


ORIGINAL ARTICLE

The Neural Dynamics of Familiar Face Recognition

Géza Gergely Ambrus ¹, Daniel Kaiser², Radoslaw Martin Cichy^{2,3,4} and Gyula Kovács¹

¹Institute of Psychology, Friedrich Schiller University Jena, Leutragraben 1, 07743 Jena, Germany, ²Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany, ³Berlin School of Mind and Brain, Humboldt-Universität Berlin, Luisenstraße 56, Haus 1, 10117 Berlin, Germany and ⁴Bernstein Center for Computational Neuroscience Berlin, Philippstraße 13/Haus 6, 10115 Berlin, Germany

Address correspondence to email: g.ambrus@gmail.com  orcid.org/0000-0002-8400-8178

Géza Gergely Ambrus and Daniel Kaiser contributed equally to the study

Abstract

In real-life situations, the appearance of a person's face can vary substantially across different encounters, making face recognition a challenging task for the visual system. Recent fMRI decoding studies have suggested that face recognition is supported by identity representations located in regions of the occipitotemporal cortex. Here, we used EEG to elucidate the temporal emergence of these representations. Human participants viewed a set of highly variable face images of 4 highly familiar celebrities (2 males and 2 females), while performing an orthogonal task. Univariate analyses of event-related EEG responses revealed a pronounced differentiation between male and female faces, but not between identities of the same sex. Using multivariate representational similarity analysis, we observed a gradual emergence of face identity representations, with an increasing degree of invariance. Face identity information emerged rapidly, starting shortly after 100 ms from stimulus onset, but was modulated by sex differences and image similarities. From 400 ms after onset and predominantly in the right hemisphere, identity representations showed 2 invariance properties: 1) they equally discriminated identities of opposite sexes and of the same sex, and 2) they were tolerant to image-based variations. These invariant representations may be a crucial prerequisite for successful face recognition in everyday situations, where the appearance of a familiar person can vary drastically.

Key words: EEG, face perception, familiarity, identity representation, MVPA

Introduction

Efficient face recognition is a key ability in human's everyday lives, and many studies have investigated its underlying neural mechanisms (Gobbini and Haxby 2007; Duchaine and Yovel 2015). Recently, much progress has been made in spatially pinpointing the neural correlates of face recognition by advances in multivariate classification techniques for fMRI data (Anzellotti and Caramazza 2014; Guntupalli et al. 2017). These techniques have allowed researchers to decode face identity from different regions of the face processing network, such as from the fusiform face area (FFA) (Gilaie-Dotan and Malach 2007; Nestor et al. 2011; Goesaert and Op de Beeck 2013;

Verosky et al. 2013; Anzellotti et al. 2014; Axelrod and Yovel 2015; Weibert et al. 2016), the anterior temporal lobe (ATL) (Kriegeskorte et al. 2007; Nasr and Tootell 2012; Anzellotti et al. 2014) or from a larger network extending from early visual areas towards the inferior frontal gyrus (Guntupalli et al. 2017; Visconti Di Oleggio Castello et al. 2017).

The temporal emergence of face identity representations, however, remains relatively unexplored. Most of our knowledge on the temporal dynamics of face recognition stems from EEG and magnetoencephalography (MEG) studies employing traditional, univariate analyses on temporally confined ERP/MEP components. Across these studies, the components associated

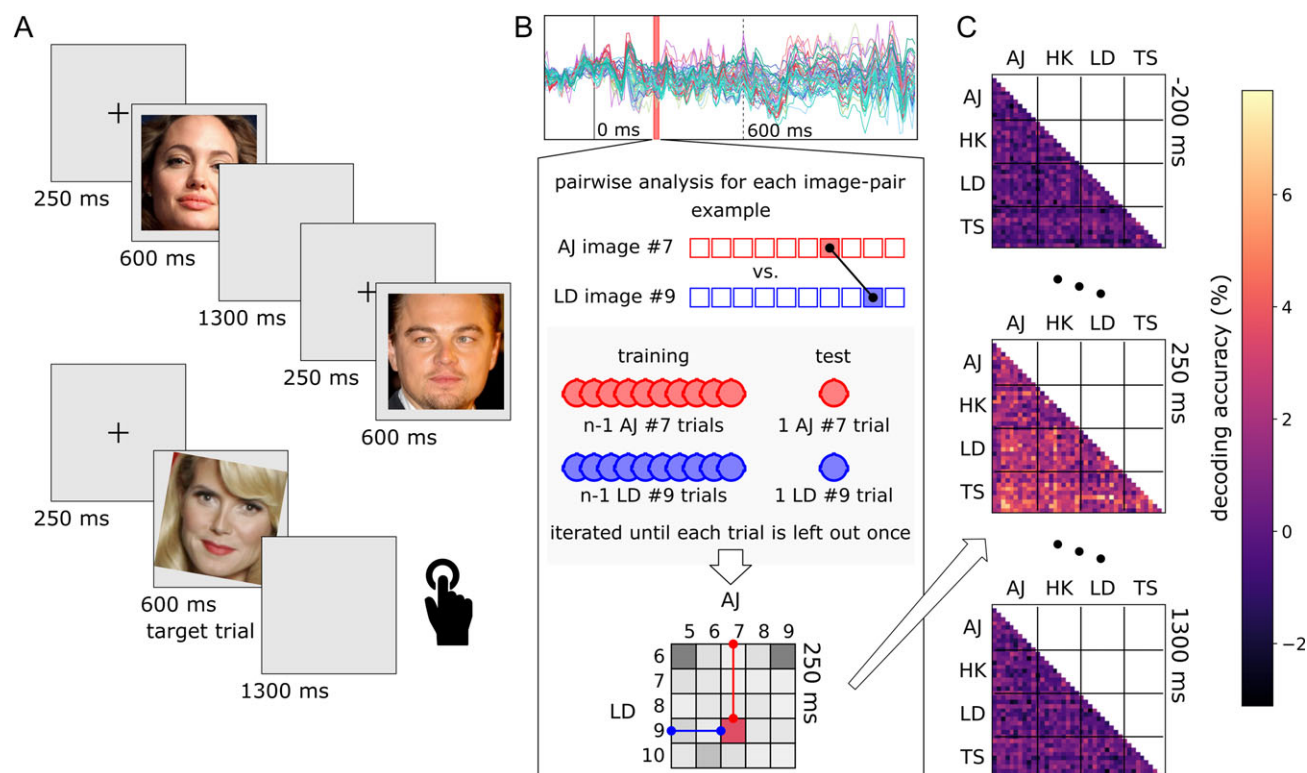


Figure 1. Design and analysis approach. (A) Trial structure and stimulus examples. Stimuli were color, “ambient”, face-cropped images of 4 highly recognized celebrities (Angelina Jolie (AJ), Heidi Klum (HK), Leonardo DiCaprio (LD), Til Schweiger (TS)) (Image credits: File:Angelina Jolie at Davos crop.jpg. (2014, April 23). Wikimedia Commons, the free media repository. Retrieved 15:17, 1 May 2018 from https://commons.wikimedia.org/w/index.php?title=File:Angelina_Jolie_at_Davos_crop.jpg&oldid=122076100. Creative Commons Attribution-Share Alike 3.0 Unported license. File:LeonardoDiCaprioNov08.jpg. (2018, January 20). Wikimedia Commons, the free media repository. Retrieved 15:18, 1 May 2018 from <https://commons.wikimedia.org/w/index.php?title=File:LeonardoDiCaprioNov08.jpg&oldid=281183411>. Creative Commons Attribution-Share Alike 3.0 Unported license. Heidi Klum: File:Project Runway (8287126501).jpg. (2017, December 8). Wikimedia Commons, the free media repository. Retrieved 15:54, January 30, 2019 from [https://commons.wikimedia.org/w/index.php?title=File:Project_Runway_\(8287126501\).jpg&oldid=271028135](https://commons.wikimedia.org/w/index.php?title=File:Project_Runway_(8287126501).jpg&oldid=271028135). Creative Commons Attribution 2.0 Generic license. These images were not part of the original stimulus set). Each trial started with a fixation cross (250 ms), followed by the stimulus image (600 ms) and a blank screen (1300 ms). Target trials containing a tilted stimulus (illustrated on the bottom) were included to ensure that participants maintained attention. (B) The logic of the multivariate pattern analysis. Top: A representative ERP recording from one participant. EEGs were segmented between -200 and 1300 ms relative to stimulus onset. Bottom: For each time point separately, linear classification analyses were performed for each combination of individual images, using a leave-one-trial-out scheme. This procedure resulted in a 40×40 matrix (i.e., 10 images for each of the 4 identities) of decoding accuracies at each time point. (C) Representational dissimilarity matrices (RDMs) from a single representative participant showing pairwise decoding accuracies at -200, 250, and 1300 ms relative to stimulus onset. The color bar presents decoding accuracy relative to chance level (50%).

with face recognition vary substantially: several reports have linked face recognition to the P100 and N170 components (Debruille et al. 1998; Heisz et al. 2006; Caharel et al. 2009; Rousselet et al. 2009; Liu et al. 2013), others have stressed the role of the later N250 and N400 components (Bentin and Deouell 2000; Schweinberger et al. 2002; Huddy et al. 2003; Tanaka et al. 2006; Curran and Hancock 2007; Gosling and Eimer 2011; Jin et al. 2012).

So far only 3 studies have used multivariate pattern analysis (MVPA) to evaluate the temporal dynamics of face identity processing (Vida et al. 2017; Nemrodov et al. 2016, 2018). All 3 investigated the temporal emergence of identity representations across changes in emotional expression, revealing that identity representations emerge relatively early within the first 200 ms after stimulus onset.

However, these previous studies suffer from 2 critical shortcomings. First, they used unfamiliar faces, whose processing is assumed to be markedly different from the processing of familiar faces, as reflected both in behavioral performance (for a review, see Johnston and Edmonds 2009) and neural activations (Natu and O’Toole 2011). Second, variability across images of

the same identity was very limited, leaving it unclear how their results generalize to everyday face recognition where individual encounters with highly variable, “ambient” face images give rise to drastic visual differences (Burton 2013; Young and Burton 2017; Kramer et al. 2018).

In the current EEG study, we provide a temporal characterization of face identity processing, which eliminates both shortcomings: First, we used images of 4 celebrities, who were highly familiar to the participants (Fig. 1A). Second, for each identity, we used 10 “ambient” images (Jenkins et al. 2011), which varied substantially in a range of properties, such as viewpoint, lighting, and expression.

Using representational similarity analysis (RSA) (Kriegeskorte and Kievit 2013), we show that the earliest representations of facial identity emerge shortly after 100 ms post-stimulus and most robustly in posterior electrodes. Later representations, emerging from 400 ms onwards and in electrodes over right occipitotemporal cortex, contained identity information for faces of the same sex and were invariant to image-based properties. Our results suggest that familiar face recognition is supported by fine-grained neural representations in the

face processing network, where identity information over time becomes increasingly invariant to other visual and conceptual properties of the face.

Methods

Participants

A total of 26 healthy participants (6 males, 20 right handed), with an average age of 25 years ($SD = 5.0$) took part in the study in exchange for partial course credits or monetary compensation. The experiment was conducted in accordance with the guidelines of the Declaration of Helsinki, and with the approval of the ethics committee of the University of Jena. Written informed consent was acquired from all participants.

Stimuli

The stimuli were ambient, color photographs of 2 female (Angelina Jolie, AJ; Heidi Klum, HK) and 2 male (Leonardo DiCaprio, LD; Til Schweiger, TS) celebrities. We selected these celebrities based on a pilot survey where we collected familiarity ratings across a range of well-known celebrities in Germany. For each identity, 10 images were selected from a pool of web-scraped photographs, pre-screened for quality. Stimulus images were cropped to a rectangle centered on the inner features of the face (Fig. 1A). To ensure substantial variation across images depicting the same identity, we selected images that minimized the structural similarity index (SSIM; Wang et al. 2004) among images of the same identity, while maximizing it among images of identities of the same sex; this was achieved by using random combination sorting with 100 000 iterations per sex category. The resulting mean SSIM values were: LD: 0.387, TS: 0.378, LD versus TS: 0.355; AJ: 0.379, HK: 0.371, AJ versus HK: 0.337. Stimuli were presented centrally on a uniform gray background on a TFT display (1680×1050 pixel resolution, refresh rate 60 Hz). The experiment was written in Psychopy (Peirce 2008).

Experimental Procedure

A total of 1760 trials (1600 nontarget and 160 target) were presented in 40 runs, each containing the 10 images of the 4 identities once, in a pseudorandom order (with the constraint that the same identity was never repeated in 2 consecutive trials). Thus, photos of one identity were seen 400 times, so that a given image of a given identity was presented 40 times during the experiment. Each trial started with a fixation cross (250 ms), followed by the stimulus image (600 ms, subtending a visual angle of 4.4° in diameter) and finally a blank display (1300 ms). Short breaks were provided after every 10 runs, but the run boundaries were not indicated otherwise. There were 4 target-trials in each run, where the image was rotated 10° clockwise or anticlockwise. Participants were instructed to press the space bar when they saw a target image (the overall detection accuracy was at $99.52 \pm 0.67\%$). These target trials served to ensure that the participants maintained their attention, and were not included in the analysis. An average experimental session lasted $81.5 (\pm 5.3)$ min.

At the beginning of each experiment, prior to mounting the electrode caps, participants were presented images of the 4 identities and were asked to name them. All our participants were able to name all 4 celebrities correctly. The images of this initial familiarity-testing phase were not part of the later EEG experiment. After the EEG recording, participants were asked to

rate their familiarity with the identities on a 7-point scale. Mean ratings were generally high (AJ: 6.12, HK: 6.30, LD: 6.15, TS: 6.11) and not statistically different for the 4 identities ($F[3,75] = 0.243$, $P = 0.866$).

EEG Recording and Preprocessing

The EEG was recorded in a dimly lit, electrically shielded, and sound-attenuated chamber. The distance between the eyes and the computer screen was set to 96 cm via a chin rest. The electroencephalogram (EEG) was recorded with a 512 Hz sampling rate (bandwidth: DC to 120 Hz) using a 64-channel Biosemi Active II system. Electrooculogram (EOG) was recorded from the outer canthi of the eyes and from above and below the left eye.

The preprocessing pipeline was implemented in MNE-python (Gramfort et al. 2013, 2014). EEG was notch-filtered at 50 Hz, band-pass filtered between 0.1 and 70 Hz, segmented from -200 to 1300 ms relative to stimulus onset, and baseline corrected with respect to the first 200 ms. Artifact rejection was carried out using the “Autoreject” algorithm (Jas et al. 2017). The resulting data was downsampled to 100 Hz to increase signal-to-noise ratio in the multivariate analyses (Grootswagers et al. 2017).

Event-Related Potentials

To test for the presence of identity-related information within the conventional ERPs we averaged data across repetitions for each facial identity, electrode and participant separately. Next, we created grand-averages of these data across 6 regions of interests, corresponding to the left and right anterior (Fp, AF, F, FC), central (FT, TP, C, CP, T) and posterior occipito-temporal electrodes (PO, P, O, I). The central electrodes (Fpz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz, Iz) were included in both the left and the right clusters; this was done to maintain sufficient electrode counts for the multivariate analyses (see below). For reasons of consistency, the same electrode clusters were used in both analyses. The posterior clusters included the electrodes typically yielding the largest face-sensitive N170 components (Rossion and Jacques 2008). First, we tested for identity selectivity by using a one-way repeated measures ANOVA with identity (4) as a factor. Second, we averaged the 2 females and 2 males face elicited ERPs and performed a paired t-test for testing sex-specific differences. Third, we tested if the ERPs differed for the 2 identities within the same sex by comparing the ERPs for the 2 female as well as for the 2 male identities with each other in t-tests.

Representational Similarity Analysis

To model the neural organization of face representations, we performed a RSA (Kriegeskorte 2008) on the EEG data. In this analysis (Fig. 1B,C), the neural dissimilarity between all pairs of face images (i.e., between all 40 individual images), was modeled as a function of different predictor matrices (see below).

Neural dissimilarity: Neural dissimilarity was extracted by performing a linear classification analysis, where pairwise decoding accuracies were used as a measure of representational dissimilarity. Classification analysis was carried out using the CoSMoMVPa toolbox (Oosterhof et al. 2016). Linear-discriminant-analysis (LDA) classifiers were trained and tested on response patterns across all 64 electrodes, separately for each time point across the epoch (downsampled to 100 Hz, i.e.,

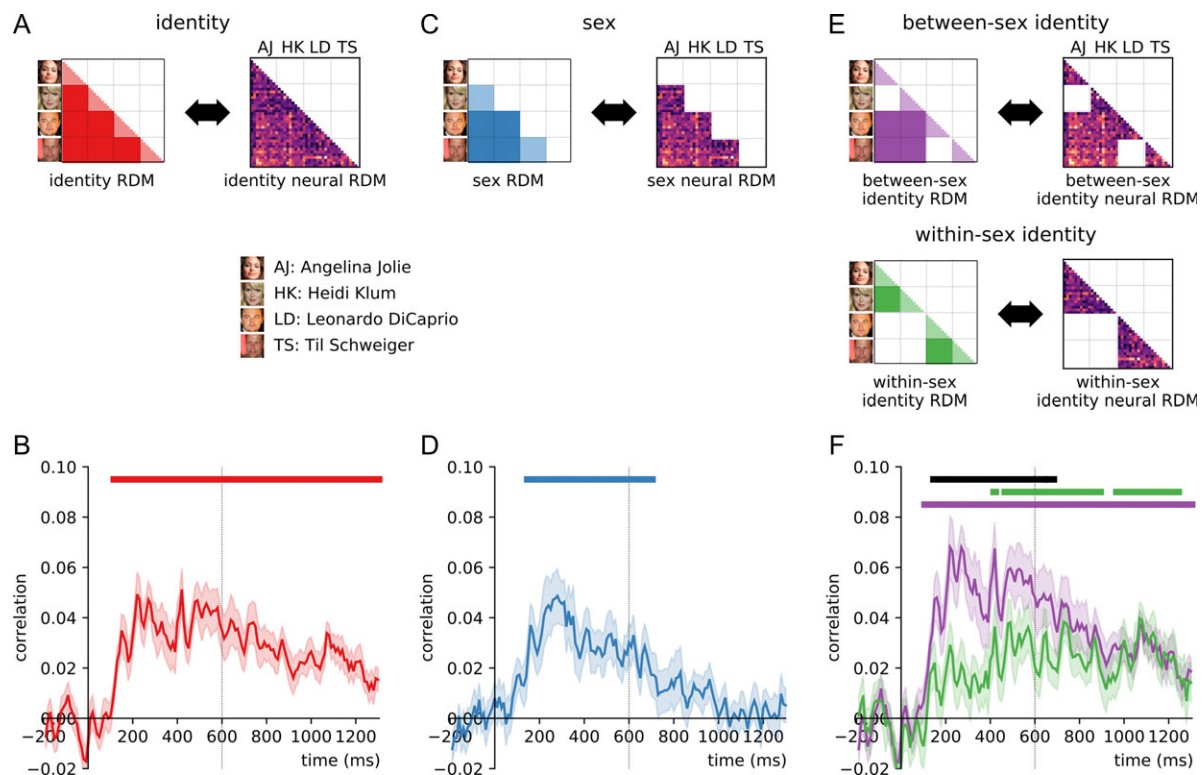


Figure 2. RSA results. To reveal identity-specific representations, we modeled the representational organization obtained from EEG signals with different predictor matrices (A, C, E). We observed temporally persisting identity information starting from 110 ms after stimulus onset (A, B). Similarly, we found strong sex information in the neural organization, emerging between 140 and 680 ms (C, D). Tracking identity information for faces of same and opposite sexes revealed that identity information for same-sex faces was relatively delayed, emerging only after 400 ms (E, F). Early identity information was significantly reduced for between-sex comparisons (black significance markers), suggesting that early identity coding partly relies on differences in face sex. Horizontal lines denote statistical significance ($P < 0.05$, corrected for multiple comparisons). Shaded ranges denote standard errors of the mean.

with a 10 ms resolution) and separately for each pair of images. Training and testing was done in a leave-one-out scheme (Fig. 1B): classifiers were trained on all but one trials for each of the 2 conditions, and tested on the left-out trials. This procedure was repeated until each trial was left out once, and classification accuracy was averaged across these repetitions. Pairwise classification time-courses were smoothed with a 30 ms (i.e., 1 consecutive time points) averaging window (Kaiser et al. 2016). This classification analysis led to one representational dissimilarity matrix (RDM; 40×40 entries, with empty diagonal) for each time point (Fig. 1C).

Modeling neural dissimilarity: To model the neural dissimilarity, we created 4 categorical predictor RDMs. Each predictor RDM covered 40×40 elements, and contained zeros where the entries represented comparisons of similar images (i.e., similar on the dimension of interest, see below) and ones, where the entries reflected comparisons of dissimilar images. To quantify correspondence between the predictor RDMs and the neural RDMs, we unfolded the lower off-diagonal elements of the matrices into 2 vectors (i.e., the diagonal of both matrices was discarded) and correlated the vectors using Spearman's correlation coefficients. These correlations were computed separately for each time point, leading to a time series of correlations that reflected the correspondence of the neural data and the predictor. Individual-participant correlations were Fisher-transformed.

Modeling identity information: For assessing differences between the 4 identities, all comparisons within a given

identity (e.g., 2 images of AJ) were marked as similar (0) and all comparisons between 2 identities (e.g., an image of AJ and an image of TS) were marked as dissimilar (1) (Fig. 2A).

Modeling sex information: For assessing differences between face sexes, all comparisons within the same sex (e.g., an image of AJ and an image of HK) were marked as similar (0), and all comparisons between the different sexes (e.g., an image of AJ and an image of TS) were marked as dissimilar (1). To avoid confounding sex information with identity information, all comparisons within the same identity (e.g., 2 different images of AJ) were excluded from this analysis (as including these comparisons would overestimate the effect of sex) (Fig. 2C).

Modeling identity information between and within sexes: To uncover interactions between sex and identity processing, we constructed identity predictor RDMs that only covered all comparisons across the sexes or within one sex. The between-sex RDM was generated from the identity predictor matrix by removing all comparisons of 2 different identities of the same sex (e.g., an image of AJ and an image of HK), leaving only comparisons within identity (0) and between identities of the opposite sex (1). The within-sex RDM was generated from the identity predictor matrix by removing all comparisons of 2 identities of different sexes (e.g., an image of AJ and an image of TS), leaving only comparisons within identity (0) and between identities of the same sex (1) (Fig. 2E). Note that this within-sex analysis tests for identity representations in more thorough way: by removing between-sex comparisons, the

more pronounced differences between faces of the opposite sex (due to face sex, and due to visual differences) are eliminated.

Sensor-space RSA: To track representational organization across electrode space, we additionally repeated the RSA across the 6 electrode clusters also used in the ERP analysis (see above). Including central electrodes (Fpz, AFz, Fz, FCz, Cz, CPz; Pz, POz, Oz, Iz) in both left- and right-hemispheric clusters yielded electrode counts of 12, 12, and 13, for the anterior, central, and posterior clusters, respectively. All technical details of the cluster-specific RSAs were identical to the analysis using all available electrodes.

Controlling for image similarity: To quantify similarity on the image level, we computed pixel similarities for all pairs of images. Each image (220 × 220 pixels in 3 color layers) was first unfolded into a vector; these vectors were then correlated for each pair of images. A pixel RDM was generated by using 1 – correlation as the dissimilarity measure. As the pixel RDM explained some variance in the face identity RDM ($R^2=0.06$), neural identity representations could in principle partly reflect pixel similarities. Hence, we used a partial correlation approach (Cichy et al. 2017; Groen et al. 2018), where we repeated the key analyses while removing the pixel RDM by partialing it out. This analysis revealed representations of face identity that are invariant to pixel-based image similarities.

Statistical Testing

To identify significant effects across time, we used a threshold-free cluster enhancement procedure (Smith and Nichols 2009) with default parameters. Multiple-comparison correction across time was based on a sign-permutation test (with null distributions created from 10 000 bootstrapping iterations) as implemented in CoSMoMVA (Oosterhof et al. 2016). The resulting statistical maps were thresholded at $Z > 1.64$ (i.e., $P < 0.05$, one sided against zero). For peaks across the time courses, we additionally report uncorrected t -values and Cohen's d as a measure of effect size.

Results

Event-Related Potentials Reflect Face Sex, But not Face Identity

Following traditional EEG studies on face perception, we first performed a univariate ERP analysis across 6 electrode clusters (Supplementary Fig. S1). ERPs were different for the 4 identities primarily in the bilateral posterior electrode clusters (main effect of identity in a 4-way ANOVA, Supplementary Fig. S1E/F, purple line) starting from 100 ms for the left and 120 ms for the right hemisphere (Supplementary Fig. S1), remaining significant throughout the length of the epoch. The other electrode clusters showed weaker and less temporally persistent differences (A–D). The difference between identities however originated from the significantly different ERPs for female and male faces from 190 ms (left) and 150 ms (right), throughout the length of the epoch. By contrast, within-sex comparisons led to no significant results at any of the time-points over any of the electrode clusters. These results support prior studies showing that ERP signals more prominently reflect face sex than face identity (Mouchetant-Rostaing et al. 2000; Freeman et al. 2010). In the following we applied MVPA to further probe the emergence of identity information with higher sensitivity.

Tracking the Emergence of Face Identity Representations

To reveal identity information in the EEG signals, we generated an identity predictor RDM, which reflected the 40 images' dissimilarity in identity (Fig. 2A). We then correlated the neural RDM with this identity RDM separately at every time point. This analysis revealed significant correlations from 110 ms onwards, peaking at around 410 ms (peak $t[25] = 5.97$ Cohen's $d = 1.17$) and lasting across the whole epoch (Fig. 2B), suggesting rapidly emerging and long-lasting face identity information in the signal. Converging results were obtained using classification analysis (see Supplementary Fig. S2).

Our stimulus set contained faces of both sexes, and faces within the same sex share more visual and conceptual properties than faces of opposite sexes (O'Toole et al. 1998). To determine whether such sex differences could be retrieved from the EEG signals, we correlated the neural RDM with a sex predictor RDM separately at every time point (Fig. 2C). This sex predictor RDM only contained between-identity comparisons, so that this analysis reflected face sex independently of identity. We found significant sex information from 140 to 680 ms, peaking at 270 ms (peak $t[25] = 4.39$ Cohen's $d = 0.86$) (Fig. 2D). This indicates that the early EEG signals also contain reliable differences between sexes, emerging at a similar time point as identity-specific information but decaying more rapidly.

The presence of sex information in the signal suggests that identity information may be processed differently as a function of the sex of the face. Specifically, as faces of the same sex are more similar in various aspects (including their visual appearance), discriminating between the 4 facial identities may overestimate the amount of genuine identity information in the signal. We thus split our analysis into comparisons between faces of opposite and of the same sex by correlating the neural RDMs with 2 separate predictor RDMs (Fig. 2E).

For one of these predictor RDMs ("between-sex") we only included comparisons between the 2 sexes, while for the other RDM ("within-sex") we only included comparisons within the same sex. As predicted by the significant contribution of sex information (Fig. 2D), we found that identity information is differently pronounced between and within the 2 sexes. We observed strong sex-dependent identity information that could be retrieved from as early as 100 ms until the end of the epoch and peaking at 260 ms (peak $t[25] = 7.40$ Cohen's $d = 1.45$). Identity information, however, differed when restricting the analysis to within-sex comparisons: it emerged significantly later, at around 400 ms, and peaked at 1050 ms (peak $t[25] = 5.97$ Cohen's $d = 1.17$) (Fig. 2F). When directly comparing identity information for the between- and within-sex comparisons, we found significantly higher identity information for the between-sex analysis between 140 and 660 ms. This suggests that early identity representations partly reflect differences in face sex. By contrast, after 660 ms, face sex did not influence identity representations, suggesting the emergence of identity representations that are invariant to commonalities and differences across the 2 sexes.

Face Identity Information Predominantly Originates From Right Posterior Sources

As highlighted by previous neuroimaging studies (Rossion et al. 2003, 2012) (for a recent review see Yovel 2016), and evident from our univariate results (see above), face-selective responses are strongest over right posterior electrodes. Using

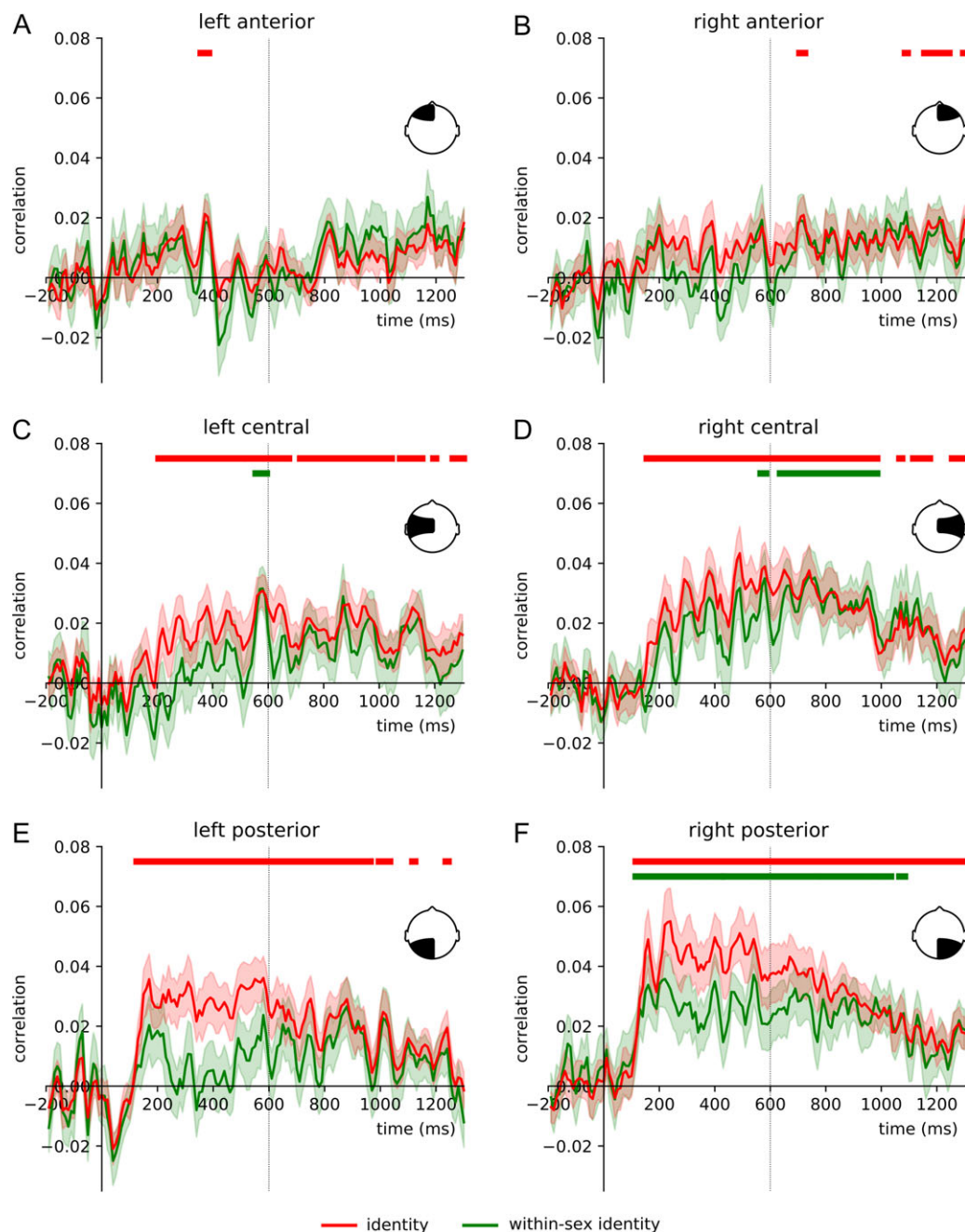


Figure 3. Sensor-space RSA results. When repeating the RSA for the 6 electrode clusters used in the ERP analysis, we found strongest identity information in the posterior clusters (E/F). This identity information was lateralized to the right hemisphere: In the right central and posterior electrode clusters (D/F), we observed significant within-sex identity information, with an early onset (110 ms) in the right posterior cluster. The corresponding left-hemispheric clusters (C/E) only yielded identity information when also the between-sex comparisons were included. The anterior clusters (A/B) did not yield substantial identity information. Horizontal lines denote statistical significance ($P < 0.05$, corrected for multiple comparisons). Shaded ranges denote standard errors of the mean.

response patterns across the whole scalp may therefore partly obscure face identity information in the multivariate analyses. We thus repeated the RSA separately for each of the 6 electrode clusters used in the univariate analysis, expecting the strongest identity information in the right posterior cluster (Fig. 3).

For the posterior electrode clusters we found the most pronounced identity information, and a marked difference between hemispheres. In the left posterior cluster, 4-way identity information (where sex may contribute to identity

encoding) emerged from 120 ms poststimulus onset and peaked at 560 ms (peak $t[25] = 4.85$ Cohen's $d = 0.95$) (Fig. 3). However, restricting the analysis to within-sex comparisons abolished identity information over this electrode cluster in the signal entirely. Similarly, in the right posterior cluster (Fig. 3F) we found robust 4-way identity information, starting from 110 ms after stimulus onset and peaking at 230 ms (peak $t[25] = 4.81$ Cohen's $d = 0.94$). Crucially however, the right posterior cluster also showed reliable within-sex identity information

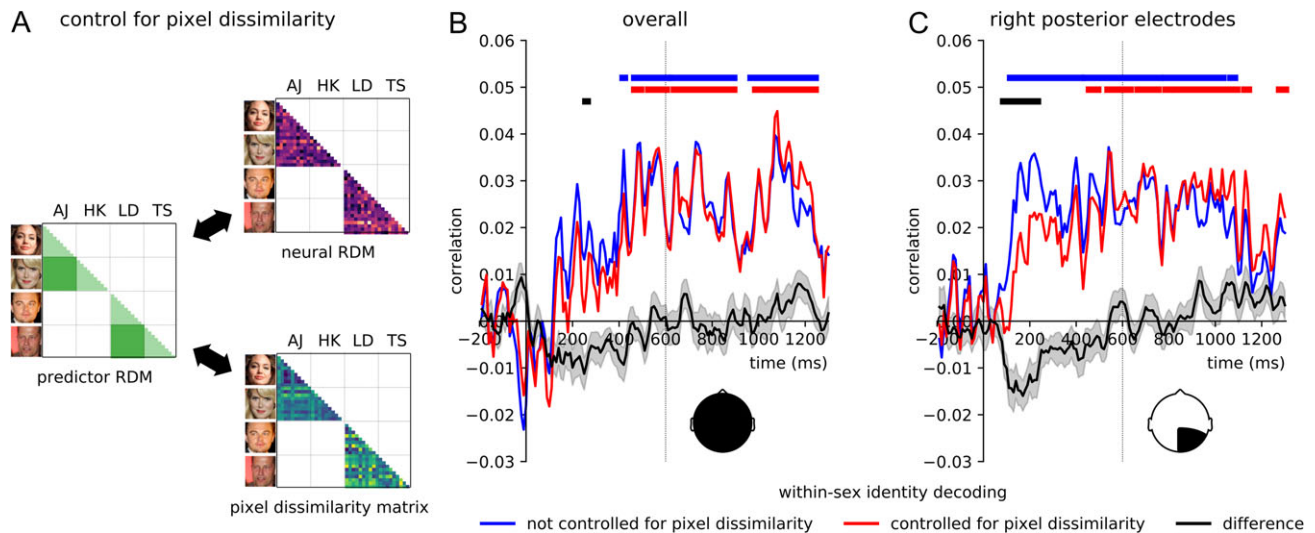


Figure 4. Controlling for image similarity. In a partial correlation analysis, we tracked within-sex identity information while controlling for the images' pixel dissimilarities (A). When using data from all electrodes, removing pixel dissimilarities did not significantly impact identity information (B). For the right posterior cluster, where early within-sex identity information was found in previous analysis (Fig. 3F), controlling for pixel dissimilarity had a significant impact (C): early identity information (90–230 ms) was significantly reduced when controlling for pixel dissimilarity, whereas later identity information was not impacted and remained significant from 460 ms after onset. These results suggest that later representations of face identity are invariant to image-based properties. Horizontal lines denote statistical significance ($P < 0.05$, corrected for multiple comparisons). Shaded ranges denote standard errors.

throughout the epoch, emerging at the same time, after 110 ms and peaking around 530 ms (peak $t[25] = 5.40$ Cohen's $d = 1.06$). This result suggests that signals recorded from electrodes close to the typically face-selective ERP recording sites of the right hemisphere contain widespread identity information, even when visual and conceptual properties are more robustly controlled for.

The left central cluster (Fig. 3C) primarily showed 4-way identity information, emerging slightly later as compared with the posterior cluster, after 200 ms, peaking at 560 ms (peak $t[25] = 5.10$ Cohen's $d = 1.0$). By contrast, the right central cluster not only yielded 4-way identity information (from 150 ms, peaking at 480 ms, peak $t[25] = 4.97$, Cohen's $d = 0.97$), but also within-sex identity information, emerging later than that of the posterior cluster, after 550 ms and peaking at 1100 ms (peak $t[25] = 3.11$, Cohen's $d = 0.61$).

Signals recorded from the 2 anterior clusters did not yield substantial identity information (Fig. 3A,B), suggesting that identity information primarily originates from sources in visual cortex.

Late Representations of Face Identity are Invariant to Image Properties

Our stimulus set was constructed to mirror natural variations across different encounters with a familiar person. This was achieved by selecting stimuli that ensured a high degree of variability within each identity (see above), so that image-based stimulus properties are unlikely to account for the emergence of identity information. To explicitly rule out this possibility, we performed a control analysis, where we additionally modeled image-based similarities between stimuli. This was done by constructing pixel RDMs, which reflected the images' dissimilarity in pixel values; these pixel RDMs were partialled out in the subsequent analysis. We focused the control analysis on the within-sex comparison, which forms the most robust test of face identity representations, and on the 2 electrode

configurations where it was most robustly found (all electrodes and right posterior electrodes).

In the analysis using all electrodes, we found no modulation of identity information after removing the pixel RDM (Fig. 4B). Similarly, sex information was not modulated when controlling for pixel similarity. By contrast, when focusing on the right posterior cluster, we found a modulation of identity information when controlling for image-based similarity (Fig. 4C). Early within-sex identity information, emerging between 90 and 230 ms was significantly reduced when controlling for pixel dissimilarity. By contrast, later within-sex identity information (from 460 ms) emerged independently of image-based properties. Together, these results suggest that later representations of face identity are robust to image-based changes, and genuinely reflect face identity. These neural representations might thus be a crucial prerequisite for efficient face recognition across visually different encounters with a person.

Discussion

In the current study, we applied representational similarity analysis to EEG signals to investigate the neural dynamics of familiar face recognition. Our results show that face identity can be rapidly recovered from EEG response patterns, even with highly variable, "ambient" face stimuli (Jenkins et al. 2011). In more fine-grained analyses, we uncovered a gradual emergence of face identity coding: Early identity information is modulated by face sex and by visual image properties. By contrast, later identity information, emerging after 400 ms and primarily in the right hemisphere, is unaffected by these factors. This finding suggests that after 400 ms representations genuinely reflect face identity. These later representations may be the basis for real-world face recognition, allowing the identification of an individual across different encounters and against similar-looking other faces.

In everyday life, the facial appearance of a single person can be highly variable. This variability makes it challenging to

match an individual encounter with a face to an identity representation stored in memory (Bruce et al. 1999; Clutterbuck and Johnston 2002; Jenkins et al. 2011; Andrews et al. 2015). The invariant identity representations revealed here are ideal for extracting face identity from different encounters, as they discriminate identities of the same sex, across variations in visual properties. The late emergence and prolonged nature (from 400 up to 1300 ms poststimulus onset) of such representations is compatible 1) with the persistent nature of stimulus-selective neural activity in macaque inferior temporal and entorhinal cortices (Fuster and Jervey 1981; Miyashita and Chang 1988; Suzuki et al. 1997; Yakovlev et al. 1998) and in the human medial temporal lobe (Kornblith et al. 2017), 2) with the involvement of mnemonic conceptual identity representations in the medial and anterior temporal cortices (Quiroga et al. 2005; Mormann et al. 2008), and 3) with recent ERP data showing the most prominent differences between highly familiar and unfamiliar faces between 400 and 600 ms (Wiese et al. 2018). Whether the late identity information in EEG signals reflects the representations of complex perceptual features that discriminate identities or whether it truly reflects semantic information about different people needs to be tested in future studies. By linking EEG results with functional neuroimaging data (Cichy et al. 2014, 2016), future studies could directly test where these representations originate. Another promising avenue is to investigate these representations for a greater number of identities (our study was limited to 4 individuals) and under varying task demands.

How do these seemingly late identity representations support rapid face recognition in the wild? While these representations are useful under great variability and in the presence of distracting face information, face recognition is sometimes easier than this: In real-life situations, we often know which person to expect, which visual properties are diagnostic of him or her, and where the person likely shows up. Under such conditions, motor responses in face recognition tasks can be faster than 400 ms (Visconti di Oleggio Castello and Gobbini 2015; Besson et al. 2017). This observation suggests that face identity can sometimes be inferred from earlier representations that do not need to be highly invariant. Future studies could thus test whether different representational stages are crucial for face recognition under varying demands.

Our study revealed a pronounced right-hemispheric lateralization of identity information: face identity information was strongest in electrodes over the right, as opposed to the left, visual cortex. Specifically, only signals recorded over right occipitotemporal cortex contained identity information which is invariant to both face sex and image-based properties. This right-lateralized topography is consistent with sources in the visual face processing network that has a strong right-hemispheric lateralization (Axelrod and Yovel 2015; Yovel 2016). Interestingly, neuroimaging work showed that specifically right-hemispheric activations predict behavioral performance in familiar face recognition (Weibert and Andrews 2015), suggesting that these identity representations could play an important role in face recognition. However, this notion has to be explicitly tested in the future, as caution needs to be applied when inferring cortical sources from EEG scalp topographies.

The timing and topography of the face identity representations revealed here suggest that they support the recognition of familiar people, rather than reflecting low-level visual image discrimination. Face identification can per definition only occur for familiar people (Hancock et al. 2000), and it is therefore

expected that identity information for previously unfamiliar people cannot be equally recovered from EEG signals. However, follow-up studies need to compare familiar and unfamiliar faces to test this prediction explicitly. We speculate that the representations described here are sculpted by our rich experience with the people we know, both on a perceptual level (where invariant perceptual representations are formed) and on a semantic level (where person knowledge is acquired). Testing how stable identity representations emerge over the course of face learning (Andrews et al. 2015) would therefore be an exciting topic for future research in this direction.

Besides identity coding, our findings also offer insights into the cortical coding of face sex. As our stimulus set contained faces of opposite sexes, we could also track the emergence of sex information. Face sex can be rapidly retrieved from EEG signals, both in univariate and multivariate analyses, and predicts cortical organization from 140 ms. This finding corroborates previous ERP studies, which have suggested that face sex is extracted early and affects a variety of face-related ERP components (Mouchetant-Rostaing et al. 2000; Ito and Urland 2003, 2005; Kloth et al. 2015). As opposed to the temporally sustained identity information, sex information displayed a more transient nature, and vanished shortly before 700 ms after onset. This difference between identity and sex information suggests that the 2 properties are coded somewhat independently at later processing stages. It is worth noting that high-level representations of faces contain a plethora of conceptual information about faces: beyond sex and identity, face information can be organized by emotion, age, attractiveness, and many other factors. Our study therefore only offers a selective snapshot of the complex cortical organization of face information, which needs to be more fully characterized in future studies.

In conclusion, we provide a characterization of the neural dynamics underlying familiar face recognition. Representations of face identity emerged gradually across the visual processing cascade. Invariant identity representations were observed after 400 ms of processing. We suggest that these representations are crucial for face recognition across different encounters with a person.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Notes

The authors would like to thank Lisa-Celine Süllwold and Bettina Kamchen for their help in participant recruitment and data acquisition. This work was supported by grants from the Deutsche Forschungsgemeinschaft (KO3918/5-1, KA4683-2/1, CI241-1/1). *Conflicts of interest:* The authors declare no competing financial interests.

References

- Andrews S, Jenkins R, Cursiter H, Burton AM, Andrews S, Jenkins R, Cursiter H, Telling AMB, Andrews S, Jenkins R, et al. 2015. Telling faces together: learning new faces through exposure to multiple instances. *Q J Exp Psychol.* 0218:2041–2050.
- Anzellotti S, Caramazza A. 2014. The neural mechanisms for the recognition of face identity in humans. *Front Psychol.* 5: 672.

- Anzellotti S, Fairhall SL, Caramazza A. 2014. Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex*. 24:1988–1995.
- Axelrod V, Yovel G. 2015. Successful decoding of famous faces in the fusiform face area. *PLoS One*. 10:e0117126.
- Bentin S, Deouell LY. 2000. Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cogn Neuropsychol*. 17:35–55.
- Besson G, Barragan-Jason G, Thorpe SJ, Fabre-Thorpe M, Puma S, Ceccaldi M, Barbeau EJ. 2017. From face processing to face recognition: comparing three different processing levels. *Cognition*. 158:33–43.
- Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P. 1999. Verification of face identities from images captured on video. *J Exp Psychol Appl*. 5:339–360.
- Burton AM.. 2013. Why has research in face recognition progressed so slowly? The importance of variability. *Q J Exp Psychol*. 66:1467–1485.
- Caharel S, Jiang F, Blanz V, Rossion B. 2009. Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. *Neuroimage*. 47:1809–1818.
- Cichy RM, Kriegeskorte N, Jozwik KM, van den Bosch JFF, Charest I. 2017. Neural dynamics of real-world object vision that guide behaviour. *bioRxiv*. 147298. doi: 10.1101/147298.
- Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. *Nat Neurosci*. 17:455–462.
- Cichy RM, Pantazis D, Oliva A. 2016. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cereb Cortex*. 26:3563–3579.
- Clutterbuck R, Johnston RA. 2002. Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*. 31:985–994.
- Curran T, Hancock J. 2007. The FN400 indexes familiarity-based recognition of faces. *Neuroimage*. 36:464–471.
- Debruille JB, Guillem F, Renault B. 1998. ERPs and chronometry of face recognition: following-up Seeck et al. and George et al. *Neuroreport*. 9:3349–3353.
- Duchaine B, Yovel G. 2015. A revised neural framework for face processing. *Annu Rev Vis Sci*. 1:393–416.
- Freeman JB, Ambady N, Holcomb PJ. 2010. The face-sensitive N170 encodes social category information. *Neuroreport*. 21:24–28.
- Fuster JM, Jervey JP. 1981. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*. 212:952–955.
- Gilaie-Dotan S, Malach R. 2007. Sub-exemplar shape tuning in human face-related areas. *Cereb Cortex*. 17:325–338.
- Gobbini MI, Haxby JV. 2007. Neural systems for recognition of familiar faces. *Neuropsychologia*. 45:32–41.
- Goesaert E, Op de Beeck HP. 2013. Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J Neurosci*. 33:8549–8558.
- Gosling A, Eimer M. 2011. An event-related brain potential study of explicit face recognition. *Neuropsychologia*. 49:2736–2745.
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, et al. 2013. MEG and EEG data analysis with MNE-Python. *Front Neurosci*. 7:267.
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS. 2014. MNE software for processing MEG and EEG data. *Neuroimage*. 86:446–460.
- Groen IIA, Greene MR, Baldassano C, Fei-Fei L, Beck DM, Baker CI. 2018. Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife*. 7:e32962 doi: 10.7554/eLife.32962.
- Grootswagers T, Wardle SG, Carlson TA. 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci*. 29:677–697.
- Guntupalli JS, Wheeler KG, Gobbini MI. 2017. Disentangling the representation of identity from head view along the human face processing pathway. *Cereb Cortex*. 27:46–53.
- Hancock PJB, Bruce V, Burton AM.. 2000. Recognition of unfamiliar faces. *Trends Cogn Sci*. 4(9):330–337.
- Heisz JJ, Watter S, Shedden JM. 2006. Automatic face identity encoding at the N170. *Vision Res*. 46:4604–4614.
- Huddy V, Schweinberger SR, Jentzsch I, Burton AM. 2003. Matching faces for semantic information and names: an event-related brain potentials study. *Brain Res Cogn Brain Res*. 17:314–326.
- Ito TA, Urland GR. 2003. Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *J Pers Soc Psychol*. 85:616–626.
- Ito TA, Urland GR. 2005. The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. *Cogn Affect Behav Neurosci*. 5:21–36.
- Jas M, Engemann DA, Bekhti Y, Raimondo F, Gramfort A. 2017. Autoreject: automated artifact rejection for MEG and EEG data. *Neuroimage*. 159:417–429.
- Jenkins R, White D, Van Montfort X, Burton AM. 2011. Variability in photos of the same face. *Cognition*. 121:313–323.
- Jin J, Allison BZ, Kaufmann T, Kübler A, Zhang Y, Wang X, Cichocki A. 2012. The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement. *PLoS One*. 7:e49688.
- Johnston R a, Edmonds AJ. 2009. Familiar and unfamiliar face recognition: a review. *Memory*. 17:577–596.
- Kaiser D, Oosterhof NN, Peelen MV. 2016. The neural dynamics of attentional selection in natural scenes. *J Neurosci*. 36:10522–10528.
- Kloth N, Damm M, Schweinberger SR, Wiese H. 2015. Aging affects sex categorization of male and female faces in opposite ways. *Acta Psychol (Amst)*. 158:78–86.
- Kornblith S, Quiñ Quiroga R, Koch C, Fried I, Mormann F. 2017. Persistent single-neuron activity during working memory in the human medial temporal lobe. *Curr Biol*. 27:1026–1032.
- Kramer RSS, Young AW, Burton AM. 2018. Understanding face familiarity. *Cognition*. 172:46–58.
- Kriegeskorte N. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:4.
- Kriegeskorte N, Formisano E, Sorger B, Goebel R. 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci*. 104:20600–20605.
- Kriegeskorte N, Kievit RA. 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci*. 17:401–412.

- Liu J, Harris A, Kanwisher N. 2013. Stages of processing in face perception: an MEG study. *Soc Neurosci Key Readings*. 5: 75–86.
- Miyashita Y, Chang HS. 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*. 331:68–70.
- Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried I, Koch C. 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J Neurosci*. 28:8865–8872.
- Mouchetant-Rostaing Y, Giard MH, Bentin S, Aguera PE, Pernier J. 2000. Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci*. 12:303–310.
- Nasr S, Tootell RBH. 2012. Role of fusiform and anterior temporal cortical areas in facial recognition. *Neuroimage*. 63: 1743–1753.
- Natu V, O'Toole AJ. 2011. The neural processing of familiar and unfamiliar faces: a review and synopsis. *Br J Psychol*. 102: 726–747. Wiley/Blackwell (10.1111).
- Nemrodov D, Niemeier M, Mok JNY, Nestor A. 2016. The time course of individual face recognition: a pattern analysis of ERP signals. *Neuroimage*. 132:469–476.
- Nemrodov D, Niemeier M, Patel A, Nestor A. 2018. The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. *eNeuro*. 5: ENEURO.0358–17.
- Nestor A, Plaut DC, Behrmann M. 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci*. 108:9998–10003.
- Oosterhof NN, Connolly AC, Haxby JV. 2016. CoSMoMvPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front Neuroinform*. 10:27.
- O'Toole AJ, Deffenbacher KA, Valentin D, McKee K, Huff D, Abdi H. 1998. The perception of face gender: the role of stimulus structure in recognition and classification. *Mem Cognit*. 26: 146–160.
- Peirce JW. 2008. Generating stimuli for neuroscience using PsychoPy. *Front Neuroinformatics*. 2:10.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature*. 435:1102–1107.
- Rossion B, Hanseeuw B, Dricot L. 2012. Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain Cogn*. 79:138–157.
- Rossion B, Jacques C. 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*. 39:1959–1979.
- Rossion B, Joyce CA, Cottrell GW, Tarr MJ. 2003. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*. 20:1609–1624.
- Rousselet GA, Husk JS, Pernet CR, Gaspar CM, Bennett PJ, Sekuler AB. 2009. Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. *BMC Neurosci*. 10:114.
- Schweinberger SR, Pickering EC, Jentzsch I, Burton AM, Kaufmann JM. 2002. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Res Cogn Brain Res*. 14:398–409.
- Smith SM, Nichols TE. 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. 44: 83–98.
- Suzuki WA, Miller EK, Desimone R. 1997. Object and place memory in the macaque entorhinal cortex. *J Neurophysiol*. 78:1062–1081.
- Tanaka JW, Curran T, Porterfield AL, Collins D. 2006. Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *J Cogn Neurosci*. 18:1488–1497.
- Versky SC, Todorov A, Turk-Browne NB. 2013. Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia*. 51:2100–2108.
- Vida MD, Nestor A, Plaut DC, Behrmann M. 2017. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proc Natl Acad Sci*. 114:388–393.
- Visconti di Oleggio Castello M, Gobbini MI. 2015. Familiar face detection in 180ms. *PLoS One*. 10:e0136548.
- Visconti Di Oleggio Castello M, Halchenko YO, Guntupalli JS, Gors JD, Gobbini MI. 2017. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Sci Rep*. 7:12237.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 13:600–612.
- Weibert K, Andrews TJ. 2015. Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*. 75:588–596.
- Weibert K, Harris RJ, Mitchell A, Byrne H, Young AW, Andrews TJ. 2016. An image-invariant neural response to familiar faces in the human medial temporal lobe. *Cortex*. 84:34–42.
- Wiese H, Tüttgenberg SC, Ingram BT, Chan CXY, Gurbuz Z, Burton AM, Young AW. 2018. A robust neural index of high face familiarity. *Psychol Sci*. in press.
- Yakovlev V, Fusi S, Berman E, Zohary E. 1998. Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. *Nat Neurosci*. 1: 310–317.
- Young AW, Burton AM. 2017. Recognizing faces. *Curr Dir Psychol Sci*. 26:212–217.
- Yovel G. 2016. Neural and cognitive face-selective markers: an integrative review. *Neuropsychologia*. 83:5–13.