Title: Rapid neural representations of personally relevant faces

Mareike Bayer*[1], Oksana Berhe[1,2], Isabel Dziobek[1], Tom Johnstone[3,4]

1 Berlin School of Mind and Brain, Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
2 Zentralinstitut für Seelische Gesundheit, Mannheim, Germany
3 Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, The University of Reading, Reading, UK
4 School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

Corresponding author

Dr. Mareike Bayer
Berlin School of Mind and Brain
Department of Psychology
Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin
Germany
Email: mareike.bayer@hu-berlin.de

Abstract

The faces of those most personally relevant to us are our primary source of social information, making their timely perception a priority. Recent research indicates that gender, age and identity of faces can be decoded from EEG/MEG data within 100ms. Yet the time course and neural circuitry involved in detecting personal relevance of faces remain unknown. We applied representational similarity analyses and simultaneous EEG-fMRI to examine neural responses to emotional faces of participants' romantic partners, friends, and a stranger. EEG-fMRI representations of personal relevance started prior to structural encoding at 100ms in visual cortex, but also in prefrontal and midline regions involved in value representation, and monitoring and recall of self-relevant information. Representations specifically related to romantic love emerged after 400ms. Our results suggest that models of face perception need to be updated to account for rapid detection of personal relevance in cortical circuitry beyond the core face processing network.

Keywords: Personal relevance, emotion, faces, EEG-fMRI, Representation

Faces are arguably the most important social and emotional stimuli we encounter in daily life. They tell us an enormous amount about our fellow humans, including whether they are strangers, friends, enemies, or loved ones, how old they are, and how they are feeling, both generally, as well as towards us specifically. So important are faces to us, that we appear to be experts at processing them, and have developed specialised neural circuitry to do so [1]. Based on a substantial body of behavioural and neuroimaging research, theories of how different types of face information are extracted from the visual stream have been proposed, with separate processing pathways being suggested for transient (e.g. emotion) and stable (e.g. identity) aspects of faces [2]. Yet testing the temporal dynamics and neural circuitry of specific aspects of face perception is hampered by the lack of temporal resolution needed to delineate several rapidly occurring processing stages. Here, we address this question using simultaneous recordings of EEG and fMRI.

Cumulative evidence from EEG and MEG research has identified the N170/M170 component (at 170 ms after stimulus onset) as a correlate of structural face encoding and representations of identity [3,4]. However, more recent research using multivariate pattern analysis and related techniques has demonstrated that various aspects of face information, including identity, emotional expressions, age, and gender, are represented already within the first 100 ms after stimulus onset [5–8]. Yet while these studies provide valuable evidence about the time course of face processing, the brain networks involved in the specific aspects of face perception remain difficult to identify. Furthermore, these studies have almost exclusively focused on faces that carry little or no personal relevance to the observer.

We thus know little about the most important type of real-life face processing, namely detecting and recognising the faces of our families and close friends, a skill in which we

are highly efficient and robust [9]. The processing of personally familiar faces involves additional brain areas outside of the visual processing stream, including regions involved in monitoring of self-relevant information (medial prefrontal cortex), episodic memories (precuneus, anterior temporal cortex), person knowledge (temporoparietal junction; TPJ), and emotion processing (amygdala, insula) [10–13]. However, it is largely unknown how quickly personal relevance can be extracted from faces, and where in visual or higher-level processing streams this occurs. EEG findings suggests that earliest effects of personal relevance of faces might occur around 170 ms after stimulus onset [9], although effects are more robust at higher-order processing stages, modulating the amplitudes of the P3 component [14]. These studies suggest that representation of personal relevance of faces might depend at least on structural encoding in the ventral visual stream, as well as possibly more elaborate cognitive processing. Interestingly, a recent study using MEG reported enhanced early (70-100ms) encoding of gender and identity in familiar faces, with less robust encoding of gender and identity in unfamiliar faces. Encoding of familiarity itself was only detectable 400 ms after stimulus onset [7]. The authors speculated, based on the timeline of their results, that amplification of familiar faces might involve early perceptual tuning that precedes structural encoding and top-down processing. However, their MEG analyses did not provide the spatial specificity to investigate the network underlying the early amplification of familiar faces; and the authors thus suggested that future multimodal studies might shed light on this important question. It is also noteworthy that the study used famous faces (actors), which are of less personal relevance and have been shown to be processed in a qualitatively different way than personally relevant faces [9,15]. Thus, despite considerable progress, it remains unclear at what stage genuine personal relevance of faces, such as those of our friends and loved ones, is decoded, which brain regions are involved, and

whether rapid prioritisation of personally relevant faces is based on tuning of sensory areas [16], or on top-down influence on these areas [17].

In the present study, we aimed to answer these questions through simultaneous recordings of EEG and fMRI. We presented heterosexual females in a romantic relationship with pictures of their romantic partner, a close male friend, and a male stranger, displaying fearful, happy and neutral facial expressions. In addition to standard unimodal analyses of EEG and fMRI, we combined these data using representational similarity analyses (RSA; Fig 1) [18]. Our results reveal two major findings: First, we show that personal relevance leads to increased activation in the face processing network, and to fast attention allocation in ERPs. Second, we show that shared representations between EEG and fMRI spatial activation start as early as 100 ms after stimulus onset not only in the visual cortex, but also in cortical areas previously associated with value representation and the monitoring of self-relevant information.

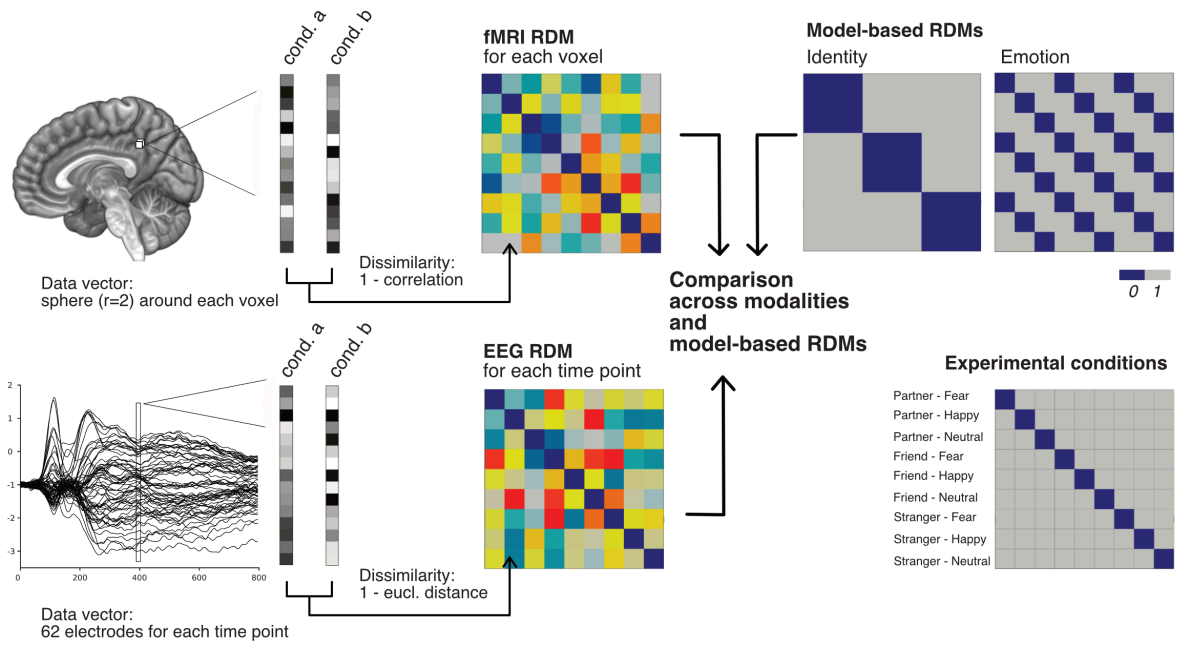

Fig. 1. Representational Similarity Analyses**.** Representational dissimilarity matrices (RDMs) are constructed for each voxel in the brain and for each data point in grand

averaged ERPs by comparing pairwise, condition-specific activations. RDMs are symmetric with a diagonal of zeros, and their size corresponds to the number of experimental conditions, here 9 x 9. Model-based RDMs reflect theoretical predictions. After construction, RDMs are compared across modalities and with model-based RDMs.

Results

Unimodal fMRI analyses

Increased activation for Partner > Stranger and Friend > Stranger was observed in right fusiform cortex, precuneus, posterior cingulate cortex, anterior cingulate cortex (ACC), right middle temporal gyrus and posterior superior temporal sulcus (Fig. S1). Additionally, the contrast Partner > Stranger showed activation in the bilateral orbitofrontal cortex, while Friend > Stranger yielded activation in the ventromedial prefrontal cortex. Emotion contrasts yielded activation for Happy > Neutral in the parieto-occipital cortex (including the intracalcarine cortex, lingual gyrus and precuneus), cerebellum and brain stem. Further clusters were located in the medial and ventromedial prefrontal cortex, orbitofrontal cortex, ACC, inferior and superior frontal gyrus, and insular cortex. Subcortically, activation was seen in the left amygdala, bilateral thalamus and left caudate. For Happy > Fear, activation was found in bilateral insula and frontal operculum, and left-lateralized inferior and middle frontal gyrus and lateral occipital cortex. See Table S1 for a complete list of activated brain regions. Region of interest analyses showed an effect of Identity on bilateral VTA activations, Fs(2,42) > 4.60, ps < .016, η_p^2s > .107, based on increased activations for Partner > Stranger, p = .01 (L VTA). In the left VTA, an interaction between Identity and Emotion reflected that activation in response to the Friend's face was modulated by the emotional expression, with increased activation for happy expressions compared to

fearful and neutral expressions, *F*s(1,21) > 7.21, *p*s < .042. In bilateral caudate activity, a main effect of Emotion, *F*s(2,42) > 2.86, *p*s = .025, η_p^2s > .161, was based on numerically higher activations in response to happy faces, but post-tests were not significant. ROI analyses of Amygdala, putamen and N. accumbens did not reveal significant results. Full statistical analyses are reported in S2.

Unimodal EEG analyses

Analyses of P1 amplitudes showed a main effect of Identity, $F(2,68) = 4.1$, $p = .025$, $\eta_p^2 = .194$, based on higher amplitudes for Partner as compared to Stranger, $p = .017$ (Fig. 3). Likewise, N170 amplitudes were modulated by the factor Identity, $F(2,34) = 9.17$, $p = .001$, $\eta_p^2 = .350$, with higher amplitudes for Partner and Friend compared to Stranger, *p*s < .02. Both P3 and LPC showed main effects of Identity, *F*s(2,34) > 13.86, *p*s < .001, η_p^2s > .449; post-tests showed higher amplitudes for Partner compared to both Friend and Stranger, *p*s < .014 (Fig. 2). See Table S3 for detailed follow-up analyses.

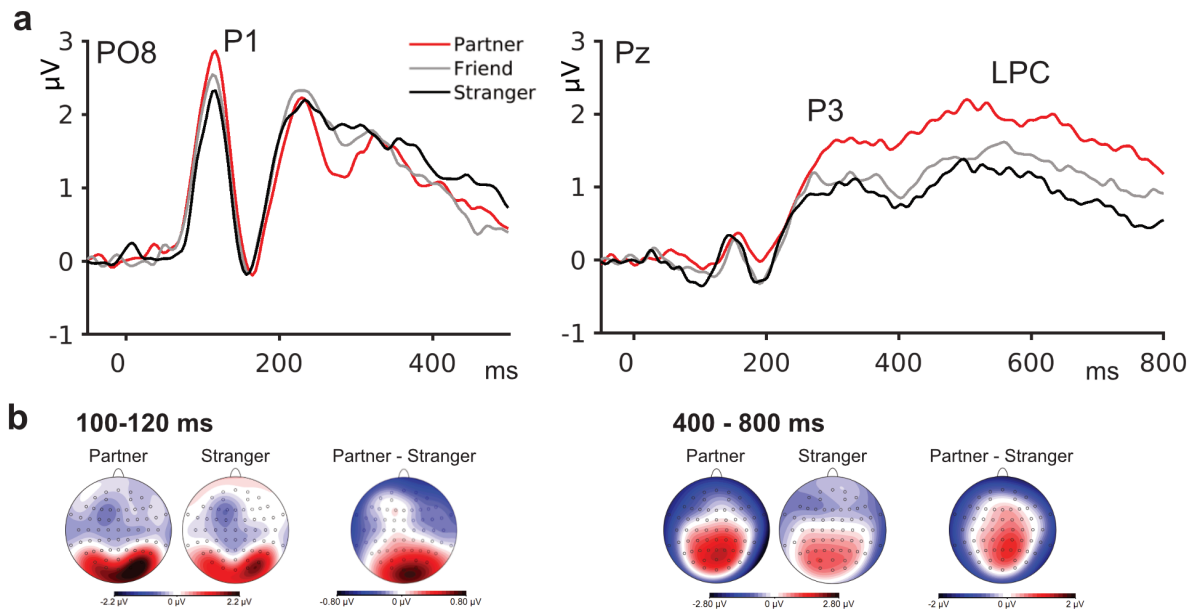

Fig. 2. Effects of personal relevance in event-related potentials. **a** Grand mean ERP waveforms for the factor Identity at electrodes PO8 and Pz, showing increased amplitudes for Partner compared to Friend/Stranger in P1 and P3/LPC amplitudes. **b**

Scalp distributions for Partner, Stranger, and their difference topography at indicated time intervals.

Representational similarity analyses: Model-based RDMs

Face Identity

Using a model-RDM for face identity as a searchlight in voxelwise fMRI analyses revealed widespread representations in the core and extended face processing network (Fig. 3). Analyses using a model-based RDM for familiarity (Partner and Friend vs. Stranger) showed a high level of consistency with representations of Identity, suggesting that face familiarity accounts for the majority of observed effects. In contrast, analysis with an RDM reflecting romantic Love (Partner vs. Friend and Stranger) revealed more focal activations, which were located in bilateral insula, opercular cortex, lateral occipital cortex and superior temporal gyrus, and in subcortical regions in putamen, caudate, amygdala, thalamus, cerebellum, and regions of the brain stem including corticospinal and corticobulbar tracts and substantia nigra. See Table S1 for a complete list of brain regions.

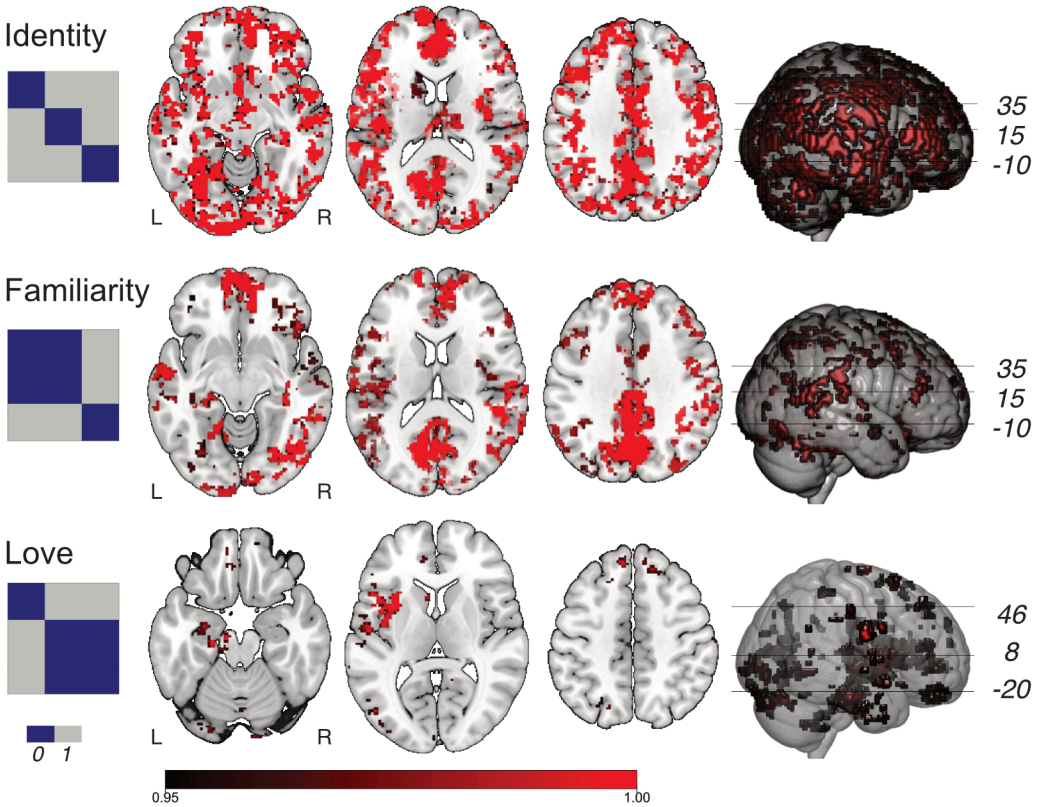

Fig. 3: Representations of conceptual RDMs for Identity, Familiarity and Love**.** Activation thresholded with FWE < 0.05 using permutations and TFCE.

Emotional expression

Comparisons with the model-based Emotion RDM revealed representations that were less widespread than effects of Identity, but included bilateral amygdala, right putamen, the orbitofrontal and ventromedial prefrontal cortex, the temporoparietal junction (TPJ) and left inferior frontal gyrus. A model-based RDM for valence (distance = 1 from neutral to both fear and happy, distance = 2 between fear and happy) showed activation in most areas of the core and extended face network, including bilateral amygdala, insula, hippocampus, ventromedial and orbitofrontal PFC (Fig. 4). The analyses of model-based emotional arousal (assuming increased values for fearful and happy compared to neutral faces) revealed no significant representations.

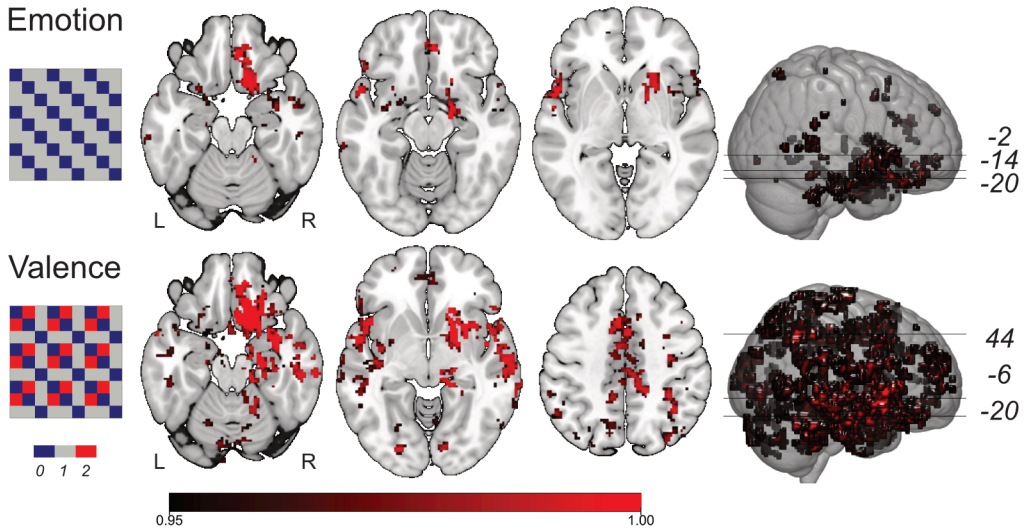

Fig. 4: Representations of conceptual model RDMs for Emotion and Valence. Activation thresholded with FWE < 0.05 using permutations and TFCE.

Combined EEG-fMRI RSA analyses

Peaks in EEG RDM structure were evident at 52 ms, 108 ms, 204 ms, 308 ms, 428 ms, and 660 ms. Accordingly, these RDMs were used as searchlight RDMs in the whole-brain analyses on single-subject fMRI data (Fig. 5).

For the EEG RDM at 52 ms after stimulus onset there were no brain regions with significantly correlated spatial RDMs. EEG representations at 108 ms showed correspondence to BOLD representations in the extended face network, including the precuneus, posterior cingulate cortex and TPJ, but also in the ventromedial PFC and in the visual cortex (intracalcarine cortex and lingual gyrus). The EEG RDM at this time point mainly represented effects of familiarity (Stranger vs. Partner and Friend), as apparent from correlations of the EEG RDM with model-based RDMs (see Table 1). The same was true for the EEG RDM at 204 ms, with more widespread corresponding EEG-fMRI representations that additionally included the ACC, fusiform gyrus, amygdala, insula and N. accumbens. At 308 ms and 428 ms, EEG RDMs showed highest

correspondence to model-based RDMs of Love (that is, Partner vs. Friend and Stranger). In EEG-fMRI, corresponding representations at 308 ms were more focal and most evident in the precuneus, posterior cingulate and superior frontal gyrus. Finally, EEG-fMRI representations at 428 ms and 660 ms after stimulus onset were again more extended, with additional significant representations in the middle temporal gyrus, TPJ, in the inferior and superior frontal gyri, as well as in regions of the midbrain. See Table 1 for correlations between EEG RDMs and model-based RDMs, and Table S1 for a complete list of brain regions.

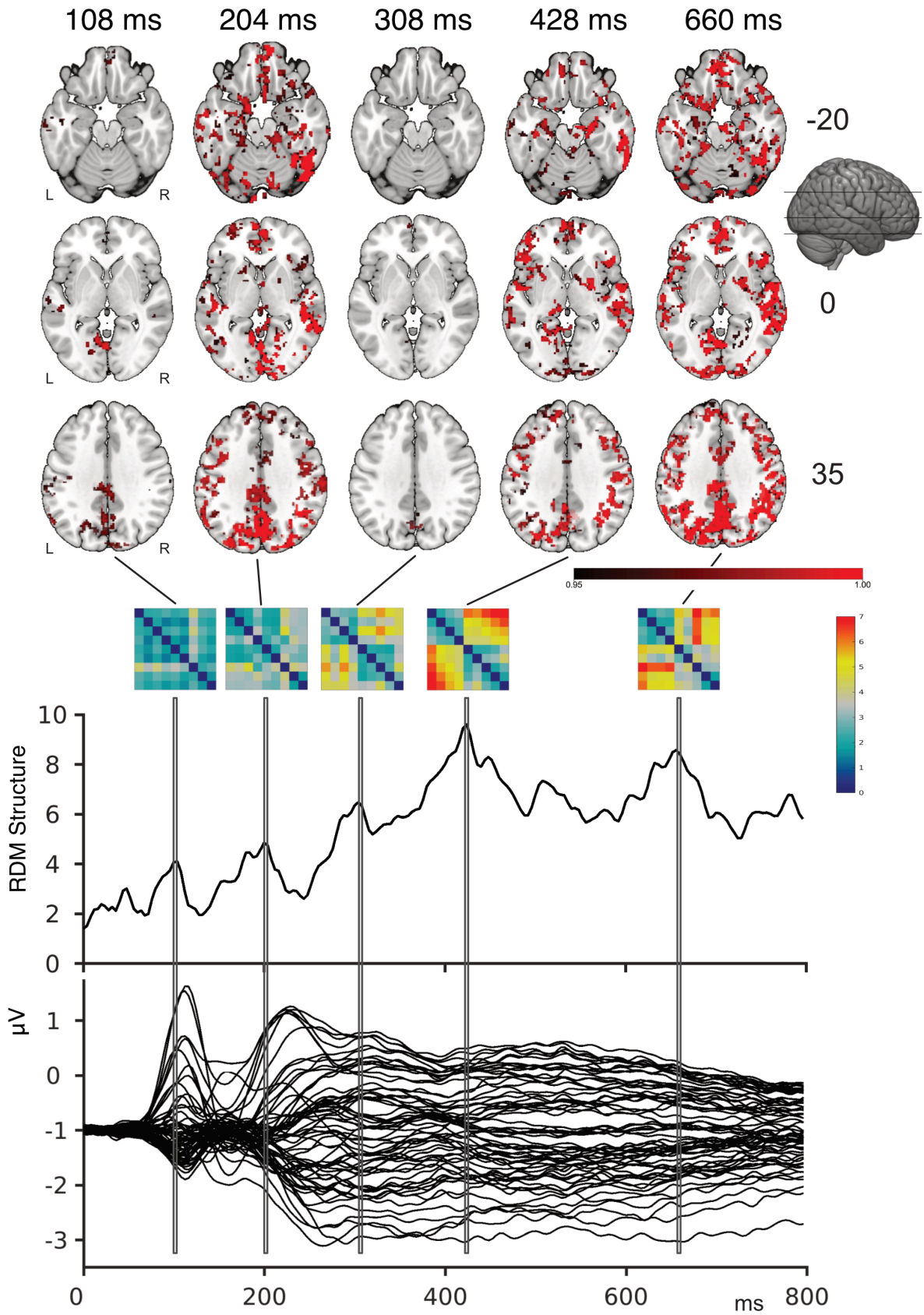

Fig. 5: Combined EEG-fMRI representations. Brain regions in which fMRI RDMs corresponded to representations in the EEG at the corresponding time points. RDMs

show EEG representations used as target RDMs in the combined EEG-fMRI analyses, selected on the basis of peaks in the RDM structure (top graph). Each line in the lower graph represents the mean time series for one electrode.

Table 1. Pearson's correlation values of model-based RDMs and EEG RDMs used in fMRI searchlight analyses

	108 ms	**204 ms**	**308 ms**	**428 ms**	**660 ms**
Identity	0.41	0.35	0.32	0.48	0.52
Familiarity	0.70	0.62	0.20	0.33	0.60
Love	0.13	0.07	0.51	0.61	0.41
Emotion	-0.14	0.00	-0.11	-0.16	-0.17

Discussion

The faces of those closest to us are arguably the most personally relevant social stimuli we commonly encounter and thus might receive a high degree of processing priority. In this study, we investigated the time course of neural processing of the faces of romantic partners and friends using simultaneous EEG-fMRI and representational similarity analyses. Our results point to the strong and rapid impact of personal relevance on face processing, with increased activation in the core and extended face processing network, as well as fast attention allocation in ERPs, and shared representations between EEG and fMRI spatial activation patterns in multiple cortical regions starting as early as 100 ms after stimulus onset.

Model-based representational similarity analyses revealed neural representations of identity in large parts of the core face processing network, including fusiform gyrus, posterior STS and inferior frontal gyrus, and in precuneus, posterior and anterior

cingulate cortex, and medial prefrontal cortex as parts of the extended face network, corroborating previous research [9,10,13]. Personal relevance was examined both in terms of personal familiarity (Partner/Friend vs. Stranger), and romantic love (Partner vs. Friend/Stranger). Personal familiarity accounted for the majority of effects of personal relevance across the face processing network. In contrast, representations corresponding to the romantic love RDM were more focal and mostly outside the face network, including subcortical regions in putamen, caudate, amygdala, thalamus, cerebellum, and regions of the brain stem including corticospinal and corticobulbar tracts and substantia nigra, but also cortical regions including the insular cortex, medial and ventromedial prefrontal cortex, inferior frontal gyrus and in the intracalcarine cortex. These results are consistent with previous findings that the processing of a loved one's face engages areas of the brain's dopaminergic reward circuitry in the dorsal striatum (putamen and caudate) [19] and substantia nigra. Furthermore, our region-of-interest analyses revealed increased BOLD activation for Partner compared to Stranger in the ventral tegmental area, which is the major source of dopaminergic neurons in the mesolimbic dopamine system. In the left VTA, analyses also showed an interaction of personal relevance and emotional expressions: While activity was generally increased in response to the partner's face compared to a stranger, activation to the Friend's face was increased only for happy expressions, showing a coding of both identity and emotional expression that reflects the degree of personal reward value.

Event-related potentials revealed that personal relevance increased activation throughout the processing time course. Increased P1 amplitudes are especially noteworthy, since the P1 was related to perceptual processing at 100 ms after stimulus onset, and thus precedes structural face encoding as indexed by the N170 component [3],

as well as subsequent face identification processes [2]. Similar activations have been reported for stimuli with emotional and motivational relevance [20], and have been related to associative learning of physical stimulus properties [21,22]. Thus, our results are not necessarily in conflict with models of face recognition based on structural encoding and associative memory [2], but might reflect reward value associated with (the physical features of) a loved one's face, extracted prior to full structural encoding.

While ERP analyses suggest early modulation of visual perceptual processes by personal relevance, the combination of EEG and fMRI data using RSA allows for the time-resolved analyses of *representations* within the fMRI. These analyses suggest fast modulation of neural processing across a far more widespread collection of cortical regions. Shared EEG-fMRI representations were first apparent as early as 100 ms after stimulus onset not only in the visual cortex, but also in the ventromedial and medial prefrontal cortex, regions involved in the monitoring of self-relevant information and value encoding [23,24]. Further early representations were observed in regions linked to multimodal sensory integration and theory of mind (TPJ) [25], and episodic and autobiographical memory (posterior cingulate cortex) [26]. The fast and widespread activation of brain areas involved in social cognition and reward encoding serves to highlight the prioritization of genuinely personally relevant information: In contrast to our findings, a previous MEG study using famous faces instead of personally relevant faces reported decoding of face familiarity only after 400 ms [7]. However, this study reported increased early representations of gender and identity for familiar compared to unfamiliar faces with latencies of 60 to 100 ms. Based on the timeline of their results, the authors speculated that these activations might be based on selective bottom-up amplification of sensory representations. In contrast, although starting somewhat later, our results show that at

100 ms after stimulus onset, corresponding fMRI representations were evident outside of visual areas, including the PFC.

Taken together, our unimodal EEG and EEG-fMRI RSA analyses suggest that personal relevance of faces can be extracted prior to full structural face encoding, involving both visual and frontal/midline brain regions. Based on this circuitry, we suggest that these effects might rely on associative learning of visual stimulus features and the reward value of personally relevant faces. This explanation is consistent with reports of fast activations of the PFC in affective associative learning of faces at latencies starting as early as 50 - 80 ms after stimulus onset [22]. Therefore, future models of face perception should consider the influence of face-unspecific mechanisms on face perception, like reward-related processes, which seem to play an important role in processing real-life faces.

At 200 ms, shortly after the stage of structural face encoding, representations additionally included the fusiform gyrus, but also amygdala, insular cortex and N. accumbens. This time course of representations of face relevance suggests that the response of subcortical relevance- and reward related structures like the amygdala and the N. accumbens might rely on the output of structural face encoding at around 170 ms after stimulus onset.

Finally, at the stage of higher-order processing from approximately 400 ms after stimulus onset, representations were identified in all regions of the core and extended face processing network, including amygdala, insular and orbitofrontal cortex [10], but also in regions identified with the theoretical RDM for romantic love, like putamen, cerebellum and regions of the brain stem. Consistent with EEG RDMs, ERPs showed evidence for differential processing of romantic partners at a similar time scale. Across modalities, our results suggest that early processing mainly reflects the amplification of

familiarity; whereas effects specific for romantic love only emerge at a higher-order processing stage. However, it is important to keep in mind that the terms 'familiarity' and 'love' in this case are not exclusively related to Friend and Partner, respectively, but reflect different aspects of close relationships [27]: Both Partner and Friend are of high personal relevance; and one can feel both friendship love towards a close friend as well as friendship towards one's partner. Thus, romantic love in our study refers to representations that are unique to a romantic relationship, over and above a close friendship.

In this study, effects related to personal relevance clearly dominated, while emotional expressions were associated with relatively weak effects. This should not come as a surprise. Being able to rapidly identify and respond to those closest to us is a fundamental ability present from early infancy [28]. It also suggests that emotional expressions are not processed in a rigid, automatic way, but that brain responses rather reflect context-specific relevance of social stimuli. Understanding the time course and neural circuitry of processing individually relevant faces might be of special importance in clinical conditions, where processing of emotional or social information is specifically disrupted. Autism spectrum conditions (ASC) are an example in which atypical face processing might reflect altered personal relevance attached to strangers, rather than a dysfunction of the neural face processing architecture [29], with potential consequences for the design of targeted interventions.

What is clear from this study is that in basic social neuroscience, as well as in clinical research, there are compelling reasons to study responses to individualised social stimuli. Such studies might yield valuable insights into how the brain engages circuitry beyond that typically considered in models of face perception, in order to prioritise

processing of genuinely personally relevant faces. After all, by far the most pervasive social stimuli we encounter in our daily lives, the ones our brains are most attuned to, are our close friends and loved ones.

Methods

The study was reviewed and approved by the University of Reading research ethics committee. All participants provided informed consent before taking part in the study.

Participants

Data were collected from 22 female participants (mean age = 19.8 years, sd = 0.9). For four participants, no EEG data was available due to technical problems and insufficient data quality. All participants were in a heterosexual romantic relationship at the time of data collection (mean duration = 20.0 months, sd = 14.6; friendship: mean duration = 35.9 months, sd = 26.5). Participants received a mean score of 105.9 points (sd = 10.7) of 135 points on the Passionate Love Scale [30]. They reported high contentment both with their relationship (mean = 8.8/10, sd = 1.3) and their friendship (mean = 8.3, sd = 1.3). All participants had normal or corrected-to-normal vision; 20 participants were right-handed. Participants were recruited through the Undergraduate Research Panel and Internet ads; they received course credit or £25 for participation.

Stimuli

Stimuli consisted of portraits of the participant's boyfriend, a male close friend, and a male stranger, displaying fearful, happy, and neutral facial expressions (3 x 3 design). All stimuli were obtained by taking screen shots during a Skype session prior to the experimental session; all participants were presented with the same stranger.

After the main experiment, participants completed ratings of stimulus valence and arousal, as well as on attractiveness (neutral expression) using 7-point Likert scales. Rating values are reported in table 1; full statistical analyses are reported in Table S4.

Table 2. Rating values (means and standard deviations) of experimental stimuli

	Attractiveness *(1-7)*	**Valence** *(-3 -3)*			**Arousal** *(1-7)*		
		Fearful	**Happy**	**Neutral**	**Fearful**	**Happy**	**Neutral**
Partner	6.1 (1.1)	-0.4 (1.5)	2.3 (0.7)	0.0 (1.8)	3.2 (1.7)	5.4 (1.5)	3.7 (1.8)
Friend	4.1 (1.4)	-1.1 (1.1)	1.6 (0.8)	-0.6 (1.1)	2.1 (1.6)	2.9 (1.9)	2.5 (1.6)
Stranger	4.1 (1.2)	-2.0 (1.1)	0.5 (1.1)	-0.6 (1.2)	1.6 (1.1)	3.2 (2.0)	2.5 (1.6)

Procedure

After receiving information about the study, participants provided informed consent and were fitted with the EEG cap. Inside the scanner, participants performed a passive face-viewing task. Face stimuli were presented for 1s. Every face stimulus was presented 40 times, resulting in 360 experimental trials. Additionally, 40 1-back trials were presented to ensure participant's attention to the faces. In these trials, participants had to indicate whether a face presented after a question mark was identical in identity and emotion to the one presented before the question mark. Stimuli were presented in 10 blocks. The stimulus sequence and the jittering of the inter-trial interval was determined using fMRI simulator [31] (mean ITI = 3000 ms, min = 2500 ms, exponential function). A central fixation cross was presented during the ITI. Stimuli were presented on a Nordic Neuro Labs goggle system at 60 Hz on an 800 x 600 pixel screen using EPrime software (Psychology Software Tools, Inc.).

Data acquisition and pre-processing

fMRI

Data were collected on a 3T Siemens Trio MRI scanner using the standard 12 channel head coil. Functional images were acquired with a T2*-weighted gradient echo EPI sequence (40 interleaved axial slices, phase encoding P to A, 2.5 x 2.5 mm voxels, slice thickness = 2.5 mm, interslice gap = 0.25mm, 100x100 in-plane matrix, TR = 2500 ms, TE = 30 ms, Flip Angle: 90°). A high-resolution T1-weighted whole-brain structural image was acquired using an MPRAGE sequence (1mm isotropic voxels, FOV = 160 x 256 x 256 mm, Flip Angle: 9°).

Data were processed with FSL 5.0 (www.fmrib.ox.ac.uk/fsl). Brain extraction was performed using the BET [32]. Data were denoised (Marchenko-Pastur Principal Component Analyses), motion corrected [33], spatially smoothed using a 5 mm FWHM Gaussian kernel [34], and grand-mean intensity normalized. Motion artefacts were removed with ICA-AROMA [35] and data were high-pass filtered (100 s). Single subject 4D data were registered to the subject's structural image using BBR [36]. Registration of the individual structural image to the 2mm MNI 152 template was performed using ANTS [37]. Finally, transformations were combined for the registration of the functional data to standard space.

EEG

Continuous EEG data were collected from 64 Ag-AgCl electrodes simultaneously to fMRI acquisition (BrainProducts system); the electrocardiogram (ECG) was recorded from an electrode placed left of the spinal column. The sampling rate was 5000 Hz; data were referenced online to electrode FCz with electrode AFz as ground. Electrode impedances were kept below 20 kΩ. A sync box (BrainProducts) synchronised the MRI and EEG

computer clocks. Offline, MR gradient artefacts were removed from continuous, baseline corrected data (using the whole artefact for baseline correction) with a sliding window of 21 artefacts [38]. Data were low-pass filtered using a FIR filter (70 Hz) and down-sampled to 250 Hz. Ballistocardiographic artefacts were identified using the ECG with a semiautomatic template matching procedure and corrected using a template subtraction approach (sliding window of 15 pulse intervals). A restricted Infomax ICA [39] was used to remove eye blinks, eye movements and residual ballistocardiographic artefacts. Data was re-referenced to average reference and segmented into epochs from -100 ms to 800 ms relative to stimulus onset, and baseline-corrected using a 100 ms pre-stimulus baseline. Trials with activity exceeding ± 100 μV or voltage steps larger than 100 μV were excluded from analyses (0.6 % of trials). Data were averaged per participant and experimental condition.

Data analyses

All voxelwise fMRI statistical tests were one-sided; all other tests were two-sided.

fMRI

A first level GLM was applied with regressors for emotion by identity conditions (9 regressors) and for 1-back trials, created by convolving the temporal profile of each experimental condition with the double gamma haemodynamic response function. Nuisance regressors without convolution were included to model breaks between blocks and artefacts (framewise displacement > 1mm). Contrasts of interest in unimodal fMRI analyses included effects of Identity and Emotion. Region-of-interest (ROI) analyses were performed for bilateral activations of amygdala, Nucleus accumbens, putamen and caudate (masks created from Harvard-Oxford subcortical structural atlas)

and ventral tegmental area (VTA; ; mask retrieved from neurovault.org/collections/1380) [40]. Masks were thresholded at 25% and binarised. Within ROIs, subject-wise covariation parameter estimates for all experimental conditions were compared with repeated-measures ANOVAs with the factors Identity and Emotion (3 x 3). Huynh-Feldt correction was applied for violations of sphericity; post-tests were Bonferroni corrected. Whole-brain analyses were thresholded with permutation tests (using an MNI gray matter mask with tissue priors thresholded at 0.3 and variance smoothing of 2mm) and threshold-free cluster enhancement to ensure a corrected Familywise Error (FWE) of < .05.

EEG

EEG analyses were performed on ERP components P1, N170 and P3 and LPC. P1 amplitudes were quantified at the average of occipital electrodes PO8, PO4, POz, PO3, PO7, O1, Oz, and O2 using semi-automatic peak detection in the time window from 90 to 130 ms after stimulus onset (mean peak latency = 107ms). N170 peak amplitudes were detected on averaged electrodes TP9, TP7, TP8, TP10, P7 and P8 from 150 to 220 ms after stimulus onset (mean peak latency = 172ms). In order to account for differences in the preceding P1 component, N170 amplitudes were subtracted from P1 amplitudes. P3 and LPC amplitudes were analysed at electrodes CP1, CPz, CP2, P1, Pz, P2 and POz in the time windows of 300 to 400ms (P3) and 400 to 800 ms (LPC). Analyses were performed with repeated-measures ANOVAs including the factors Identity (3), Emotion (3) and, for P3 and LPC, electrode (7). Huynh-Feldt correction was applied for violations of sphericity; post-tests were Bonferroni corrected for multiple comparisons.

RSA analyses

RSA analyses were performed using the CoSMoMVPA toolbox [41] and the toolbox for representational similarity analyses [42] on Matlab R2017b.

fMRI

Representational dissimilarity matrices (RDMs) were constructed for each voxel in the brain, separately for each subject, based on z-values of each of the 9 experimental conditions included in the mixed-effects GLM. For each voxel and condition, we created a vector (based on a sphere around each voxel, radius = 2 voxels). We quantified the dissimilarity between each pair of experimental conditions as 1 – Pearson's R of the two corresponding vectors. This resulted in a 9 x 9 RDM for each voxel and each participant.

EEG

RDM matrices were constructed using grand-averaged waveforms. For each time point from stimulus onset to 800 ms after stimulus onset, the distance between pairs of experimental conditions was quantified as their Euclidean distance across all 62 scalp electrodes. Euclidean distance was used in order to account for amplitude differences, which convey essential information in event-related potentials. Analyses resulted in a 9 x 9 RDM for each time point (Fig. S2).

In order to select EEG time points as target RDMs in the fMRI analyses, we derived a measure of internal structure of each time point's RDM by computing the Euclidean distance of RDM cell values to the arithmetic mean of cell values. As a result, time points with pronounced differences of dissimilarity values (RDM cell entries) between condition pairs receive high values, whereas smaller differences result in low values. For these calculations, we used one half of the (symmetrical) RDM, excluding the diagonal zeros.

Conceptual model RDMs

Conceptual model RDMs for Emotion and Identity were created based on the assumption of high similarity within experimental categories and low similarity across categories (coded as 0 and 1, respectively). We also computed conceptual model RDMs for face familiarity (Partner/Friend vs. Stranger) and romantic love (Partner vs. Friend/Stranger; see Figure 4, and for emotional valence (distance between happy/fearful vs. neutral = 1, distance happy vs. fearful = 2) and arousal (distance happy/fearful vs. neutral = 1, distance happy vs. fearful = 0).

Joint EEG-fMRI

For combination of EEG and fMRI data, we performed representational similarity analyses using EEG RDMs as a searchlight with each participant's individual fMRI RDMs for each voxel using Pearson's R. Analyses were performed only on subjects where both fMRI and EEG data was available (n= 18). Resulting brain maps of similarity were combined across subjects using permutation tests and threshold-free cluster enhancement in order to ensure FWE of < .05.

Acknowledgements

We thank Michael Lindner and Catriona Scrivener for assistance with data collection and Luca Brivio for assistance with data analysis. M.B. and T.J. were supported by the Berlin School of Mind and Brain (Humboldt-Universität zu Berlin).

Author contributions

M.B., T.J. and I.D. designed the experiment, M.B. and T.J. collected data, conducted analyses and drafted the manuscript, M.B. and O.B. analysed EEG data. All authors provided feedback and revised the manuscript.

The authors declare no conflict of interest.

Code and data availability

We will make code, data and brain maps available upon request and will make them publicly available following publication.

References

1. Kanwisher, N. Domain specificity in face perception. *Nature Neuroscience* (2000). doi:10.1038/77664
2. Bruce, V. & Young, A. Understanding face recognition. *Br. J. Psychol.* (1986). doi:10.1111/j.2044-8295.1986.tb02199.x
3. Eimer, M. The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport* **11,** 2319–24 (2000).
4. Rossion, B. & Jacques, C. in *The Oxford Handbook of Event-Related Potential Components* (2012). doi:10.1093/oxfordhb/9780195374148.013.0064
5. Vida, M. D., Nestor, A., Plaut, D. C. & Behrmann, M. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proc. Natl. Acad. Sci.* (2017). doi:10.1073/pnas.1614763114
6. Nemrodov, D., Niemeier, M., Mok, J. N. Y. & Nestor, A. The time course of individual face recognition: A pattern analysis of ERP signals. *Neuroimage* (2016).

doi:10.1016/j.neuroimage.2016.03.006

7. Dobs, K., Isik, L., Pantazis, D. & Kanwisher, N. How face perception unfolds over time. *Nat. Commun.* (2019). doi:10.1038/s41467-019-09239-1

8. Dima, D. C., Perry, G., Messaritaki, E., Zhang, J. & Singh, K. D. Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces. *Hum. Brain Mapp.* (2018). doi:10.1002/hbm.24226

9. Ramon, M. & Gobbini, M. I. Familiarity matters: A review on prioritized processing of personally familiar faces. *Vis. cogn.* (2018). doi:10.1080/13506285.2017.1405134

10. Gobbini, M. I. & Haxby, J. V. Neural systems for recognition of familiar faces. **45,** 32–41 (2007).

11. Natu, V. & O'Toole, A. J. The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology* (2011). doi:10.1111/j.2044-8295.2011.02053.x

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

13. Visconti Di Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D. & Gobbini, M. I. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Sci. Rep.* (2017). doi:10.1038/s41598-017-12559-1

14. Guerra, P. *et al.* Affective processing of loved familiar faces : Integrating central and peripheral electrophysiological measures. *Int. J. Psychophysiol.* **85,** 79–87 (2012).

15. Sugiura, M. Neuroimaging studies on recognition of personally familiar people. *Front. Biosci.* (2014). doi:10.2741/4235

16. Visconti di Oleggio Castello, M. & Gobbini, M. I. Familiar Face Detection in 180 ms. *PLoS One* **10,** e0136548 (2015).

17. Gobbini, M. I. & Haxby, J. V. Neural response to the visual familiarity of faces. *Brain Res. Bull.* (2006). doi:10.1016/j.brainresbull.2006.08.003

18. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* (2008). doi:10.3389/neuro.06.004.2008

19. Cacioppo, S., Bianchi-Demicheli, F., Hatfield, E. & Rapson, R. L. Social neuroscience of love. *Clin. Neuropsychiatry* **9,** 3–13 (2012).

20. Pourtois, G., Schettino, A. & Vuilleumier, P. Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biol. Psychol.* **92,** 492–512 (2013).

21. Bayer, M., Grass, A. & Schacht, A. Associated valence impacts early visual processing of letter strings: Evidence from ERPs in a cross-modal learning paradigm. *Cognitive, Affective and Behavioral Neuroscience* (2018). doi:10.3758/s13415-018-00647-2

22. Steinberg, C., Bröckelmann, A. K., Rehbein, M., Dobel, C. & Junghöfer, M. Rapid and highly resolving associative affective learning: Convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological Psychology* (2013). doi:10.1016/j.biopsycho.2012.02.009

23. Berridge, K. C. & Kringelbach, M. L. Pleasure Systems in the Brain. *Neuron* (2015). doi:10.1016/j.neuron.2015.02.018

24. Northoff, G. *et al.* Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *Neuroimage* **31,** 440–457 (2006).

25. Patel, G. H., Sestieri, C. & Corbetta, M. The evolution of the temporoparietal

junction and posterior superior temporal sulcus. *Cortex* (2019). doi:10.1016/J.CORTEX.2019.01.026

26. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* **137,** 12–32 (2014).
27. Hendrick, C. & Hendrick, S. A Theory and Method of Love. *J. Pers. Soc. Psychol.* (1986). doi:10.1037/0022-3514.50.2.392
28. Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. & Fabre-Grenet, M. Mother's face recognition by neonates: A replication and an extension. *Infant Behav. Dev.* (1995). doi:10.1016/0163-6383(95)90009-8
29. Pierce, K. The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. *Brain* **127,** 2703–2716 (2004).
30. Hatfield, E. & Sprecher, S. Measuring passionate love in intimate relationships. *J. Adolesc.* **9,** 383–410 (1986).
31. Rorden, C. fMRI Simulator. (2011). at <https://www.mccauslandcenter.sc.edu/crnl/tools/fmristim>
32. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17,** 143–155 (2002).
33. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17,** 825–41 (2002).
34. Smith, S. M. & Brady, J. M. SUSAN—A New Approach to Low Level Image Processing. *Int. J. Comput. Vis.* (1997). doi:10.1023/A:1007963824710
35. Pruim, R. H. R. *et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* (2015). doi:10.1016/j.neuroimage.2015.02.064

36. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* (2001). doi:10.1016/S1361-8415(01)00036-6

37. Avants, B. B., Tustison, N. & Song, G. Advanced Normalization Tools (ANTS). *Insight J.* (2009). doi:http://hdl.handle.net/10380/3113

38. Allen, P. J., Josephs, O. & Turner, R. A Method for Removing Imaging Artifact from Continuous EEG Recorded during Functional MRI. *Neuroimage* **12,** 230–239 (2000).

39. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7,** 1129–59 (1995).

40. Smittenaar, P., Kurth-Nelson, Z., Mohammadi, S., Weiskopf, N. & Dolan, R. J. Local striatal reward signals can be predicted from corticostriatal connectivity. *Neuroimage* (2017). doi:10.1016/j.neuroimage.2017.07.042

41. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front. Neuroinform.* (2016). doi:10.3389/fninf.2016.00027

42. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* (2014). doi:10.1371/journal.pcbi.1003553