How face perception unfolds over time

Katharina Dobs [1,2,3*], Leyla Isik [1,2,3], Dimitrios Pantazis [1,2], Nancy Kanwisher [1,2,3]

[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA

[2] McGovern Institute of Brain Research, Massachusetts Institute of Technology, Cambridge, MA

[3] The Center for Brains, Minds, and Machines, Massachusetts Institute of Technology, Cambridge, MA

* Correspondence: katharina.dobs@gmail.com

Abstract

Within a fraction of a second of viewing a face, we have already determined its gender, age and identity. A full understanding of this remarkable feat will require a characterization of the computational steps it entails, along with the representations extracted at each. Here we used magnetencephalography to ask which properties of a face are extracted when, and how early in processing these computations are affected by face familiarity. Subjects viewed images of familiar and unfamiliar faces varying orthogonally in gender and age. Using representational similarity analysis, we found that gender and age information emerged significantly earlier than identity information, followed by a late signature of familiarity. Importantly, gender and identity representations were enhanced for familiar faces early during processing. These findings start to reveal the sequence of processing steps entailed in face perception in humans, and suggest that early stages of face processing are tuned to familiar face features.

Introduction

A brief glimpse at a face quickly reveals rich multidimensional information about the person in front of us. How is this impressive computational feat accomplished? A key property of a complex computation is that it proceeds via stages and hence unfolds over time. Thus, knowing which information is extracted when is of the essence for understanding the computations underlying face perception. Surprisingly, it remains unknown 1) when after stimulus onset information about different face dimensions, such as gender, age, or identity, are extracted, and

2) how early these computations are affected by the familiarity of the face. Here we use magnetencephalography (MEG) to answer these questions.

While extensive prior evidence indicates that humans detect and recognize faces very rapidly [1-3], much less is known about the precise temporal dynamics of extraction of information about different dimensions of face information. One possibility is that different dimensions of face information are extracted at different stages of processing. For example, gender information might be extracted before identity information, following a coarse-to-fine trajectory [1,4]. Alternatively, different face dimensions could be processed at the same time, suggesting greater interdependence of their processing [5,6]. Resolving the time course by which information about gender, age, and identity emerge will importantly constrain computational models of face perception.

Our second question is whether these face dimensions are processed differently for familiar versus unfamiliar faces, and if so how early familiarity affects processing. A striking yet unexplained finding about face perception is that familiar faces are processed more robustly and efficiently than unfamiliar faces [7]. But the neural mechanisms underlying this effect remain unknown. According to one hypothesis, visual experience with specific faces tunes the bottom-up processing filters for face features, thereby enhancing representations of familiar faces [8]. This hypothesis predicts that familiarity should enhance face representations early in processing, at the same time when those representations are first being extracted. Alternatively (or in addition [9]), familiarity effects in face processing could arise via activation of associated person knowledge and memories, which would then enhance perceptual representations in a top-down manner [10,11]. This hypothesis predicts that familiarity should enhance face representations at some point after those representations are first extracted. Determining which (or both) of these accounts is correct will provide an important step towards understanding the neural mechanisms underlying the behavioral familiarity enhancement effect, and will further inform more general and long-standing questions of how specific prior experience affects the processing of objects [12-14].

To determine how face processing unfolds over time, we applied multivariate analysis methods to MEG data from subjects viewing images of familiar and unfamiliar celebrities who varied orthogonally in gender and age. We used representational similarity analysis (RSA) to reveal the temporal dynamics of the representation of gender, age, and identity for familiar and unfamiliar faces. We reasoned that a bottom-up account of familiarity effects would predict an early enhancement for familiar faces, whereas a top-down account of familiarity effects would involve later processing stages.

Results

We recorded MEG data from 16 subjects while viewing five images of each of eight familiar (American) and eight unfamiliar (German) celebrities and monitoring for consecutive repetitions of identical images (i.e., 1-back task; Fig. 1a). Celebrities varied orthogonally in gender and age. Images for each identity were chosen to incorporate natural variability in various dimensions, such as pose, hair style, eye gaze or lightning. Subjects viewed 28 repetitions of each of 80 face stimuli, each presented for 200 ms in individual trials. To reveal the time course of face processing, we performed multivariate pattern analysis on the MEG signals in a time-resolved manner (Fig. 1b) [15,16]. We first extracted a set of principal components (PCs) for each subject, based on MEG responses across sensors, trials, timepoints (-100 to 800 ms with respect to image onset; 1 ms resolution) and conditions. Then, using the resulting PCs, we trained and tested support vector machines (SVM) on every pair of stimuli for each time point. Dissimilarity for each pair of stimuli was computed as five-fold cross-validated decoding accuracy, resulting in one 80 x 80 MEG representational dissimilarity matrix (RDM) per subject and time point.

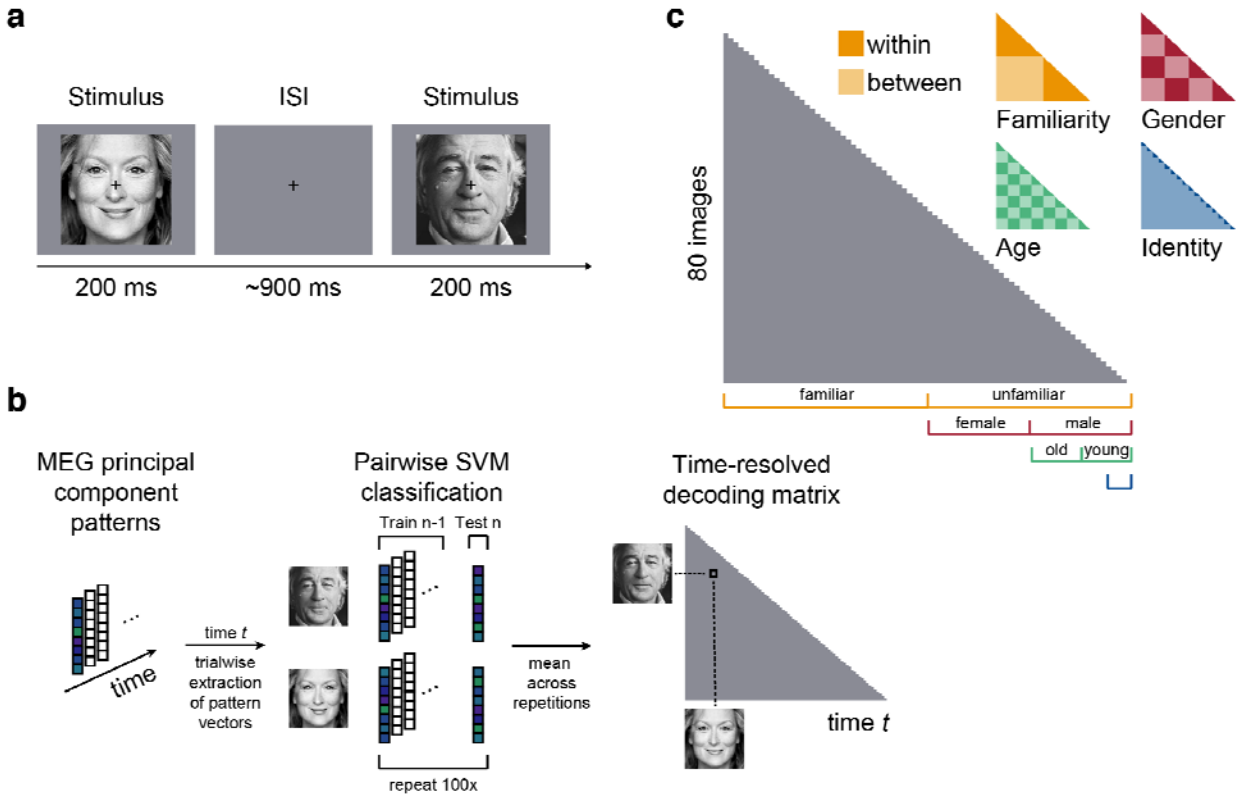

Figure 1 | Task and multivariate MEG analyses. (**a**) Subjects viewed 80 images of faces while performing a 1-back task on the image. Each image was presented for 200 ms, followed by a variable [800 – 1000 ms] interstimulus interval (ISI). (**b**) MEG analyses were performed in a time-resolved manner on principal components extracted from all MEG sensors, separately for each subject (see Methods). For each time point *t*, we extracted the pattern of response across PCs for each condition and each trial and performed pairwise cross-validated SVM classification. The resulting decoding accuracy values resulted in a 80 x 80 representational dissimilarity matrix (RDM) for each time point. (**c**) To perform representational similarity analysis (RSA), we constructed model RDMs for each face dimension (1 corresponding to "between" and 0 corresponding to "within" category, respectively).

Behavioral performance during MEG task. To ensure that subjects maintained attention and processed the presented images throughout the course of the experiment, we asked them to perform a one-back task, pressing a button when the identical image (not identity) was repeated consecutively. Subjects were highly sensitive to an image repetition (mean sensitivity index d' $\pm$ SEM: 4.28 $\pm$ 0.16) and responded quickly (mean response time $\pm$ SEM: 458 $\pm$ 10 ms after target stimulus onset). To test whether subjects processed familiar faces more efficiently than unfamiliar faces, we compared the responses to familiar versus unfamiliar images. While subjects' sensitivity did not differ between familiar and unfamiliar face images (p = 0.67; two-sided signed-rank test), subjects responded significantly faster to familiar than to unfamiliar faces (p = 0.021, familiar: 454 $\pm$ 11 ms, unfamiliar: 462 $\pm$ 9 ms). These behavioral results confirm that the processing of familiar faces is enhanced compared to unfamiliar faces, even when the task requires only image-level (not identity-level) processing.

Time course of face image decoding. To determine when neural representations can first discriminate any visual information, we computed the average across all pairwise decoding accuracy values, separately at each time point. This analysis yielded a time course of neural image decoding accuracy (Fig. 2a). Individual images could be discriminated by visual representations early (decoding first reached significance at 46 ms), reached a peak at 103 ms (73.3% mean decoding accuracy) and remained significantly above chance until 706 ms after stimulus onset ($p < 0.05$; cluster-corrected sign permutation test). This time course is highly similar to the course of image decoding reported in previous MEG studies [15-17] and shows how neural responses are resolved at the level of individual face images. To test how persistent neural responses to face image representations were, we further performed temporal generalization analysis (see Supplementary Information).

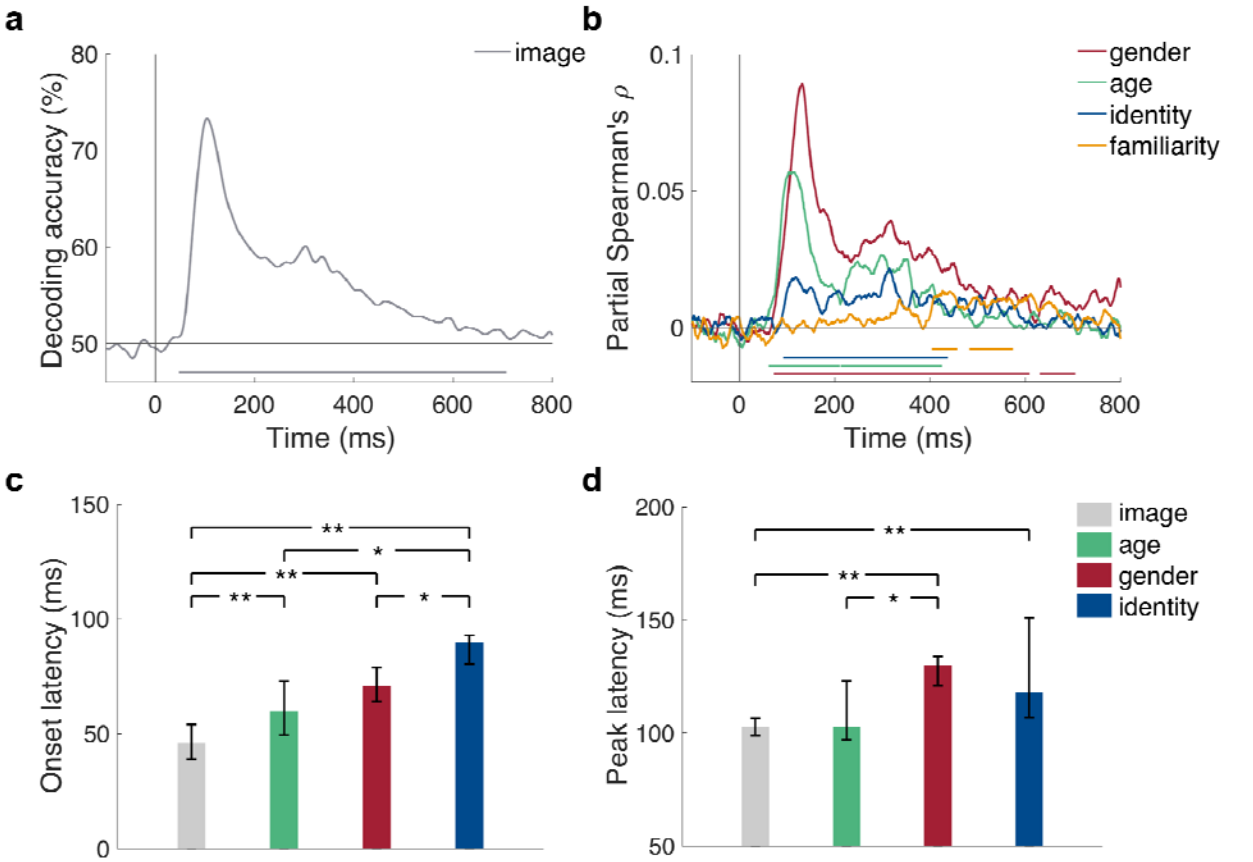

Figure 2 | Decoding of face images and dimensions from MEG signals. (**a**) Time course of image decoding where 0 indicates image onset (n = 16). (**b**) Time course of partial Spearman correlations between MEG RDMs and model RDMs (see Fig. 1c) for gender (red), age (green), identity (blue) and familiarity (orange), partialling out all other models and low-level features (see Methods). Lines below plots indicate significant times using cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$). (**c**) Onset and (**d**) peak latencies for decoding of images, gender, age and identity. Error bars indicate bootstrapped 95% confidence intervals. Stars above bars indicate significant differences across conditions (one-sample two-sided bootstrap test, **: $p < 0.01$; *: $p < 0.05$; FDR-corrected).

Early representations of face dimensions revealed by RSA. To determine when neural representations discriminated face dimensions at higher categorization levels (e.g., gender) or even at the level of identity or familiarity, we created a model RDM (e.g., “1” for between and “0” for within gender stimulus pairs) for every face dimension (i.e., gender, age, identity and familiarity; Fig. 1c). Because some face dimensions (e.g., gender) might be associated with differences in low-level image properties (e.g., long hair versus short hair for female and male, respectively), we further created a low-level feature RDM based on an early layer of a deep,

convolutional neural network trained on faces (i.e., VGG-face; see Methods for details and for comparison to other low-level feature models). For every face dimension, we computed partial Spearman correlations between corresponding model and MEG RDMs at every time point and for each subject, while partialling out all other model RDMs and the low-level feature model (Fig. 2b). We found that visual MEG representations were able to discriminate faces by age (significant time points: 60 – 208, 214 – 368 ms), gender (72 – 608, 630 – 705 ms) and specific identity (91 – 437 ms; all cluster-corrected sign permutation tests; cluster-defining threshold $p < 0.05$; and corrected significant level $p < 0.05$;). Age and gender information were extracted first from MEG representations, and arose around 20 ms earlier than identity information (Fig. 2c; $p < 0.05$; one-sample two-sided bootstrap test, FDR-corrected). This finding suggests that coarse distinctions between faces in categorical dimensions are extracted before finer distinctions on an identity level; in line with a coarse-to-fine processing of face dimensions [1,4]. Importantly, age, gender and identity dimensions were discriminated significantly later than individual images ($p < 0.01$) suggesting that these facial dimensions emerged early but later during processing than low-level image discrimination. Interestingly, we found that neural representation of gender and identity peaked at similar latencies ~125 ms after stimulus onset (Fig. 2d). This finding suggests that, while peak latencies often vary for different types of object categories [16] (presumably processed in different spatial locations), different dimensions of the same category (here faces) can peak at similar times.

We further found that MEG representations separated familiar from unfamiliar identities at much later latencies (403 – 457, 482 – 573 ms; cluster-corrected sign permutation test, cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$) than perceptual categories such as gender or age, and after specific identity information is extracted (all $p < 0.01$; one-sample two-sided bootstrap test, FDR-corrected). This finding indicates that a late signature of generic familiarity can be read out from MEG signals, long after image-invariant identity information is extracted. The basis of this familiarity signature is not clear, and could reflect the activation of memories associated with a given familiar individual, an emotional response to a familiar face, or a generic familiarity response.

Familiarity enhances face information at early stages. Behavioral evidence shows that familiar faces are processed more robust than unfamiliar faces, but it is unknown how early in processing this occurs. To answer this question, we conducted the same RSA analysis as above, but did so separately for familiar and unfamiliar faces (Fig. 3a). Note that this separation reduces the data available for each analysis fourfold, thus reducing the signal to noise ratio

(SNR). Despite this reduction, MEG representations still discriminated familiar faces by gender (Fig. 3b; 71 – 206, 226 – 492 ms), age (Fig. 3c; 92 – 200, 220 – 271 ms) and identity (Fig. 3d; 96 – 168, 252 – 406, 447– 516, 519 – 570 ms; cluster-corrected sign permutation test, cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$). In contrast, neural representations of unfamiliar faces were less pronounced but could still be discriminated by gender (Fig. 3b; 102 – 415, 488 – 560 ms) and age (Fig. 3c; 70 – 137, 241 – 397 ms; cluster-corrected sign permutation test, cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$), and were no longer discriminable by identity (Fig. 3d). Crucially, we found that the encoding of gender and identity, but not age, was significantly enhanced for familiar compared to unfamiliar faces (Fig. 3b, 3d, black lines in plots; cluster-corrected sign permutation test, cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$). These enhancements occurred early with respect to the onset of gender and identity encoding (gender: 66 ms – 142 ms (onset 71 ms); identity: 106 – 159 ms (onset 96 ms)). Overall, familiarity enhancement arose early during processing, suggesting that early stages of visual processing are tuned to familiar face features. However, we cannot fully exclude the contribution of fast feedback processes even at these early stages (c.f. [17], who report recurrent neural processes at early stages).

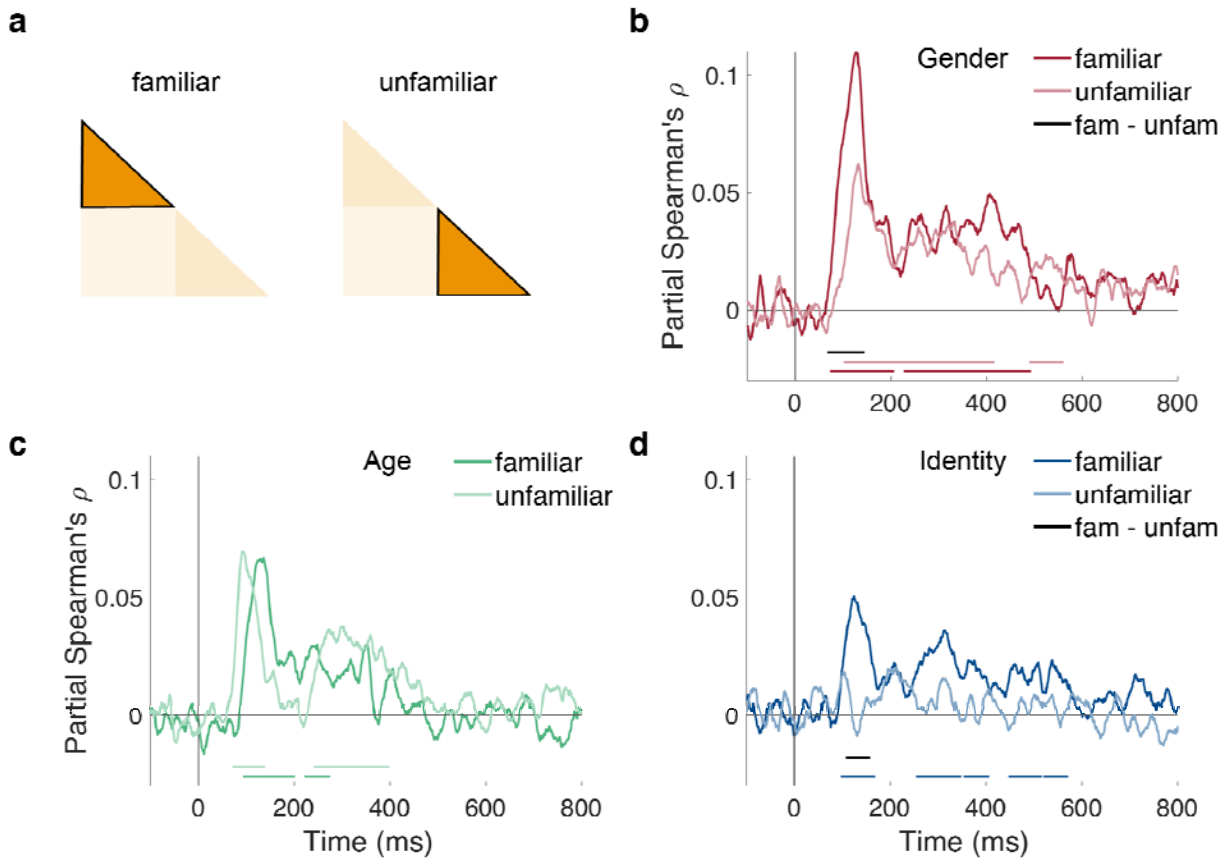

Figure 3 | Effect of familiarity on face representations. (**a**) We conducted separate RSA analyses for familiar and unfamiliar faces. Note that this reduces the amount of data fourfold. (**b-d**) Time course of partial Spearman correlations between MEG RDMs and gender (**b**), age (**c**) and identity (**d**) separated for familiar and unfamiliar faces, partialling out other models and low-level features (see Methods). Colored lines below plots indicate significant times, and black lines indicate significant difference between conditions both using cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$).

MEG responses are correlated with behavior. Not all information that can be read out from brain activity is used by the brain to guide behavior [18,19]. Is the face information we report here related to behavior? To find out, fourteen of the sixteen subjects participated in a behavioral multi-arrangement paradigm (see Methods) after the MEG experiment to assess their perceived dissimilarity between the face stimuli. We correlated the resulting behavioral RDM of each subject with the MEG RDMs for each time point for that subject, while partialling out low-level stimulus features. The resulting correlation time courses revealed a significant partial correlation between MEG and behavior (Fig. 4a; 85 – 274, 294 – 420 ms after stimulus onset; cluster-corrected sign permutation test, cluster-defining threshold $p < 0.05$, and corrected significance

level $p < 0.05$) with a peak at 123 ms (partial Spearman's partial ρ: 0.06) after stimulus onset. To assess how much of the explainable MEG variance was captured by behavior, we further computed an estimate of the noise ceiling (gray shaded area in Fig. 4a) given the variability across the restricted set of fourteen subjects (gray shaded area in Fig. 4a). While the correlation between behavior and MEG reveals the shared variance between both modalities, it does not indicate how much each of the face dimensions contributes to this shared variance. To answer this question, we conducted a model-based commonality analysis (Fig. 4b; see Methods). This approach is based on variance partitioning and identifies the variance uniquely shared between MEG and behavior and a given model RDM (e.g., the gender model), termed commonality coefficients. We restricted this analysis to the time window during which we found significant correlations between MEG and behavior (i.e., 85 – 420 ms after stimulus onset; note the changed x-axis in Fig. 4b). Given the late correlation with the MEG data and the familiarity model (see Fig. 2b), we did not include the familiarity model in this analysis. We found that gender, age and identity each uniquely contributed to the shared variance (commonality coefficients are shown in Fig. 4b). We further show the explained variance between MEG and behavior as reference (gray line in Fig. 4b). Note that commonality coefficients reported with this kind of analysis are often significant but very small [20]. Together, we found that behavior was predictive for MEG responses and that the shared variance predominantly reflected gender information, followed by age and then identity, though all three were significant.

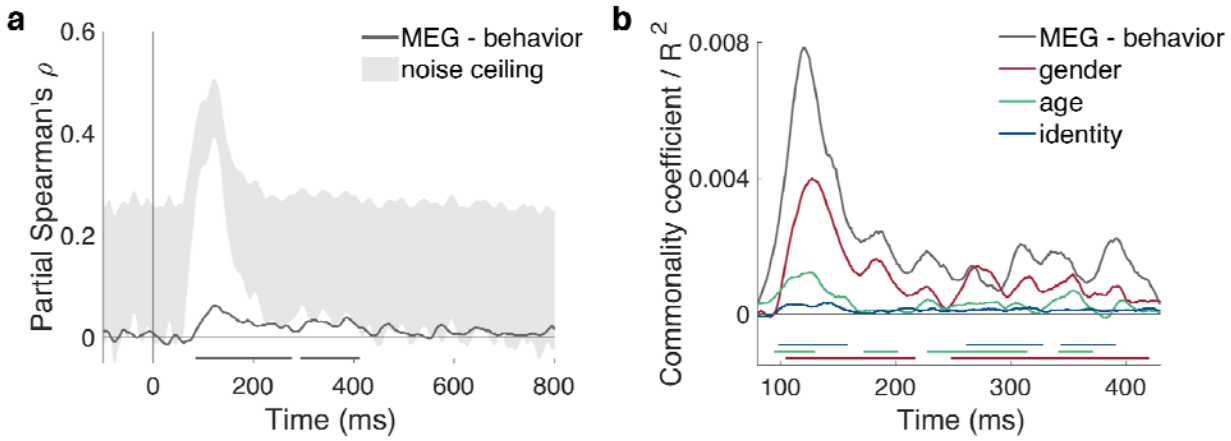

Figure 4 | MEG and behavioral RDM comparisons. (**a**) Time course of correlation (partial Spearman, partialling out low-level features) between behavioral and MEG RDMs (n = 14). Gray-shaded area indicates estimated noise ceiling based on the variability across subjects. (**b**) Model-based commonality analysis showing the portion of shared variance between MEG and behavior (gray line) uniquely explained by gender (i.e., commonality coefficient; red), age (green) and identity (blue). Note that this

analysis was restricted to the time window of the significant correlation between MEG and behavior (as shown in **a**). Colored lines below plots indicate significant times using cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$).

Discussion

This study answers two fundamental questions about the time course of face processing in humans. First, we find that extraction of information about gender and age begins after image-level decoding but before extraction of identity information. Second, we show that familiarity of the face enhances representations of gender and identity very early in processing. These two previously unanswered questions reveal the temporal dynamics underlying face processing and provide powerful constraints on computational models of face perception. Next we relate these findings to prior work in monkeys and humans, as well as computational models of face perception.

While a few prior studies have investigated the time course of face perception using multivariate pattern analyses in humans [21-24], our work goes beyond previous findings in two important respects. First, previous studies have focused on a single facial dimension (e.g., identity [22-24]; viewpoint [21]), and hence could not address the relative timing of extraction of multiple dimensions of face perception. To our knowledge, only one prior study [23] investigated the representation of identity within and across gender, and reported no difference in onset latencies. Their analysis, however, included identity differences in the gender comparison (only two identities per gender) whereas we investigated gender information unconfounded from identity (i.e., by partialling out the identity model). While these prior studies also find very early decoding of face identity information, consistent with our results, they do not reveal the relative timing of extraction of different dimensions from faces.

Second, and most importantly, our study is the first to reveal how and when familiarity affects the representations of different face dimensions. The neural mechanisms underlying the powerful behavioral differences between the perception of familiar versus unfamiliar faces [7,25] have remained an important unsolved mystery. In particular, is the familiarity enhancement effect due to tuning of bottom-up perceptual filters, or to top-down feedback [8,9,11]? Previous studies could not address this question as they used only unfamiliar face images [22-24], indirect analyses of steady-state visual evoked responses [26] or functional magnetic resonance imaging with low temporal resolution [27-31] . Our finding that identity and gender information are significantly enhanced for familiar faces, very early on in processing, and virtually as early as

this information becomes available, provides a possible neural mechanism for the strong behavioral enhancement in perception of familiar versus unfamiliar faces. Namely, our results suggest that familiarity affects face perception by altering feed-forward face processing, not exclusively through feedback from later stages after the personal identity of a face has been extracted.

How do our findings relate to prior findings on the spatial organization of face processing in the brain? In our study, we consider whole-brain information and do not restrict the time course to certain spatial locations given the low spatial resolution of MEG. Despite the lack of spatial resolution, the finding that gender and age were extracted before identity information suggests that these aspects of face perception are processed at different stages of face processing. Evidence from fMRI is equivocal about where specific face dimensions such as gender [32,33] or identity [34-36] information is represented in the brain. Given the fleeting presence of face dimensions in our MEG data, it is possible that fMRI misses some of this information due to its low temporal resolution. Interestingly, previous studies found that people with developmental prosopagnosia, who have trouble recognizing faces, have no impairments in gender processing, suggesting that these two facial dimensions might be processed in distinct neural areas, potentially in parallel [37,38]. In contrast, behavioral studies showed that gender processing influences face identification [6]. Our results reconcile these two findings by suggesting that gender and age are processed earlier in the processing hierarchy, at stages possibly less affected by prosopagnosia, yet able to influence subsequent identity processing in typical subjects [6]. Taken together, it remains unclear where face dimensions, such as gender, age or identity, under investigation in this study are represented in the brain. In future, this question might be answered by combining MEG with fMRI using a “fusion” approach [16,17] to link our finding to regions in the brain.

While human neuroimaging techniques are mainly limited to either high temporal or high spatial resolution, neurophysiological recordings in non-human primates provide an ideal opportunity to simultaneously measure face representations with high temporal and spatial resolution. Indeed, the macaque face perception system is similar [39,40] and possibly homologous [41] to the human face processing system. Neurophysiological studies with non-human primates find that categorical distinctions between faces and other categories develop earlier than face identity information [40,42], but it is still unknown when other face dimensions, such as gender or age, emerge. Consistent with human MEG data [21], facial representations in macaques were also found to gradually build up and become more invariant to viewpoint at successive processing stages, measured both spatially and temporally [39,43], again showing the

usefulness of the macaque's face perception system as a model to study human face perception. With regard to familiarity, our findings are in line with a recent study in macaques reporting early quantitative differences and late qualitative differences in processing of familiar versus unfamiliar faces [44]. However, the paradigms and stimuli that have been used so far in humans and macaque studies are too different to provide a precise correspondence between species.

Our findings place important constraints on computational models of face perception and further suggest new hypotheses to probe in such models. Recently, deep convolutional neural networks (CNNs), have provided successful models of object and scene perception in humans and macaques [45-47]. However, it remains unclear how useful CNNs are as model for human face processing. In future work, it will be interesting to test how well different computational models can explain our findings. In particular, the fact that we found little temporal cross-decoding (see Supplementary Information) suggests that a sequence of non-linear operations are performed on the representations of each face dimension. Comparisons of our data to different computational models might shed light on the operations and transformations performed at different stages of face processing in humans, as has been successfully done for facial viewpoint decoding in macaques [48]. Another important implication of our data is that early stages of face processing are apparently tuned to familiar faces, a phenomenon that could also be tested in computational models. Analyzing these questions could provide a path toward the development of a computationally precise, image-computable model of face processing in humans.

While the results presented here further our understanding of human face processing dynamics, they also have several limitations. First, our decision to use natural image in this study for ecological validity may introduce greater low-level image confounds compared to highly controlled, artificial face stimuli. Although we partialled out low-level features as measured by an early layer of a CNN trained on faces (a conservative choice, given that this model had the greatest overall correlation with the MEG data), we cannot be sure that all low-level features have been captured by this model. Second, by investigating the onset of extraction of several face dimensions, we cannot draw conclusions about when the processing is completed, because the earliest latency of significant decoding reflects an upper bound for the beginning of the process. In fact, the correlations with gender, age and identity were relatively sustained until at least 400 ms after stimulus onset. Third, the mere existence of representations revealed by multivariate pattern analysis does not in itself imply that these representations are relevant to behavior [18,19]. Here, we correlated MEG to behavioral similarity

of our subjects and found that all face dimensions explained unique variance between MEG and behavior. While this is an important step towards linking MEG representations with behavior, more direct links such as correlates to online behavior during MEG recording would provide stronger evidence. However, our results do make testable predictions about human face processing behavior. For example, in line with our results, studies investigating the speed of human categorization behavior have shown that identity decisions were made faster than familiarity decisions [1] and that gender decision occurred faster for familiar than unfamiliar faces [49]. Our results go beyond these findings to predict that behavioral discriminations of gender (and age, if tested in a binary fashion) should be made faster than discriminations of identity.

In sum, our findings of how face processing unfolds over time in humans show that the extraction of face dimensions follow a coarse-to-fine time trajectory and support the hypothesis that the face processing system is tuned to familiar face features in a bottom-up manner. These findings inform broader questions about how prior experience affects processing of other stimuli beyond faces, such as object shapes [12-14], offer powerful constraints on computational models of face perception, and provide new predictions to be tested in future work.

Material and Methods

Participants. Twenty-one healthy volunteers with normal or corrected-to-normal vision participated in the study. Five subjects were excluded before data analysis due to at least one of the following exclusion criteria: excessive motion during the recording, behavioral performance below two standard deviations of the mean, or incomplete recordings due to technical issues. Data from 16 subjects (eight female; mean age 25.9, SD = 4.33) remained for the MEG analysis. The chosen sample size was based on previous studies using multivariate decoding of EEG/MEG data [16,17,23]. Fourteen of these 16 subjects additionally participated in an online behavioral follow-up experiment. All subjects provided informed, written consent prior to the experiment. The Massachusetts Institute of Technology (MIT) Committee on the Use of Humans as Experimental Subjects approved the experimental protocol (COUHES No 1606622600) and the study was conducted according to the Declaration of Helsinki.

Experimental design and stimuli. To investigate the temporal dynamics of face processing, subjects viewed face images of different identities while monitoring for consecutive repetitions of identical images (i.e., 1-back task; Fig. 1a) in the MEG. We chose eight “familiar” (i.e., famous actors in the US) and eight “unfamiliar” (i.e., German actors) celebrities as identities, which

varied orthogonally in gender and age, such that half were female and half were male and half of them were “young” (i.e., maximum age was 36 years) and half were “old” (i.e., minimum age was 59 years).

To ensure that all subjects were familiar with the “familiar” identities, subjects completed an online screening task prior to the study. In this screening, we presented them with one image for each of the 16 identities (different from the images used in the MEG study) and asked if they were familiar with the person shown. Only subjects who recognized each of the eight familiar identities (e.g., by giving their names or contexts in which they remembered the person) were included in the study.

Final stimuli used in the MEG study consisted of five gray-scale images of each of the 16 identities for a total of 80 stimuli. For each identity, we selected five images from the internet which varied in several aspects, such as expression (at least two smiling and two neutral facial expressions), eye gaze (one averted to the left, one averted to the right, two directed gaze and one gaze aligned with rotated head), pose (one with head slightly rotated to the side), lightning, hair, etc. We then standardized all images to a template by rotating, scaling and cropping them based on the position of the nose tip, the mouth center and both eyes and saved them as gray-scale images.

During the MEG experiment, subjects viewed trials of face images (Fig. 1a). Each trial started with the presentation of a face image for 0.2 s followed by a 0.8 - 1 s interstimulus interval (ISI; uniformly sampled between 0.8 and 1 s) during which a gray screen was presented. Subjects were instructed to respond via button press to a consecutive repetition of an identical image during image presentation or during ITI. To avoid artifacts due to eye movements or blinking, subjects were instructed to fixate a black fixation cross in the upper center of the screen during image presentation (i.e., presented between the tip of the nose and the eyes of a face) and ISI. They were further asked to blink at the same time when giving a button response, as these trials were not included in the data analysis.

Subjects viewed 28 blocks of trials in which each of the 80 images was presented once randomly interleaved with 20 task trials (1-back task) for a total of 100 trials per block. Task trials were pseudo-randomized such that each of the 80 images was additionally shown seven times as task trial for a total of 35 presentations. The experiment lasted around 70 minutes.

MEG recording and preprocessing. MEG data were collected using a 306-channel Elekta Triux system with a 1000 Hz sampling rate, and were filtered online between 0.01 and 330 Hz. The position of the head was tracked during MEG recording based on a set of five head position

indicator coils placed on particular landmarks on the head. We preprocessed the raw data with Maxfilter software (Elekta, Stockholm) to remove head motion and to denoise the data using spatiotemporal filters. We then used Brainstorm (version 3.4 [50]) to extract trials from -200 to 800 ms with respect to image onset. In Brainstorm, every trial was baseline-corrected by removing the mean activation from each MEG sensor between -200 ms and stimulus onset and principal component analysis was used to remove eye blink artifacts which were automatically detected from frontal sensor MEG data. We used a 6000 fT peak-to-peak rejection threshold to discard bad trials, imported the remaining trials in Matlab (version 2016a; The Mathworks, Natick, MA) and smoothed them with a 30 Hz low-pass filter. To further decrease noise and to reduce computational costs, for each subject we concatenated data of each MEG sensor over time and applied principal component analysis to the MEG sensor data (keeping all components that explained 99.99% of the variance in the data). This step reduced the set of features from 306 MEG sensors to around 70 principal components per subject and we conducted all further analysis on this reduced set. We then baseline-corrected every trial by removing the mean activation between -200 ms and stimulus onset from each principal component. These principal component scores for each trial and each time point were used for the subsequent analyses.

MEG multivariate pattern analysis. We used multivariate pattern analysis to extract temporal information about the face stimuli from the MEG data (Fig. 2). To obtain a similarity measure for each pair of stimuli, we used cross-validated pairwise classification accuracy of linear support vector machines (SVM; libsvm [51]). Classification analysis was performed separately for each subject in a time-resolved manner (i.e., independently for each time point). A pattern in the analysis consisted of the principal component scores for one trial and one condition at a given time point. In the first step, we sub-averaged all trials of one condition by randomly assigning each trial to one of five splits and averaging the trials in each split (~5-7 trials per split when considering bad trials). We then divided the groups into training and testing data randomly selecting one group for testing and the remaining groups for training (i.e., five-fold cross-validation). We then conducted a binary classification of all 3170 pairwise comparisons (i.e., 80x79/2 combinations) between conditions. This classification procedure was repeated 100 times. The average decoding accuracies over repetitions served as value in the 80 x 80 decoding matrix, termed representational dissimilarity matrix (RDM). This RDM is symmetric and the diagonal is undefined. The entire procedure resulted in one MEG RDM for each subject and time point.

To get a measure of how well each face stimulus can be discriminated from all other images in the MEG (i.e., image decoding), we averaged all pairwise decoding accuracies in the lower triangular of each RDM. This resulted in one average decoding accuracy value per subject and time point. The time course of image decoding further serves as benchmark of time course of low-level image processing in the MEG data. To investigate how persistent neural responses were to face images, we further extended the SVM decoding procedure with a temporal generalization approach [16,52,53]. Details and results of this analysis can be found in the Supplementary Information.

Representational similarity analysis. To analyze the representation of face dimensions in the MEG data, we used representational similarity analysis (RSA). We created model RDMs for each face dimension which were 80 x 80 binary matrices where “1” corresponded to a between category stimulus comparison (e.g., male vs female for the gender model) and “0“ to a within category stimulus comparison (e.g., female vs female). This procedure resulted in four face models corresponding to the familiarity, gender, age and identity dimensions of our stimuli. To compute correlations between each model and the MEG data, we extracted the lower off-diagonal of each of these matrixes as vectors. For each model and subject, we computed the partial coefficients (Spearman correlation) between the model and the MEG RDM at each time point partialling out all other face models. This step was crucial because some of the models are correlated (e.g., between identity comparisons comprised between gender comparisons) and partialling out the other models thus allowed us to disentangle contributions of the models from each other.

To further exclude the contribution of low-level features of our stimuli to the results, we additionally partialled out a low-level feature model. This low-level feature model was computed by extracting features for each of the 80 stimuli from the second convolutional layer of a deep, convolutional artificial neural network (CNN) trained on thousands of face images (VGG-face [54]). We used 1 – Pearson correlation as a measure of dissimilarity between the CNN units of each pair of stimuli, resulting in a 80 x 80 RDM based on low-level image features. Note that we also compared other models of low-level features (e.g., HMAX C2 [55,56], Gist [57], pixel-based similarity), which produced similar results; we report here the VGG-face model because it reached the maximum correlation with the MEG data and hence explains the most data (as accountable by low-level features).

We investigated the effect of familiarity on face processing by dividing the MEG and model RDMs into within familiar and within unfamiliar RDMs, respectively. Each of these RDMs

was a 40 x 40 RDM constituting of only familiar or only unfamiliar face images. We then performed the same analysis as for the full set of stimuli (see above). To further test differences between familiar and unfamiliar face processing, we subtracted the time courses of correlation for unfamiliar faces from the time courses obtained for familiar faces for each subject and statistically compared these difference time courses to zero (see Statistical inference below).

Behavioral similarity experiment. Fourteen of the sixteen subjects additionally performed a behavioral multi-arrangement task [58] on the same stimuli on a separate day after the MEG experiment. Subjects performed the multi-arrangement experiment online using their own computer and by logging into an online platform to run behavioral experiments (www.meadows-research.com). Subjects had to enter an anonymous, personal code that was provided to them via email to start the experiment. In the experiment, all 80 stimuli that the subject had previously seen in the experiment were arranged as thumbnails around a white circle in the center of the screen. Subjects were instructed to arrange these thumbnails based on their perceived similarity ("similar images together, dissimilar images apart", without explicit instructions on which feature to use) by dragging and dropping them in the circle. The experiment terminated automatically when a sufficient signal to noise ratio was reached (i.e., evidence weight was set to 0.5). The average duration of the experiment was ~70 minutes. After the completion of the experiment, the pairwise squared on-screen distances between the arranged thumbnails was computed, thus representing a behavioral RDM. For each subject, we extracted the lower off-diagonal data from the behavioral RDM and correlated this vector with the corresponding MEG RDMs for each time point. To assess the unique contribution of each model to the shared variance between MEG and behavioral RDMs, we additionally performed commonality analysis, a variance partitioning approach that estimates the shared variance between more than two variables [20, 59]. Briefly, we computed the variance uniquely contributed from each face model (e.g., gender) by calculating two correlation coefficients: First, for each subject, we calculated the partial correlation between MEG and behavioral RDMs, while partialling out all models (gender, age, identity and low-level feature model). Second, we calculated the partial correlation between MEG RDM and behavioral RDM while partialling out all face models and the low-level feature model but leaving one face model out (e.g., gender). The difference between these two partial correlation coefficients represents the unique variance contributed by that model referred to as "commonality coefficient". This step was repeated for every MEG time point resulting in a commonality coefficient time course for each face model.

Statistical inference. For all analyses, we used non-parametric statistical tests that do not rely on assumptions on the distributions of the data [60,61]. For statistical inference of decoding accuracy (image decoding) or partial correlation (e.g., model correlation) time series, we performed permutation-based cluster-size inference (i.e., a cluster refers to a set of contiguous time points). The null hypothesis corresponded to 50% chance level for decoding accuracies, and 0 for correlation values or correlation differences. Significant temporal clusters were defined as follows. First, we permuted the condition labels of the MEG data by randomly multiplying subject responses by +1 or -1 (i.e., sign permutation test). We repeated this procedure 1000 times resulting in a permutation distribution for every time point. Second, time points that exceeded the 95th percentile of the permutation distribution served as cluster-inducing time points (i.e., equivalent to $p < 0.05$; one-sided). Lastly, clusters in time were defined as the 95th percentile of the maximum number of contiguous, significant time points across all permutations (i.e., equivalent to $p < 0.05$; one-sided).

Onset and peak latency analysis. To test for statistical differences in onset or peak latencies between different face dimensions, we performed bootstrap tests. We bootstrapped the subject-specific time courses (e.g., measures as decoding accuracy, partial correlation or commonality coefficient) 1000 times to obtain an empirical distribution of the onset (i.e., minimum significant time point post stimulus onset) and peak latencies (i.e., maximum correlation value between 80 and 180 ms post stimulus onset). We restricted the time window for the peak analysis to 180 ms post stimulus onset, since we were interested in the first peak occurring after stimulus onset, unconfounded from later peaks (e.g., due to stimulus offset responses [62]). The 2.5th and the 97.5th percentile of these distributions defined the 95% confidence interval for onset and peak latency, respectively. For differences between latencies, we computed 1000 bootstrap samples of the difference between two latencies (e.g., onset) resulting in an empirical distribution of latency differences. The number of differences that were smaller or larger than zero divided by the number of permutations defined the p-value (i.e., two-sided testing). These p-values were corrected for multiple comparisons using false discovery rate at a 0.05 level.

Data and code availability. The stimuli used in this study can be downloaded from the Open Science Framework (https://osf.io/gk6f5/). Data will be made available upon request. We used Elekta MaxFilter software, publically available brain analysis software (Brainstorm) and experimental platform (Meadows), the libsvm toolbox and standard Matlab functions for analysis.

Acknowledgements

This work was supported by a Feodor-Lynen postdoctoral fellowship of the Humboldt Foundation to K.D., NIH grant Grant DP1HD091947 to N.K and National Science Foundation Science and Technology Center for Brains, Minds, and Machines.

Author contributions

K.D., L.I. and N.K. designed the experiment. K.D. collected the data. K.D. and D.P. analyzed the data. K.D., L.I., D.P. and N.K. wrote the article.

Competing financial interests

The authors declare no competing financial interest.

References

1. Besson, G. *et al.* From face processing to face recognition: Comparing three different processing levels. *Cognition* **158,** 33–43 (2017).
2. Crouzet, S. M., Kirchner, H. & Thorpe, S. J. Fast saccades toward faces: Face detection in just 100 ms. *J. Vis.* **10,** 1–17 (2010).
3. Martin, J. G., Davis, C. E., Riesenhuber, M. & Thorpe, S. J. Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search. *Sci. Rep.* **8,** 12482 (2018).
4. Goffaux, V. *et al.* From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing. *CerCor* **21,** 467–476 (2010).
5. Rossion, B. Is sex categorization from faces really parallel to face recognition? *Vis. Cogn.* **9,** 1003–1020 (2002).
6. Zhao, M. & Hayward, W. G. Integrative processing of invariant aspects of faces: Effect of gender and race processing on identity analysis. *J. Vis.* **13,** 15–15 (2013).
7. Young, A. W. & Burton, A. M. Are We Face Experts? *Trends Cogn. Sci.* **22,** 100–110 (2018).
8. Visconti di Oleggio Castello, M. & Gobbini, M. I. Familiar Face Detection in 180ms. *PLoS ONE* **10,** e0136548–18 (2015).
9. Kramer, R. S. S., Young, A. W. & Burton, A. M. Understanding face familiarity. *Cognition* **172,** 46–58 (2018).

10. Gobbini, M. I. & Haxby, J. V. Neural response to the visual familiarity of faces. *Brain Res. Bull.* **71,** 76–82 (2006).
11. Schwartz, L. & Yovel, G. The roles of perceptual and conceptual information in face recognition. *J. Exp. Psychol.-Gen.* **145,** 1493–1511 (2016).
12. Meyer, T., Walker, C., Cho, R. Y. & Olson, C. R. Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nat. Neurosci.* **17,** 1388–1394 (2014).
13. Op de Beeck, H. P. & Baker, C. I. The neural basis of visual object learning. *Trends Cogn. Sci.* **14,** 22–30 (2010).
14. Woloszyn, L. & Sheinberg, D. L. Effects of Long-Term Visual Experience on Responses of Distinct Classes of Single Units in Inferior Temporal Cortex. *Neuron* **74,** 193–205 (2012).
15. Carlson, T., Tovar, D. A., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: The first 1000 ms. *J. Vis.* **13,** 1–1 (2013).
16. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17,** 1–10 (2014).
17. Mohsenzadeh, Y., Qin, S., Cichy, R. M. & Pantazis, D. Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *eLife* **7,** e36329 (2018).
18. Grootswagers, T., Cichy, R. M. & Carlson, T. A. Finding decodable information that can be read out in behaviour. *NeuroImage* 1–37 (2018). doi:10.1016/j.neuroimage.2018.06.022
19. Williams, M. A., Dang, S. & Kanwisher, N. G. Only some spatial patterns of fMRI response are read out in task performance. *Nat. Neurosci.* **10,** 685–686 (2007).
20. Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I. & Cichy, R. M. The representational dynamics of task and object processing in humans. *eLife* **7,** 509 (2018).
21. Kietzmann, T. C., Gert, A. L., Tong, F. & König, P. Representational Dynamics of Facial Viewpoint Encoding. *J. Cogn. Neurosci.* **29,** 637–651 (2017).
22. Nemrodov, D., Niemeier, M., Patel, A. & Nestor, A. The Neural Dynamics of Facial Identity Processing: insights from EEG-Based Pattern Analysis and Image Reconstruction. *eNeuro* **5,** e0358–17.2018– 1–17 (2018).
23. Nemrodov, D., Niemeier, M., Mok, J. N. Y. & Nestor, A. The time course of individual face recognition: A pattern analysis of ERP signals. *NeuroImage* **132,** 469–476 (2016).

24. Vida, M. D., Nestor, A., Plaut, D. C. & Behrmann, M. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *PNAS* 388–393 (2017). doi:10.6084/m9.figshare.4233107.v1
25. Jenkins, R., White, D., Van Montfort, X. & Burton, A. M. Variability in photos of the same face. *Cognition* **121,** 313–323 (2011).
26. Collins, E., Robinson, A. K. & Behrmann, M. Distinct neural processes for the perception of familiar versus unfamiliar faces along the visual hierarchy revealed by EEG. *NeuroImage* **181,** 120–131 (2018).
27. Natu, V. S. & O'Toole, A. J. Spatiotemporal changes in neural response patterns to faces varying in visual familiarity. *NeuroImage* **108,** 151–159 (2015).
28. 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.* 1–14 (2017). doi:10.1038/s41598-017-12559-1
29. Ramon, M., Dricot, L. & Rossion, B. Personally familiar faces are perceived categorically in face-selective regions other than the fusiform face area. *Eur. J. Neurosci.* **32,** 1587–1598 (2010).
30. Weibert, K. *et al.* An image-invariant neural response to familiar faces in the human medial temporal lobe. *Cortex* **84,** 34–42 (2016).
31. Rossion, B., Schiltz, C. & Crommelinck, M. The functionally defined right occipital and fusiform 'face areas' discriminate novel from visually familiar faces. *NeuroImage* **19,** 877–883 (2003).
32. Kaul, C., Rees, G. & Ishai, A. The Gender of Face Stimuli is Represented in Multiple Regions in the Human Brain. *Front. Hum. Neurosci.* **4,** 1–12 (2011).
33. Podrebarac, S. K., Goodale, M. A., van der Zwan, R. & Snow, J. C. Gender-selective neural populations: evidence from event-related fMRI repetition suppression. *Exp. Brain Res.* **226,** 241–252 (2013).
34. Anzellotti, S., Fairhall, S. L. & Caramazza, A. Decoding Representations of Face Identity That are Tolerant to Rotation. *CerCor* **24,** 1988–1995 (2013).
35. Dobs, K., Schultz, J., Bülthoff, I. & Gardner, J. L. Task-dependent enhancement of facial expression and identity representations in human cortex. *NeuroImage* **172,** 689–702 (2018).
36. Kriegeskorte, N., Formisano, E., Sorger, B. & Goebel, R. Individual faces elicit distinct response patterns in human anterior temporal cortex. *PNAS* **104,** 20600–20605 (2007).

37. Chatterjee, G. & Nakayama, K. Normal facial age and gender perception in developmental prosopagnosia. *Cogn. Neuropsychol.* **29,** 482–502 (2012).
38. DeGutis, J., Chatterjee, G., Mercado, R. J. & Nakayama, K. Face gender recognition in developmental prosopagnosia: Evidence for holistic processing and use of configural information. *Vis. Cogn.* **20,** 1242–1253 (2012).
39. Freiwald, W. A. & Tsao, D. Y. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330,** 845–851 (2010).
40. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. & Livingstone, M. S. A Cortical Region Consisting Entirely of Face-Selective Cells. *Science* **311,** 670–674 (2006).
41. Lafer-Sousa, R., Conway, B. R. & Kanwisher, N. G. Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques. *J. Neurosci.* **36,** 1682–1697 (2016).
42. Tsao, D. Y. & Livingstone, M. S. Mechanisms of Face Perception. *Annu. Rev. Neurosci.* **31,** 411–437 (2008).
43. Meyers, E. M., Borzello, M., Freiwald, W. A. & Tsao, D. Intelligent Information Loss: The Coding of Facial Identity, Head Pose, and Non-Face Information in the Macaque Face Patch System. *J. Neurosci.* **35,** 7069–7081 (2015).
44. Landi, S. M. & Freiwald, W. A. Two areas for familiar face recognition in the primate brain. *Science* **357,** 591–595 (2017).
45. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6,** 1–13 (2016).
46. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19,** 356–365 (2016).
47. Groen, I. I. *et al.* Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *eLife* **7,** (2018).
48. Yildirim, I., Freiwald, W. & Tenenbaum, J. Efficient inverse graphics in biological face processing. Preprint at https://www.biorxiv.org/content/early/2018/04/02/282798 (2018).
49. Balas, B., Cox, D. & Conwell, E. The Effect of Real-World Personal Familiarity on the Speed of Face Information Processing. *PLoS ONE* **2,** e1223–5 (2007).
50. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience* **2011,** 1–13 (2011).

51. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2,** 27 (2011).
52. Isik, L., Meyers, E. M., Leibo, J. Z. & Poggio, T. The dynamics of invariant object recognition in the human visual system. *J. Neurophysiol.* **111,** 91–102 (2014).
53. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18,** 203–210 (2014).
54. Parkhi, O. M., Vedaldi, A. & Zisserman, A. Deep face recognition. in **1,** 6 (2015).
55. Riesenhuber, M. & Poggio, T. Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* **12,** 162–168 (2002).
56. Serre, T., Wolf, L. & Poggio, T. Object Recognition with Features Inspired by Visual Cortex. (2006).
57. Torralba, A. & Oliva, A. Statistics of natural image categories. *Network: Comput. Neural Syst.* **14,** 391–412 (2003).
58. Kriegeskorte, N. & Mur, M. Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Front. Psychol.* **3,** 245 (2012).
59. Seibold, D. R. & McPHEE, R. D. Commonality analysis: A method for decomposing explained variance in multiple regression analyses. *Human Communication Research* **5,** 355–365 (1979).
60. Pantazis, D., Nichols, T. E., Baillet, S. & Leahy, R. M. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *NeuroImage* **25,** 383–394 (2005).
61. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164,** 177–190 (2007).
62. Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J. & Turret, J. High temporal resolution decoding of object position and category. *J. Vis.* **11,** 9–9 (2011).