

Screening Phase: Assessment tools

In total nine screening tests were administered. In order to select participants eligible for the testing phase, we used two main tools. These were: i) the long version of the Cambridge Face Memory Test (CFMT+ Duchaine & Nakayama, 2006; Russell et al., 2009) as an indicator of unfamiliar face learning skills, and ii) a short questionnaire investigating participants' familiarity with the identities that were used in the main experiment. To qualify for the present study, participants' CFMT+ score had to be above 77 or below 56 (respectively one standard deviation above and below the mean performance, see Duchaine & Nakayama, 2006 and Russell et al., 2009); in addition, participants needed to be familiar with the famous identities used in the EEG experiment to achieve a valid dissociation between familiar and unfamiliar stimuli in both groups. Further tests in our battery were the following: the Glasgow Face Matching Test (Burton et al., 2010), an adapted version of the Bielefelder Famous Faces Test (Dobel et al., 2008; Fast et al., 2008), the Cambridge Car Memory Test (Dennett et al., 2012b), two subtests from the Freiburg

Vision Test (Bach, 1996, 2006), and finally an autism questionnaire (Baron-Cohen et al., 2001). These screening tests are described in detail in the Supplementary Materials.

Screening Phase: Procedure

The order of test administration was constant across participants. The duration of whole screening phase was approximately 2.5 hours.

Testing Phase

Testing Phase: Participants

Sample size for our effect of interest was defined based on an a priori power analysis (see Supplementary Materials). Out of the 129 participants who took part in the screening phase, 43 participants (all right-handed) qualified for further participation based on their results in the long form of the Cambridge Face Memory Test (CFMT+) (Duchaine & Nakayama, 2006; Russell et al., 2009). The selected sample consisted of 21 high performers and 22 low performers (See Table 1). Cutoff criteria were taken from accumulated former research of the department with comparable samples (N = 76, M_{CFMT+} = 66.46, SD_{CFMT+} = 10.41) and were similar to earlier published results (see **Figure S1**, Supplementary Materials) (Bobak et al., 2016; Russell et al., 2009; Russell et al., 2012). A participant was assigned as 'high performer' if a score of 77 or higher was reached, that is, at least one standard deviation above the mean in the CFMT+; conversely, low performers obtained a score of 56 or smaller, that is, at least one standard deviation below the mean of the CFMT+ (Duchaine & Nakayama, 2006; Russell et al., 2009).

Table 1. *Demographic data of high and low performers in the CFMT+.*

	N	Males	Mean age (SD)	Mean CFMT+ score (SD)
High CFMT+ performers	21	7	24.71 (3.56)	84.65 (4.58)
Low CFMT+ performers	22	7	23.04 (3.40)	51.24 (4.27)
Whole screened sample	129	14	23.86 (3.54)	66.99 (12.49)

Testing Phase: EEG experiment

Both the stimulus materials and the experimental procedures used in the testing phase are the same as described in Wuttke & Schweinberger (2019). Therefore, we refer the interested reader to this work, and we only report a summary here.

Experimental stimuli consisted in a total of 120 face images, of which 30 showed faces of 10 famous people (e.g., very well-known celebrities), 30 showed faces of 10 unfamiliar people (e.g. relatively unknown celebrities), and the remaining 60 showed their respective anti-faces. A constant number of 20 faces was used both to create the *average face* (an average of 20 images each showing a different identity) and to create each *original face* (an average of 20 images each showing the same identity), to avoid morphing artifacts that are prevalent when high-resolution individual images are morphed with smoothed average faces. Anti-faces were generated by morphing an original face image with the respective average face (norm) along the same trajectory in the MDFS, but in the opposite direction. Each of these identities was displayed in three possible levels of typicality (or distance-to-norm, DTN), achieved by morphing with the norm: 0.33, 0.66, and 0.99 for faces, and −0.33, −0.66, and −0.99 for anti-faces. For a graphical illustration of the relationship between original and anti-faces, as well as the effects of DTN manipulations, see **Figure 1** (see also Faerber et al., 2015, Fig.1; and Wuttke & Schweinberger, 2019, Fig. 1).

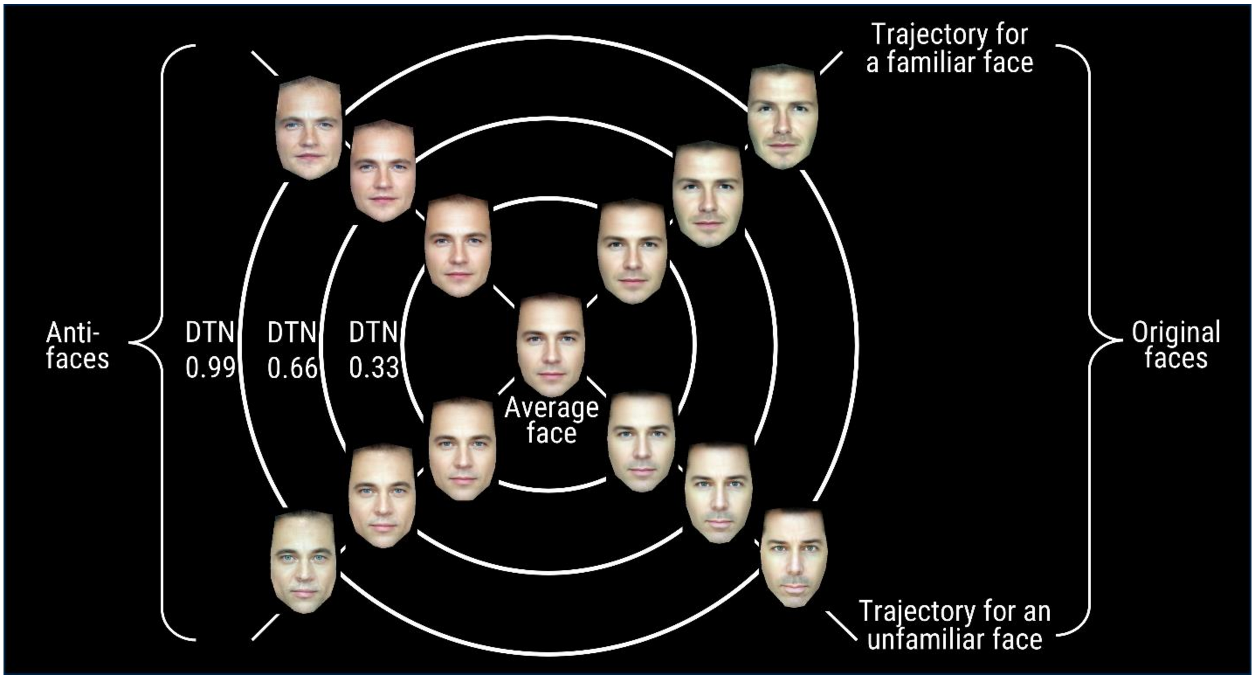

Figure 1. *Overview of the stimuli and paradigm. The trajectory of one exemplary familiar (David Beckham) and one exemplary unfamiliar face (Richard Lounello) are depicted. Each trajectory goes from the original face (right side) through the average face (center) to the respective anti-face. Three levels of distance to norm were used (0.33, 0.66, and 0.99).*

During the EEG experiment, participants rated the typicality of the presented faces using a 6-point-Likert scale (1 = very atypical, 6 = very typical) on a standard computer keyboard. They were instructed to indicate how much the physical appearance of the face differed from all the faces familiar to them, and to perform these ratings fast and spontaneously. Stimuli were presented on a ViewSonic Graphics Series G90fB screen with 19" and 60 Hz refresh rate. First, participants performed 16 practice trials using other face stimuli (subsequently discarded from the analyses). Then, they performed 480 trials (2 familiarity conditions [familiar, unfamiliar] x 2 polarities [original face, anti-face] x 3 DTN levels [0.33, 0.66, 0.99] x 4 repetitions x 10 identities), resulting in 40 trials per condition. Target face stimuli were presented for 1000 ms each, in a randomized order. A fixation cross preceded each stimulus (500 ms), and a blank screen followed (1500 ms). Participants were prompted to respond faster in case their response time was longer than 2000 ms. The experiment was divided into blocks of 40 trials, between which the participants could take self-paced breaks. In total, the EEG experiment lasted about one hour.

Testing Phase: Statistical Analyses of Behavioral and EEG data

We performed t-tests to assess significant differences between high- and low-performers in the screening tests. In case the Shapiro-Wilk test and the Levene-test detected a violation of the assumptions of normality and homoscedasticity respectively, then we performed nonparametric Wilcoxon-Man-Whitney U tests instead. We ran these analyses in R (Version 4.0.4, R Core Team, 2020).

As for the testing phase data, we collected typicality ratings, reaction times (RTs) and electrophysiological signal during rating task performance. As ratings were our main interest, RTs were not analyzed further.

Electrophysiological data were recorded using a 32-channel EEG with BioSemi Active II system (BioSemi, Amsterdam, The Netherlands). Active sintered Ag/Ag-Cl electrodes were attached with the help of an elastic cap in the 10-20 system. Signals were recorded with a 512 Hz sample rate from DC to 155 Hz from the following sites: 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, I2. A “zeroRef” was used and ground and reference electrodes were replaced at CMS/DRL circuit (see also www.biosemi.com/faq/cms&drl.htm).

The multiple source eye correction algorithm implemented in BESA 5.1.8.10 (Berg & Scherg, 1994; MEGIS Software GmbH, Graefelfing, Germany) was used to correct the effects of ocular artefacts and blinks. To reduce eye movements, participants had to keep their fingertips on the response keys while performing the task. Trials with non-ocular artefacts were rejected manually and by using the BESA™ artefact rejection tool (amplitude threshold 75 µV, gradient criterion 50 µV). We segmented EEG recordings from −200 ms until 1000 ms relative to stimulus onset, with the first 200 ms as baseline. Trials were averaged separately for each channel and experimental condition. Averaged ERPs were low-pass filtered at 30 Hz (zero phase shift, 12 dB/octave), and recalculated to average reference. After artefact rejection, a mean of 37.8 out of 40 trials for familiar OF, 37.9/40 for “familiar” AF, 37.9/40 for unfamiliar OF, and 37.8/40 for unfamiliar AF contributed to an average ERP.

ERPs were quantified using mean amplitudes for the occipito-temporal P200 components (190–260 ms) and the ventral-temporal N250 (260–400 ms) and late negativity (LN, 400–800 ms). Mean amplitudes were also taken for the N170 (140–180 ms) and the late positive component (LPC, 500–700 ms), all relative to a 200 ms pre-stimulus baseline. The time windows for N170, P200, N250, and LPC were selected in accordance with a previous publication using the same paradigm (Wuttke & Schweinberger, 2019), and were confirmed by visual inspection of the grand mean ERPs. To select electrodes of interest, we used amplitude maxima of a particular component in the present grand means and previous research as criteria (Bindemann et al., 2008; Schweinberger, Pickering, Burton, & Kaufmann, 2002; Schweinberger, Huddy, & Burton, 2004). Accordingly, N170 was assessed at P7, P8, P200 at P7, P8, P9, and P10, N250 and LN was captured at P7, P8, P9, P10, PO9. PO10, TP9, and TP10, and the LPC was measured at Cz and Pz.

We performed repeated measures ANOVAs with the following within-factors: topography (e.g., site and hemisphere, depending on the ERP component), a combined familiarity-polarity factor (familiar original faces, familiar anti-faces, unfamiliar original faces, unfamiliar anti-faces), and DTN (0.33, 0.66, and 0.99) and the between-factor group (high vs. low performer). Note that participants subjectively perceived only original familiar faces as actually familiar, whereas the other face types looked unfamiliar to them. Analyses were performed in SPSS and R version 4.1.2 (R Core Studio, 2020). Here we mainly report ERP reports to significant effects that involved group differences in the experimental factors of familiarity-polarity, and DTN, and do not report topographical effects per se. We refer the interested reader to the Supplementary Information for a complete report.

Testing Phase: Procedure

The order of test administration was constant across participants. The duration of whole testing phase was approximately one hour.

Results

Screening tests

High performers in the CFMT+ were also superior to low performers in the other face processing tests, including famous face recognition (as measured by the BFFT, W = 78.5, p < .0001) and face matching (GFMT, W = 44.5, *p* <.0001). While a certain consistency between face processing measures is not surprising, two of our screening results were unexpected. First, high performers also performed marginally but not significantly better in object recognition (tested in the CCMT, t(40) = -1.993, *p* = .053). Second, the two groups did not differ in terms of autistic traits (AQ questionnaire, t(38) = -0.047, *p* = .962). Furthermore, while we found no differences in visual acuity (W = 178.5, *p* = 0.4804), contrast sensitivity was slightly but significantly higher in CFMT+ high performers (t(34) = -2.736, *p* = .009; cf. Fig. 2).[1] See Table 2 for a complete report

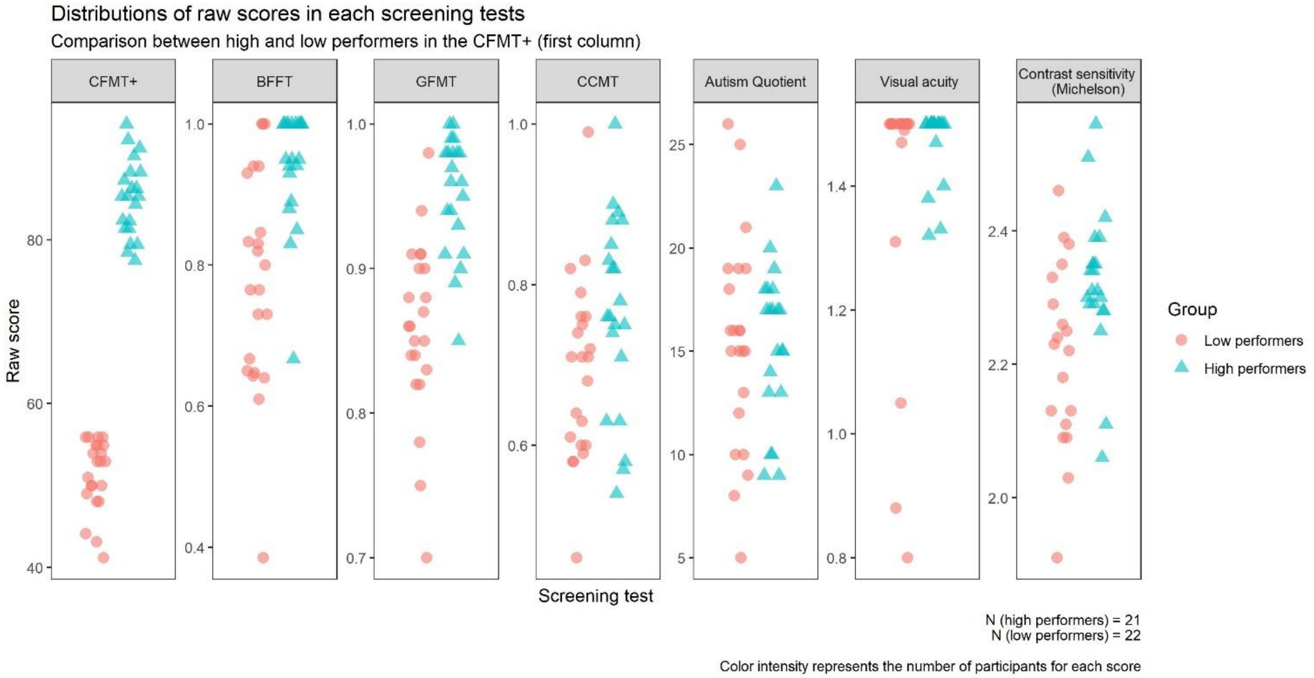

Figure 2: *Performance of individual participants in each test: unfamiliar face learning (CFMT+), familiar face recognition (BFFT), face matching (GFMT), object recognition (CCMT), autism quotient and vision tests (decimal visual acuity and contrast sensitivity). The Y axis represents the raw performance score in the original units, whereas the X axis represents the tests used. Both colour and shape code participant group (high and low performers in the CFMT+, see first facet on the left). Higher values reflect better performance, except the autism quotient where higher values indicate scoring higher on autism spectrum. For reference to colour in this figure, please refer to the online version of this article.*

---

[1] We conducted extensive further analyses which, in short, ruled out the possibility that the group differences in face perception reported below would be entirely explained by contrast sensitivity. In the interest of readability, these additional analyses are reported in the Supplementary Information.

of the screening results, and **Figure 2** for a visualization of the performance of individual participants in each test.

Table 2. *Results of the screening tests: long form of the Cambridge Face Memory Test (CFMT+), Bielefelder Famous Faces Test (BFFT), Glasgow Face Matching Test (GFMT), Autism Questionnaire (AQ), Vision tests. Unpaired t-tests were performed if there was no violation of the assumptions. If else, Wilcoxon-Mann-Whitney U test was conducted and median and range are reported instead of the mean and standard deviation. Table structure was adopted from Crawford, Garthwaite, and Porter (2010).*

Task	Test assumptions of normal distribution and variance homogeneity						Unpaired t-tests					
	Levene test		Shapiro-Wilk test High performers		Shapiro-Wilk test Low performers				High performers		Low performers	
	F(df)	p	W	p	W	p	t (df)[a]	p	N	Mean[b] (SD)	N	Mean (SD)
CFMT+	0.12 (1, 41)	.731	.97	.820	.89	.015	**0[c]**	**<.001**	21	85.06[d] (4.65)	22	51.34 (4.35)
BFFT	8.08 (1, 41)	<.001	.73	<.001	.94	0.231	**78.5[c]**	**<.001**	21	0.94[d] (0.08)	22	0.78 (0.16)
GFMT	1.64 (1, 41)	.207	.91	.049	.96	.550	**44.5[c]**	**<.001**	21	0.95[d] (0.04)	22	0.86 (0.06)
CCMT	0.14 (1, 41)	.709	.96	.538	.96	.516	-2.00 (41)	.053	21	0.77 (0.12)	22	0.69 (0.11)
AQ	0.93 (1, 41)	.341	.94	.257	.97	.737	-0.05 (41)	.963	21	15.43 (3.74)	22	15.36 (5.18)
Acuity decVA	2.29 (1, 38)	.138	.55	<.001	.54	<.001	178.5[c]	.480	21	1.47[d] (0.06)	19	1.39 (0.22)
Vision Contrast Michelson	1.86 (1, 38)	.180	.92	.073	.98	.978	**-2.77 (38)**	**.009**	21	2.32 (0.11)	19	2.21 (0.14)

[a] If the assumptions were violated, Wilcoxon-Mann-Whitney U test is reported.

[b] If the assumptions were violated, median and range are reported instead of mean and sd.

[c] Wilcoxon-Mann-Whitney U

[d] Median and range

Behavioral data

In the main Experiment while measuring EEG, participants performed a typicality rating task on familiar and unfamiliar celebrities. Behavioral typicality ratings are summarized in Figure 3. We performed an ANOVA on typicality ratings, including DTN (0.33, 0.66, 0.99) as a within factor, and group as a between factor (CFMT+ low performers vs. CFMT+ high performers). We also included a combined factor familiarity-polarity (which had 4 levels that reflected orthogonal combinations of familiarity: familiar, unfamiliar and polarity: original face, anti-face). It is notable that perceived typicality did not differ overall between high and low performers, $F(1, 41) = 2.69$,

$\varepsilon_{G\text{-}G} = .532$, $p = .108$, $\eta_p^2 = .062$. In line with the manipulation intention, faces of larger DTN were rated to be less typical, as evident in a significant main effect of DTN, $F(1.06, 43.31) = 111.05$, $\varepsilon_{G\text{-}G} = .528$, $p < .001$, $\eta_p^2 = .730$ (all pairwise comparison's $p < .001$). Additionally, in line with the manipulation intention, a main effect of the combined familiarity-polarity factor indicated that familiar original faces were rated to be more typical compared to all other combinations, $F(1.39, 57.02) = 14.41$, $\varepsilon_{G\text{-}G} = .464$, $p < .001$, $\eta_p^2 = .260$ (contrast testing familiar original faces against the other conditions: $t(41) = 3.63$, $p < .001$). Descriptively this difference seems to be even larger in the group of high performers, but the interaction of DTN and familiarity-polarity with group was not significant, $F(1.39, 57.02) = 1.54$, $\varepsilon_{G\text{-}G} = .464$, $p = .224$, $\eta_p^2 = .036$. Contrasts additionally indicated that familiar anti-faces systematically differed from both unfamiliar face conditions, as they were rated to be less typical than unfamiliar original faces and unfamiliar anti-faces ($p < .001$). Strikingly, there was an interaction between DTN and the factor familiarity-polarity, $F(2.17, 88.82) = 9.89$, $\varepsilon_{G\text{-}G} = .361$, $p < .001$, $\eta_p^2 = .194$, with the difference between familiar OFs and all other faces increasing from DTN level 0.33 to 0.99, as depicted in **Figure 3** (larger difference at DTN level 0.66 compared to 0.33, $p = .023$, and larger difference at DTN level 0.99 compared to 0.66, $p < .001$). This means that, with increasing DTN, faces were perceived to be less typical, but lesser so for familiar original faces.

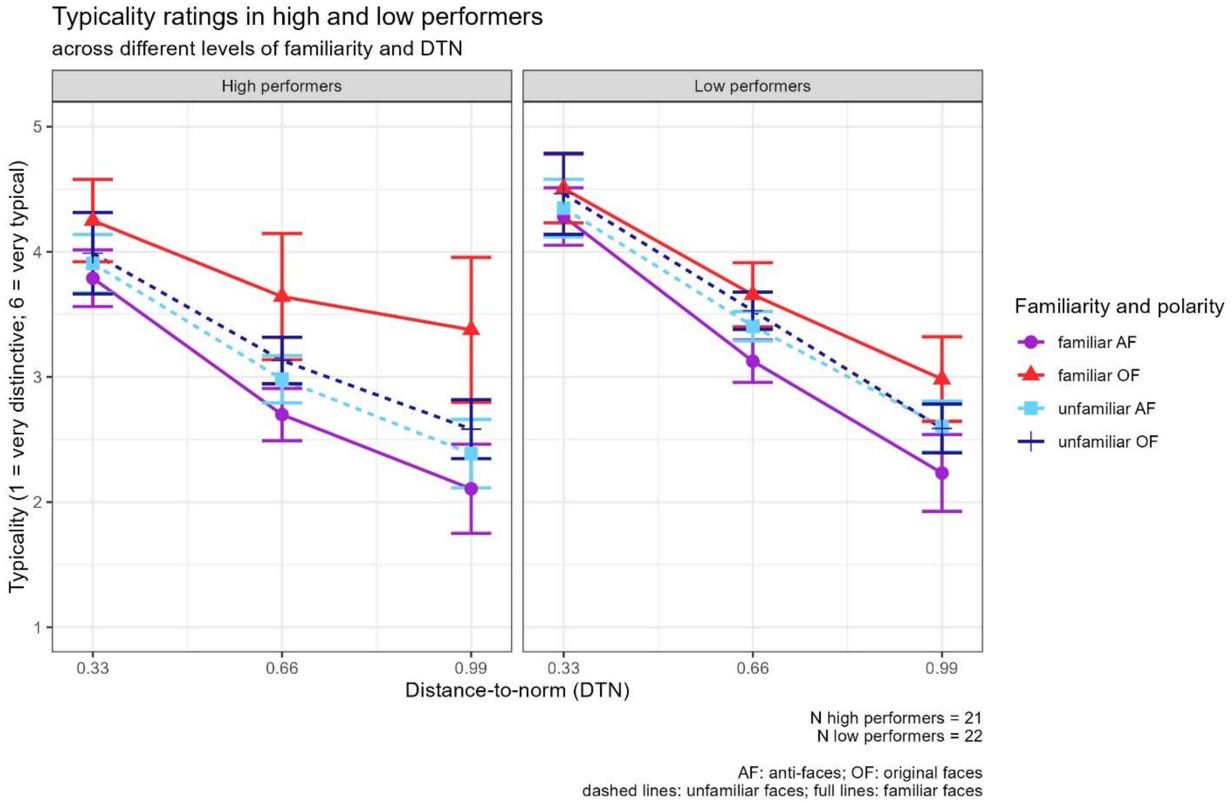

Figure 3. *Typicality ratings (mean and CI) obtained during the EEG-rating task. Error bars indicate 95% within-subject confidence intervals (cf., Baguley, 2012). For reference to colour in this figure, please refer to the online version of this article.*

Electrophysiological data

For all ERP-components mixed design ANOVAs were calculated including a topographical predictor (either site or hemisphere or both, based on the respective ERP component), the within factors "familiarity-polarity” (4 levels: familiar original faces, familiar anti-faces, unfamiliar original faces, unfamiliar anti-faces) and DTN (3 levels: 0.33, 0.66 or 0.99), and the between factor group (high or low performers). Only significant effects or interactions with the manipulated factors (familiarity-polarity or DTN) will be reported, with a focus on group differences. As purely topographical effects of site or hemisphere are of minor importance for our research question, these effects are reported in the Supplementary Materials, together with a complete report of the ANOVA results on all ERP components. We show an overview of our analysis component by component in **Figure 4**, which illustrates the effects detailed in the following sections.

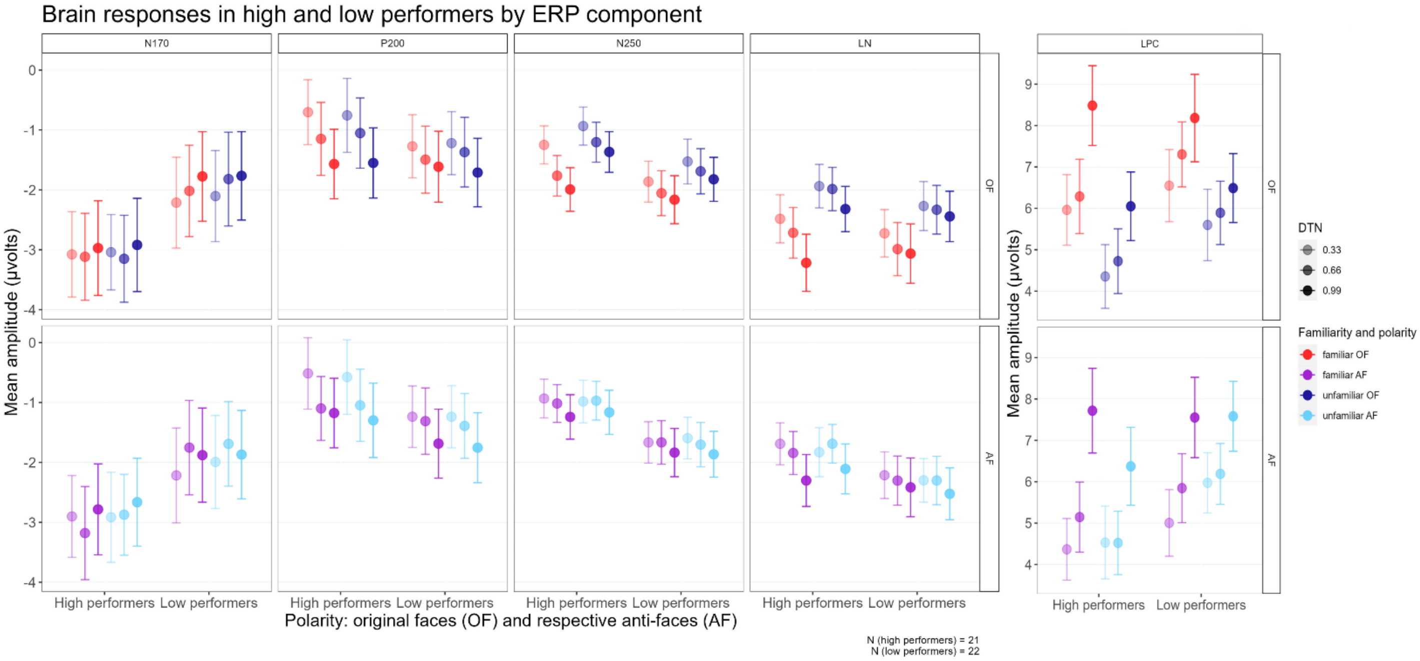

Figure 4. *Overview of all ERP components analyzed, separated by DTN levels (shades), polarity-familiarity levels (color and y axis) and performance group (x axis). The first significant effect of DTN can be found in the N170 in low performers, but DTN effects become more pronounced in the P200, especially for high performers. A significant difference for familiar original faces (red) compared to all other faces is first evident in the N250, and it remains prominent in the LN and the LPC. Error bars reflect 95% within-subject confidence intervals (cf., Baguley, 2012). Note the different scaling used for the late positive component. Please refer to the online version of this article for a colour version of this figure.*

N170 (140-180 ms)

The results of the N170 analysis revealed a significant main effect of DTN, $F(2, 82) = 4.56$, $p = .013$, $\eta_p^2 = .100$, with decreasing negativity (decreasing amplitude) for increasing DTN. Importantly, this effect differed between groups, as indicated by a significant interaction between DTN and group, $F(2, 82) = 3.65$, $p = .030$, $\eta_p^2 = .082$. The main effect of group was not significant ($p = .051$). Sliding difference contrasts revealed a significant decrease of the N170 amplitude from DTN level 0.33 to 0.66 ($p = .004$) for low performers only. For high performers there was a non-significant trend for DTN effect, but only for level 0.66 vs. 0.99 ($p = .056$), all other contrasts were not significant ($p > .123$).

P200 (190-260 ms)

Replicating previous research, we observed a prominent effect of DTN in the P200, $F(2, 82) = 62.75$, $p < .001$, $\eta_p^2 = .605$. In line with the view that P200 is a marker of face typicality, P200 amplitude declined with increasing DTN. Central to our research question, this effect differed substantially between groups, as reflected in the interaction between group and DTN, $F(2, 82) = 5.33$, $p = .007$, $\eta_p^2 = .115$. In both high and low performers all DTN levels differed between each other (*ps < .002*). Critically, the DTN effect was more pronounced in high performers between levels 0.33 and 0.66 (*p = .005*) (Note that groups did not differ in the decrease from level 0.66 to 0.99, *p = .106*). We additionally found a main effect of the combined familiarity-polarity factor, $F(3, 123) = 2.77$, $p = .044$, $\eta_p^2 = .063$. Planned contrasts revealed a non-significant difference between familiar original faces and all the other face types ($p = .053$), as well as a significant difference between original faces and anti-faces, regardless of their familiarity ($p = .021$), with original anti-faces eliciting larger amplitudes than original faces. The familiarity-polarity effect did not differ depending on performance group in this time range.

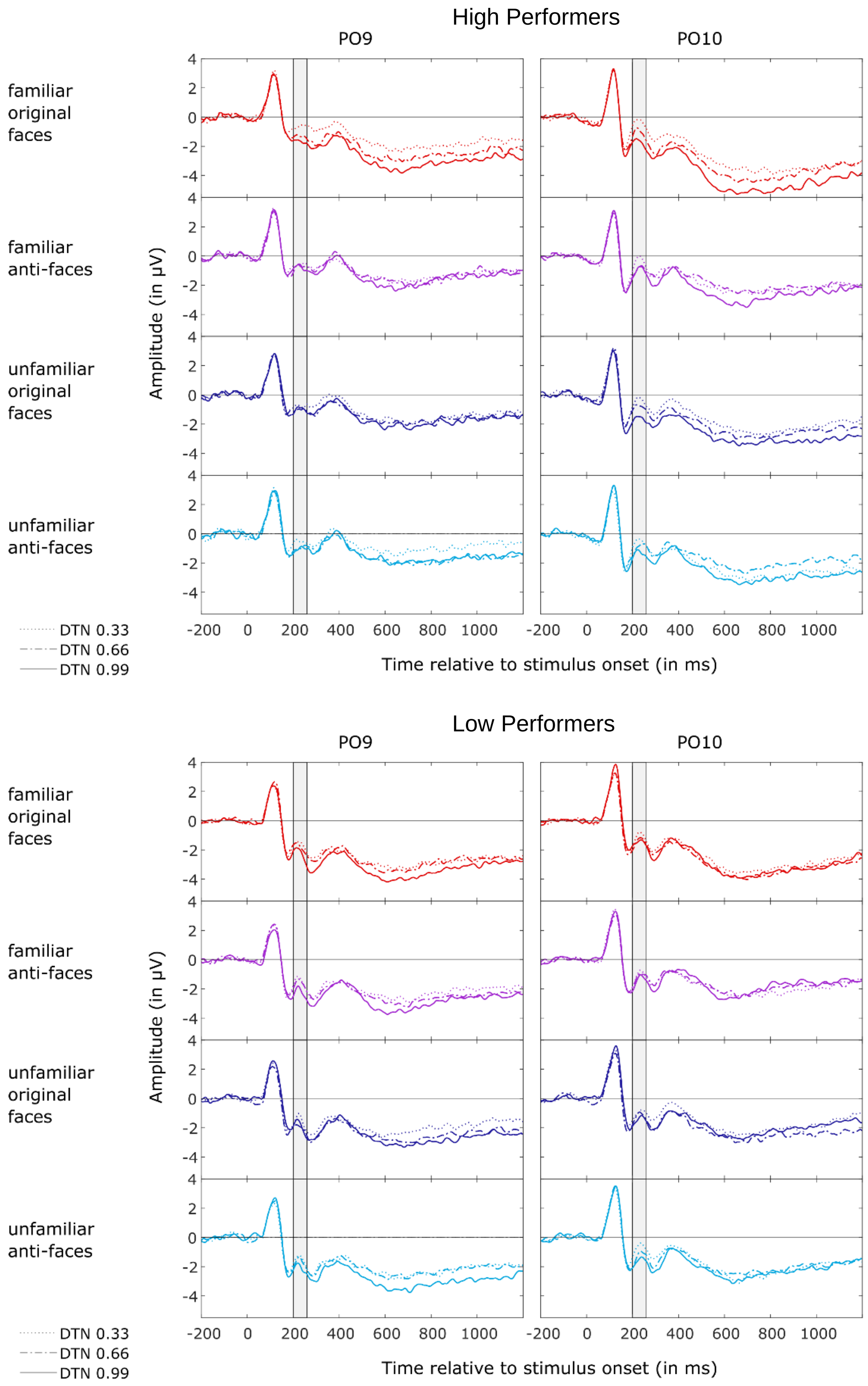

Figure 5. P200 DTN effect at electrodes PO9 and PO10 for high and low performers of face recognition. Note that only familiar original faces (OF) were familiar to the participants, whereas their anti-faces (AF) should not be interpreted as familiar.

N250 (260-400 ms)

The analysis of the N250 revealed a significant main effect of the familiarity-polarity factor, $F(3, 123) = 49.72$, $p <.001$, $\eta_p^2 = .548$. Planned contrasts revealed a difference between familiar original faces and all the other face types ($p < .001$), between familiar and unfamiliar original faces ($p < .001$), and between original faces and anti-faces ($p < .001$).

Strikingly, we observed a significant familiarity-polarity by group interaction, $F(3, 123) = 5.46$, $p = .001$, $\eta_p^2 = .118$. Planned contrasts showed that in both performance groups there was a difference between original familiar faces and all the other face types ($p < .001$ in each group, with original faces eliciting larger amplitudes). Crucially, this difference was even larger in high, as compared to low performers ($p < .003$). In **Figure 6** this effect is evident when comparing the "familiar original" ERP curves in red to all other conditions (also cf. **Figure 7** for a plot of mean amplitudes of the difference between familiar and unfamiliar original faces). We interpret this difference as the effect of being truly familiar with a face (original, familiar faces), that is larger in high CFMT+ performers.

Additionally, the effect of DTN was significant, $F(1.42, 58.21) = 10.87$, $\varepsilon_{G\text{-}G} = .710$, $p < .001$, $\eta_p^2 = .208$. Familiarity-polarity and DTN interacted significantly, $F(6, 246) = 2.34$, $p = .032$, $\eta_p^2 = .054$. Post-hoc contrasts revealed a significant difference between original familiar faces and the other face types at all DTN levels (all ps < .001), but this effect increased significantly from DTN level 0.33 to DTN level 0.66 ($p = .007$; sliding difference contrasts over DTN levels ps < .083).

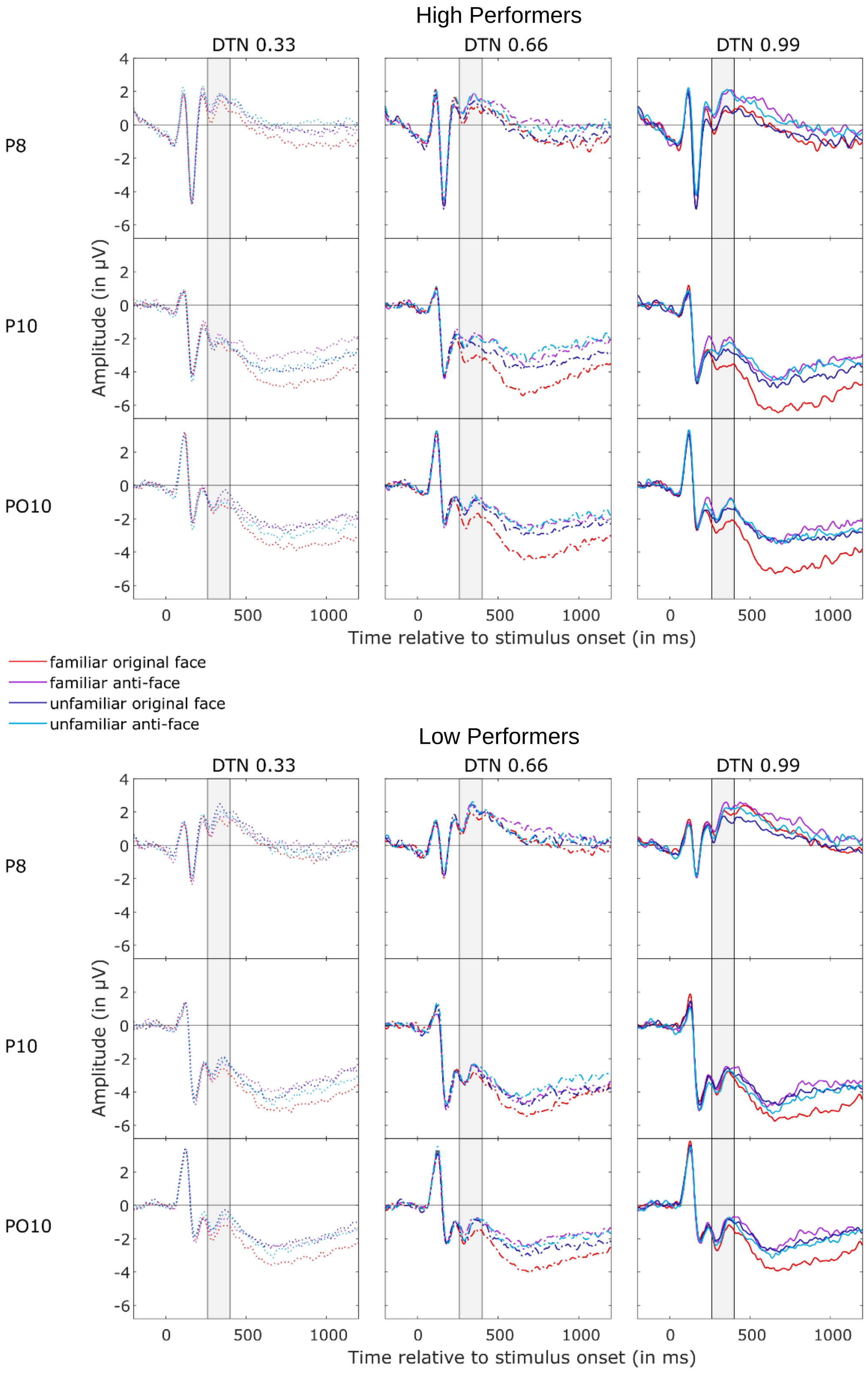

Figure 6. N250 familiarity effect at electrodes P8, P10, and PO10 for high and low performers of face recognition. Note that only familiar original faces (OF) were familiar to the participants, whereas their anti-faces (AF) should not be interpreted as familiar.

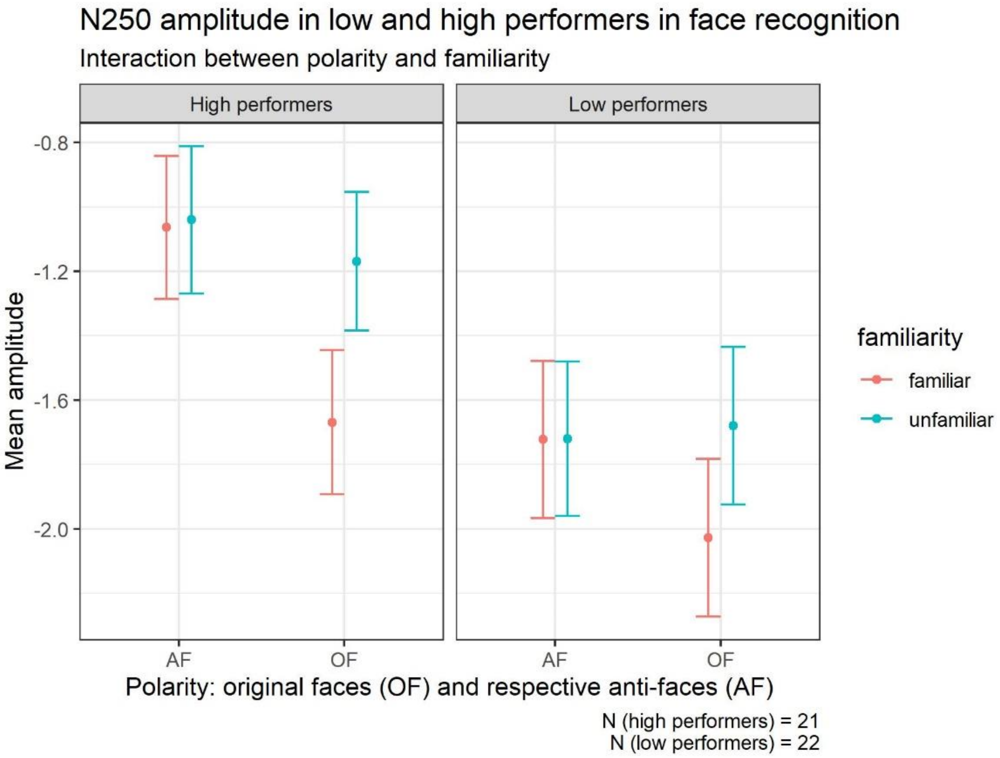

Figure 7. *Planned contrast for the group differences in the familiarity x polarity interaction in the N250. AF (anti-face), OF (original face). Means and within-subjects confidence intervals (cf., Baguley, 2012) are displayed.*

Late negativity (400-800)

The analysis of the late negativity revealed a significant main effect of familiarity-polarity, $F(2.45, 100.28) = 61.03$, $\varepsilon_{G\text{-}G} = .815$, $p < .001$, $\eta_p^2 = .598$, which was very similar to that observed in the N250 range. This effect again differed between groups, $F(2.45, 100.28) = 3.06$, $\varepsilon_{G\text{-}G} = .815$, $p < .041$, $\eta_p^2 = .069$. Post-hoc contrasts showed that the difference between familiar, original faces and the other face types was enhanced in high, as compared to low performers ($p < .001$). Moreover, the effect of DTN was significant, $F(1.56, 64.05) = 11.45$, $\varepsilon_{G\text{-}G} = .781$, $p < .001$, $\eta_p^2 = .218$, but did not interact with group. With increasing DTN the amplitude increased, as shown by sliding difference contrasts (all *p*s < .007).

Late positive component (500-700 ms)

For the late positive component the analysis revealed a significant main effect of familiarity-polarity, $F(2.33, 95.42) = 21.45$, $\varepsilon_{G\text{-}G} = .776$, $p < .001$, $\eta_p^2 = .343$. Again, post-hoc contrasts showed a difference between original familiar faces and the rest, as well as between familiar and unfamiliar faces (both *p*s < .001). This effect did not differ significantly between groups, $F(2.33, 95.42) = 2.66$, $\varepsilon_{G\text{-}G} = .776$, $p = .067$, $\eta_p^2 = .061$. The main effect of DTN was again significant, $F(1.57, 64.23) = 41.33$, $\varepsilon_{G\text{-}G} = .783$, $p < .001$, $\eta_p^2 = .502$. With increasing DTN the amplitude increased, as revealed by sliding difference contrasts (*p*s < .001). Furthermore, there was a significant interaction between familiarity-polarity and DTN, $F(6, 246) = 3.36$, $p = .003$, $\eta_p^2 = .076$. Familiar original faces differed from other face types at each DTN level (*p*s < .001). Whilst familiar original faces elicited larger amplitudes as compared to unfamiliar original faces at each level of DTN (all *p*s < .001), familiar anti-faces elicited smaller amplitudes at DTN level 33 ($p = .034$) but larger amplitudes at level 99 ($p = .038$), as compared to unfamiliar anti-faces.

Discussion

Overview and replication of past research

On the basis of Valentine's nMDFS (Valentine, 1991), we aimed to better understand how faces with varying DTN and familiarity levels are processed in individuals with high and low recognition skills, as assessed by the CFMT+ (long version of the Cambridge face memory test). We were able to i) replicate electrophysiological correlates of distance-to-norm (DTN) and familiarity processing, namely the P200 and N250 across both groups, and ii) showed interindividual differences in both the P200-typicality processing and the N250-familiarity processing.

Behaviourally, we observed performance differences between high and low performers of face recognition in other face processing tasks, such as recognizing famous people and matching unfamiliar faces, but not in tasks involving object processing (although marginally poorer car memory was observed in low performers of face recognition). High performers also displayed slightly higher contrast sensitivity on average – a group difference, which was significant, whereas we found no evidence for differences in visual acuity nor in subclinical autistic traits. As regards typicality perception of the faces in the main experiment, original, familiar faces were rated as more typical than all other faces, replicating a special role of familiar face processing as evident in previous findings (Burton et al., 2011; Burton et al., 2015; Wuttke & Schweinberger, 2019; Young & Burton, 2018). While this familiarity effect increased with increasing DTN for both high and low performers, it appears descriptively stronger for high performers. The perception of anti-faces of familiar faces as slightly more distinctive than the original familiar faces has also been reported before (Wuttke & Schweinberger, 2019). A tentative interpretation of the systematic difference of "familiar" anti-faces is that morphing familiar faces, which are perceived as more typical than unfamiliar faces (DTN and facial attractiveness being equal) would enhance a-typical features (Faerber et al., 2016, also cf. Figure 1 in Faerber et al., 2016). In other words, creating an anti-face of these faces would selectively change those features that determined the increased typicality perception of familiar faces into its diametral opposite. Thus, one can speculate that this "typicality boost" is reversed by the anti-face morphing procedure, thus leading to increased distinctiveness of the respective anti-faces. This again somehow supports the idea that familiar faces are represented in a special way and calls for studies focussed on anti-faces of familiar vs. unfamiliar identities (Burton et al., 2015; Young & Burton, 2018).

At the neural level, we observed several interesting effects. In contrast to by Wuttke & Schweinberger (2019) we observed DTN effects as early as in the N170 time window. This difference could be due to our larger sample size (43 compared to 24 participants), resulting in an increased power to detect this effect (see Kaufmann et al., 2013, who detected a trend in this direction with 28 participants). Of potential interest, whereas this present N170 DTN effect was prominent in low performers, it was not significant in high performers. Crucially, we replicated Wuttke & Schweinberger (2019) as regards the observation of decreased P200 amplitudes with increasing DTN (Halit et al., 2000; Itz et al., 2014; Kaufmann & Schweinberger, 2012; Latinus & Taylor, 2006; Limbach et al., 2018; Schulz et al., 2012a; Schulz et al., 2012b), and also with respect to larger N250 amplitudes for familiar original faces (Gosling & Eimer, 2011; Herzmann et al., 2004; Kaufmann et al., 2009; Pierce et al., 2011; Schweinberger et al., 1995; 2002; Tanaka et al., 2006; Zheng et al., 2012). The sensitivity of the P200 to second-order face configuration and to familiarity has also been long known in ERP literature (reviewed in Schweinberger & Neumann, 2016). Most importantly, we found: i) larger P200 amplitude differences between DTN levels in high performers, as compared to low performers, and ii) larger differences in the N250 and late negativity between familiar, original faces and all the other face types (unfamiliar) in high, as compared to low performers.

General Vision Differences Contribute to Face Recognition Performance

In line with other published works (Baldson et al., 2018; Burton et al., 2010; Fysh & Bindemann, 2018; Itz et al., 2018; McCaffery et al., 2018), high and low performers also differed in their ability to recognize celebrities from their face and the ability to tell together and apart two unfamiliar face identities. This suggests that, despite these tasks likely tap into different mechanisms (Itz et al., 2017), there seems to be at least some degree of overlap across face processing tasks. However, group differences in other aspects of visual perception are less clear-cut. In line with reports that face and object processing are more associated than previously thought (Hendel et al., 2019; White & Burton, 2022), high performers were marginally better at learning and recognizing car exemplars. This is likely due to very similar demands between the CFMT+ and the CCMT (Dennett et al., 2012b; Harrington et al., 2022; Shakeshaft & Plomin, 2015), which feature the same task structure and require similar mental strategies. Together with task demands similarities, however, high and low performers might also differ in a general factor representing efficient visual processing (Faghel-Soubeyrand et al., 2022). This factor could potentially differentiate high and low performers in a non-clinical sample like ours (Hendel et al., 2019). Since contrast sensitivity was significantly higher in high performers, one can speculate that low-level aspects of vision, at least partially drive face recognition performance. While this idea seems to be supported by clinical studies (Hirji et al., 2020), our N250 analyses suggest that differences in neural correlates of face recognition persist even when contrast sensitivity is added as a covariate.

The present high and low performer groups did not differ in visual acuity. Indeed, severe visual acuity deficits, especially since childhood, can result in impaired face recognition (Mauer et al., 2007). However, in our study participants' vision deficiency was likely not comparably strong, and was corrected with glasses or contact lenses during CFMT+ performance. Thus, the lack of differences in visual acuity we report might not be very informative and is not particularly unexpected. Nonetheless, it seems that personality variables such as subclinical autistic traits play a minor role in face recognition performance, in contrast to low-level visual features. As a comment, research positing a face recognition deficit in autism (Klin et al., 1999; Weigelt et al.,

2012) might benefit from a systematic investigation of low-level visual differences between autistic and neurotypical individuals (cf. Simmons et al., 2009, for a review).

Differences in Neural Effects of DTN and Familiarity between High and Low Performers

We had two main hypotheses. One was that high and low performers differ in their norm-based coding, with high performers showing stronger DTN effects. The second was that enhanced familiarity processing in high performers is reflected in larger neural correlates of face familiarity, in comparison to low performers.

In support of the idea that high performers in face recognition encode faces at higher DTN, as compared to low performers, these participants seemed to have a more sensitive perception of typicality, as shown by more diversified typicality ratings for different face types. These patterns, presented in **Figure 3**, suggest an expanded or better tuned face space in high performers, for whom typicality perception is more sensitive. This is mirrored at the neural level in increased DTN effects in the P200 component in high performers: in particular, the decreased P200 amplitude with increasing DTN was even more pronounced in high performers. This might represent a neural substrate for their fine-grained typicality perception, and complements studies reporting increased identity aftereffects (more efficient norm-based coding) in good recognizers (Engfors et al., 2017; Rhodes et al., 2014). To our knowledge, this difference is investigated here for the first time, and the present study thus represents initial evidence that differences in individual face space contribute to recognition performance.

Notably, typicality processing appears to be relevant not only for face recognition, but also in the formation of impressions of other people. One important finding is that more typical faces (e.g., faces morphed towards the average), overall are rated more likeable and attractive than less typical faces, an effect called “beauty-in-averageness” (e.g., Langlois & Roggman, 1990; Zäske et al., 2020). In order to explain this finding, it has been suggested that processing of face typicality follows an efficient coding strategy (Ryali et al., 2020): Faces are encoded in relation to their difference to a norm face (statistically typical face), which is an efficient way to represent faces because it minimizes coding costs. The less typical a face is, the higher the mental effort should be to encode it, because the difference to the norm is larger. As a result, such faces receive negative affective ratings (e.g., less likely/attractive). Relating the efficient coding hypothesis to the current individual differences in the P200 DTN effect, the question arises whether high vs. low performers might differ in their efficiency in representing faces and eventually their social judgements. This might be a fruitful research direction for future studies.

Our second hypothesis, that high performers should show, by definition, increased behavioural and neural correlates of familiarity, was overall supported too. Behaviourally, our selected CFMT+ performers also performed better at another face memory task, the BFFT, and an unfamiliar face matching task, the GFMT. Critically, at the neural level, we found significant group differences in the N250 familiarity-polarity effect, which is in line with previous findings relating N250 amplitudes and face memory accuracy (Kaltwasser et al., 2014, p. 872, Table 2). As shown in Figure 7, high performers showed a larger N250 amplitude than low performers, when observing familiar original faces, the only faces that were really familiar to the observers, vs. all other face types. This finding, on one hand, further supports the notion of the N250 as a marker of face familiarity (Kaufmann et al., 2009; Schweinberger et al., 2002; Tanaka et al., 2006; Wiese et al., 2019b). On the other hand, the N250 represents a key neural substrate of superior face encoding in high performers, which has been reported in previous studies (Herzmann et al., 2010; Huang et al., 2017; Kaltwasser et al., 2014; Kaufmann et al., 2013, but see Belanova et al., 2018). To put our findings in perspective, it must be mentioned that the group difference in familiarity processing was not as robust as, for instance, the reported group difference in DTN processing (see footnote in results section). The claim that familiarity processing is enhanced in high performers is nevertheless supported by enhanced responses to original familiar faces in high performers in the late negativity time range.

Neural Individual Differences in Face Processing across Milliseconds

In this section we sketch a synthetic time course of the neural differences between high and low performers in two effects of special interest, namely those of DTN and the interaction between polarity and familiarity (i.e., the difference between truly familiar and unfamiliar faces).

Group differences in DTN effects seem to appear between 140 and 240 milliseconds (N170-P200 time range), and reach their maximum size after 190 ms. Then, high and low performers do not seem to differ in their DTN neural effects after 260 ms. While the DTN effects in the P200 replicated the general pattern reported in a previous study (Wuttke & Schweinberger, 2019) and were substantially larger in high performers, DTN effects in the preceding N170 were less consistent, and were in fact larger in low performers. If this finding is substantiated in future research, it might provide another perspective on performance-related differences in the time course and specific neural processing of DTN information in face perception.

We were also interested in group differences between original familiar faces and the other face types across time. Such group differences did not emerge before 250 ms after stimulus onset, but hereafter high performers showed increased occipito-temporal negativity to truly familiar (vs. not familiar) faces between 260 and 400 ms (N250), a difference which persisted until about 800 ms (late negativity).

Conclusions

The present research shows that event-related brain potentials can offer substantial insights into the understanding of individual differences in face recognition. Face learning and memory performance appears to be substantially related to neural processing of typicality, as reflected in the P200 ERP - with high performers showing evidence for an “expanded” face space. Neural processing of familiarity, as reflected in the N250 and LN ERP components, also appears to be related to individual differences in face recognition, although perhaps to a lesser extent than expected. The present study established that these relationships are not simply explained by differences in autistic traits or in low-level visual processing, but they also provide evidence to suggest that these factors should be systematically considered in this type of research.

References

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *Quarterly Journal of Experimental Psychology*, *70*(8), 1620-1632.

Arrington, M., Elbich, D., Dai, J., Duchaine, B., & Scherf, K. S. (2022). Introducing the female Cambridge face memory test–long form (F-CFMT+). *Behavior Research Methods*, 1-14.

Bach, M. (1996). The Freiburg Visual Acuity Test-automatic measurement of visual acuity. *Optometry and Vision Science*, *73*(1), 49-53.

Bach, M. (2006). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965-971.

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods, 44*(1), 158–175.

Balsdon, T., Summersby, S., Kemp, R. I., & White, D. (2018). Improving face identification with specialist teams. *Cognitive Research: Principles and Implications*, *3*(1), 1-13.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17.

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex*, *108*, 92-111.

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PloS one*, *11*(2), e0148148.

Boutet, I., Taler, V., & Collin, C. A. (2015). On the particular vulnerability of face recognition to aging: a review of three hypotheses. *Frontiers in Psychology*, *6*, 1139.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305-327.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286-291.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943-958.

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*(4), 482–496.

Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: current perspectives. *Eye and Brain*, 165-175.

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face processing in developmental prosopagnosia. *Brain*, *137*(6), 1781-1798.

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012a). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, *23*(11), 1279-1287.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012b). The Cambridge Car Memory test: A task matched in format to the Cambridge face memory test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587-605.

Dobel, C., Putsche, C., Zwitserlood, P., & Junghöfer, M. (2008). Early left-hemispheric dysfunction of face processing in congenital prosopagnosia: an MEG study. *PloS One*, *3*(6), e2326.

Dolzycka, D., Herzmann, G., Sommer, W., & Wilhelm, O. (2014). Can training enhance face cognition abilities in middle-aged adults?. *PloS one*, *9*(3), e90249.

Dougherty, B. E., Martin, S. R., Kelly, C. B., Jones, L. A., Raasch, T. W., & Bullimore, M. A. (2009). The development of a battery of functional tests for low vision. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *86*(8), 955.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576-585.

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain*, *135*(2), 542-554.

Engfors, L. M., Jeffery, L., Gignac, G. E., & Palermo, R. (2017). Individual differences in adaptive norm-based coding and holistic coding are associated yet each contributes uniquely to unfamiliar face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 281.

Faerber, S. J., Kaufmann, J. M., & Schweinberger, S. R. (2015). Early temporal negativity is sensitive to perceived (rather than physical) facial identity. *Neuropsychologia*, 75, 132-142.

Faerber, S. J., Kaufmann, J. M., Leder, H., Martin, E. M., & Schweinberger, S. R. (2016). The role of familiarity for representations in norm-based face space. *PloS One*, *11*(5), e0155380.

Faghel-Soubeyrand S., Ramon M., Bamps E., Zoia M., Woodhams J., Richoz A., Caldara R., Gosselin F, Charest I., (2022), The neural code behind face recognition abilities. bioRxiv 2022.03.19.484245

Fast, K., Fujiwara, E., & Markowitsch, H. J. (2008). Der famous faces test. *Hogrefe: Göttingen*.

Fisher, K., Towler, J., & Eimer, M. (2016). Reduced sensitivity to contrast signals from the eye region in developmental prosopagnosia. *Cortex*, *81*, 64-78.

Freiwald, W. A. (2020). The neural mechanisms of face processing: cells, areas, networks, and models. *Current Opinion in Neurobiology*, *60*, 184-191.

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637-1651.

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British Journal of Psychology*, *109*(2), 219-231.

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: Implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233.

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia*, *49*(9), 2736-2745.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know?. *Proceedings of the Royal Society B*, *285*(1888), 20181319.

Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport*, *11*(9), 1871-1875.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223-233.

Hendel, R. K., Starrfelt, R., & Gerlach, C. (2019). The good, the bad, and the average: Characterizing the relationship between face and object processing across the face recognition spectrum. *Neuropsychologia*, *124*, 274-284.

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*(5), 688-701.

Herzmann, G., Kunina, O., Sommer, W., & Wilhelm, O. (2010). Individual differences in face cognition: brain–behavior relationships. *Journal of Cognitive Neuroscience*, *22*(3), 571-589.

Hirji, S. H., Liebmann, J. M., Hood, D. C., Cioffi, G. A., & Blumberg, D. M. (2020). Macular Damage in Glaucoma is Associated With Deficits in Facial Recognition. *American Journal of Ophthalmology*, *217*, 1–9.

Hotta, T., Satoh, S., Kosaka, N., & Kohda, M. (2017). Face recognition in the Tanganyikan cichlid Julidochromis transcriptus. *Animal Behaviour*, *127*, 1-5.

Huang, W., Wu, X., Hu, L., Wang, L., Ding, Y., & Qu, Z. (2017). Revisiting the earliest electrophysiological correlate of familiar face recognition. *International Journal of Psychophysiology*, *120*, 42-53.

Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuroimage*, *15*(2), 353-372.

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14*(2), 132-142.

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*, 736-747.

Itz, M. L., Golle, J., Luttmann, S., Schweinberger, S. R., & Kaufmann, J. M. (2017). Dominance of texture over shape in facial identity processing is modulated by individual abilities. *British Journal of Psychology*, *108*(2), 369-396.

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.

Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 861-878.

Kaufmann, J. M., Schweinberger, S. R., & Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience*, *21*(4), 625-641.

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*(1), 21-33.

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*(7), 1310-1319.

Kendrick, K. M., Atkins, K., Hinton, M. R., Heavens, P., & Keverne, B. (1996). Are faces special for sheep? Evidence from facial and object discrimination learning tests showing effects of inversion and social familiarity. *Behavioural Processes*, *38*(1), 19-35.

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, *140*(15), 1617-1622.

Klin, A., Sparrow, S. S., De Bildt, A., Cicchetti, D. V., Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders*, *29*(6), 499-508.

Kloth, N., Rhodes, G., & Schweinberger, S. R. (2017). Watching the brain recalibrate: Neural correlates of renormalization during face adaptation. *NeuroImage*, *155*, 1-9.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*(2), 115-121.

Latinus, M., & Taylor, M. J. (2006). Face processing stages: impact of difficulty and the separation of effects. *Brain Research*, *1123*(1), 179-187.

Limbach, K., Kaufmann, J. M., Wiese, H., Witte, O. W., & Schweinberger, S. R. (2018). Enhancement of face-sensitive ERPs in older adults induced by face recognition training. *Neuropsychologia*, *119*, 197-213.

Limbach, K., Itz, M. L., Schweinberger, S. R., Jentsch, A. D., Romanova, L., & Kaufmann, J. M. (2022). Neurocognitive effects of a training program for poor face recognizers using shape and texture caricatures: A pilot investigation. *Neuropsychologia*, *165*, 108133.

Maurer, D., Mondloch, C. J., & Lewis, T. L. (2007). Effects of early visual deprivation on perceptual and cognitive development. *Progress in Brain Research*, *164*, 87-104.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1), 1-15.

Müller, V. I., Höhner, Y., & Eickhoff, S. B. (2018). Influence of task instructions and stimuli on the neural network of face processing: An ALE meta-analysis. *Cortex*, *103*, 240-255.

Nador, J. D., Zoia, M., Pachai, M. V., & Ramon, M. (2021). Psychophysical profiles in super-recognizers. *Scientific Reports*, *11*(1), 1-11.

Newcombe, F., de Haan, E. H., Ross, J., & Young, A. W. (1989). Face processing, laterality and contrast sensitivity. *Neuropsychologia*, *27*(4), 523-538.

Norton, D., McBain, R., & Chen, Y. (2009). Reduced ability to detect facial configuration in middle-aged and elderly individuals: associations with spatiotemporal visual processing. *Journals of Gerontology: Series B*, *64*(3), 328-334.

Noyes, E., Hill, M. Q., & O'Toole, A. J. (2018). Face recognition ability does not predict person identification performance: Using individual data in the interpretation of group results. *Cognitive Research: Principles and Implications*, *3*(1), 1-13.

Olivares, E. I., Iglesias, J., Saavedra, C., Trujillo-Barreto, N. J., & Valdés-Sosa, M. (2015). Brain signals of face processing as revealed by event-related potentials. *Behavioural Neurology*, 514361.

Owsley, C., & Sloane, M. E. (1987). Contrast sensitivity, acuity, and the perception of 'real-world' targets. *British Journal of Ophthalmology*, *71*(10), 791-796.

Pierce, L. J., Scott, L. S., Boddington, S., Droucker, D., Curran, T., & Tanaka, J. W. (2011). The n250 brain potential to personally familiar and newly learned faces and objects. *Frontiers in Human Neuroscience*, *5*, 111.

Popova, T., & Wiese, H. (2022). The time it takes to truly know someone: Neurophysiological correlates of face and identity learning during the first two years. *Biological Psychology*, *170*, 108312.

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*, *110*(3), 461-479.

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., & Ewing, L. (2014). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 897.

Ross, D. A., Deroche, M., & Palmeri, T. J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*, *21*(1), 47-70.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252-257.

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: No special role for surface reflectance processing. *Neuropsychologia*, *50*(2), 334-340.

Ryali, C. K., Goffin, S., Winkielman, P., & Yu, A. J. (2020). From likely to likable: The role of statistical typicality in human social assessment of faces. *Proceedings of the National Academy of Sciences*, *117*(47), 29371-29380.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. PNAS *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), 12887-12892.

Sheehan, M. J., & Tibbetts, E. A. (2011). Specialized face learning is associated with individual recognition in paper wasps. *Science*, *334*(6060), 1272-1275.

Schweinberger, S. R., Pfütze, 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*(3), 722.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, *14*(3), 398-409.

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. Cortex, 80, 141-153.

Schulz, C., Kaufmann, J. M., Kurt, A., & Schweinberger, S. R. (2012a). Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces. *NeuroImage*, *63*(1), 491-500.

Schulz, C., Kaufmann, J. M., Walther, L., & Schweinberger, S. R. (2012b). Effects of anticaricaturing vs. caricaturing and their neural correlates elucidate a role of shape for face learning. *Neuropsychologia*, *50*(10), 2426-2434.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. https://doi.org/10.1016/j.visres.2009.08.005

Sommer, W., Stapor, K., Kończak, G., Kotowski, K., Fabian, P., Ochab, J., ... & Ślusarczyk, G. (2021). The N250 event-related potential as an index of face familiarity: a replication study. *Royal Society Open Science*, *8*(6), 202356.

Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, *18*(9), 1488-1497.

Tibbetts, E. A., Den Uyl, J., Dwortz, M., & McLean, C. (2019). The development and evolution of specialized face learning in paper wasps. *Animal Behaviour*, *147*, 1-7.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161-204.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*(10), 1996-2019.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*(3), 1060-1084.

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, 1-14.

Wiese, H., Ingram, B. T., Elley, M. L., Tüttenberg, S. C., Burton, A. M., & Young, A. W. (2019a). Later but not early stages of familiar face recognition depend strongly on attentional resources: Evidence from event-related brain potentials. *Cortex*, *120*, 147-158.

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019b). A robust neural index of high face familiarity. *Psychological Science*, *30*(2), 261-272.

Wiese, H., Hobden, G., Siilbek, E., Martignac, V., Flack, T. R., Ritchie, K. L., Young, A. W., & Burton, A. M. (2022a). Familiarity is familiarity is familiarity: Event-related brain potentials reveal qualitatively similar representations of personally familiar and famous faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(8), 1144–1164.

Wiese, H., Anderson, D., Beierholm, U., Tüttenberg, S. C., Young, A. W., & Burton, A. M. (2022b). Detecting a viewer's familiarity with a face: Evidence from event-related brain potentials and classifier analyses. *Psychophysiology*, *59*(1), e13950.

Wuttke, S. J., & Schweinberger, S. R. (2019). The P200 predominantly reflects distance-to-norm in face space whereas the N250 reflects activation of identity-specific representations of known faces. *Biological Psychology*, *140*, 86-95.

Young, A. W., & Burton, A. M. (2018). Are We Face Experts? *Trends in Cognitive Sciences*, *22*(2), 100–110.

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398-410.

Zäske, R., Skuk, V. G., & Schweinberger, S. R. (2020). Attractiveness and distinctiveness between speakers' voices in naturalistic speech and their faces are uncorrelated. *Royal Society Open Science*, *7*(12), 201244.

Zheng, X., Mondloch, C. J., & Segalowitz, S. J. (2012). The timing of individual face recognition in the brain. *Neuropsychologia*, *50*(7), 1451-1461.