SCIENTIFIC REPORTS
natureresearch

Check for updates

OPEN

Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis

Daniel Kaiser✉ & Karen Nyga

When we see a face, we rapidly form an impression of its attractiveness. Here, we investigated how rapidly representations of facial attractiveness emerge in the human brain. In an EEG experiment, participants viewed 100 face photographs and rated them for their attractiveness. Using time-resolved representational similarity analysis on the EEG data, we reveal representations of facial attractiveness after 150–200 ms of cortical processing. Interestingly, we show that these representations are related to individual participants' personal attractiveness judgments, suggesting that already early perceptual representations of facial attractiveness convey idiosyncratic attractiveness preferences. Further, we show that these early representations are genuinely related to attractiveness, as they are neither explained by other high-level face attributes, such as face sex or age, nor by features extracted by an artificial deep neural network model of face processing. Together, our results demonstrate early, individually specific, and genuine representations of facial attractiveness, which may underlie fast attractiveness judgments.

When we see a face, we almost immediately can tell whether we find it attractive or not[1]. Beyond such first impressions, facial attractiveness affects people's everyday lives in fundamental ways: for instance, an attractive face grants advantages in various aspects, such as increased success in dating[2,3], receiving help more often[4,5], and being more successful on the job market[6,7].

Given these varied effects of facial attractiveness, a large body of research has focused on understanding the factors that make a face attractive. One line of research has tried to establish objective physical markers of facial attractiveness[8,9], revealing that faces are more attractive when they are more similar to the average[10,11], more symmetric[11–13], or have favourable sexual characteristics[14]. This research has led into advances in computer vision, providing algorithms that can predict how attractive humans will find a particular face[15,16]. However, there is considerable agreement that there is also a subjective component to facial attractiveness, with responses varying substantially between observers[17–19]. Indeed, our impression of facial attractiveness may depend on both: an objectively attractive physical composition of visual features and an idiosyncratic appreciation of these features.

How does the brain achieve the transition from physical stimulus properties into an individual representation of facial attractiveness? To answer this question, previous studies have used event-related potentials (ERPs) obtained from EEG recordings to investigate when brain responses to more or less attractive faces differ. Although in many such studies facial attractiveness influences multiple ERP components, a key question is when facial attractiveness is *first* represented in EEG signals. The answers to this question are quite mixed. Some studies highlight relatively early ERP modulations depending on facial attractiveness, for instance at the N170 processing stage[20–26], suggesting that attractiveness is analysed during perceptual face processing. Other studies suggest that attractiveness is first analysed during the N250 processing stage[27–30], at the time when face identity is represented. Finally, some studies only find late ERP modulations, at the P3 stage or during later slow waves starting from around 350–400 ms[31–37], suggesting that facial attractiveness is analysed during post-perceptual stages of cognitive processing.

Given these mixed findings, at which time do brain signals first reflect face attractiveness? One factor that complicates the interpretation of previous studies clearly is the variety in tasks (with some studies employing explicit attractiveness ratings and others employing orthogonal tasks), which makes it hard to appreciate whether

Department of Psychology, University of York, Heslington, York YO10 5DD, UK. ✉email: danielkaiser.net@gmail.com
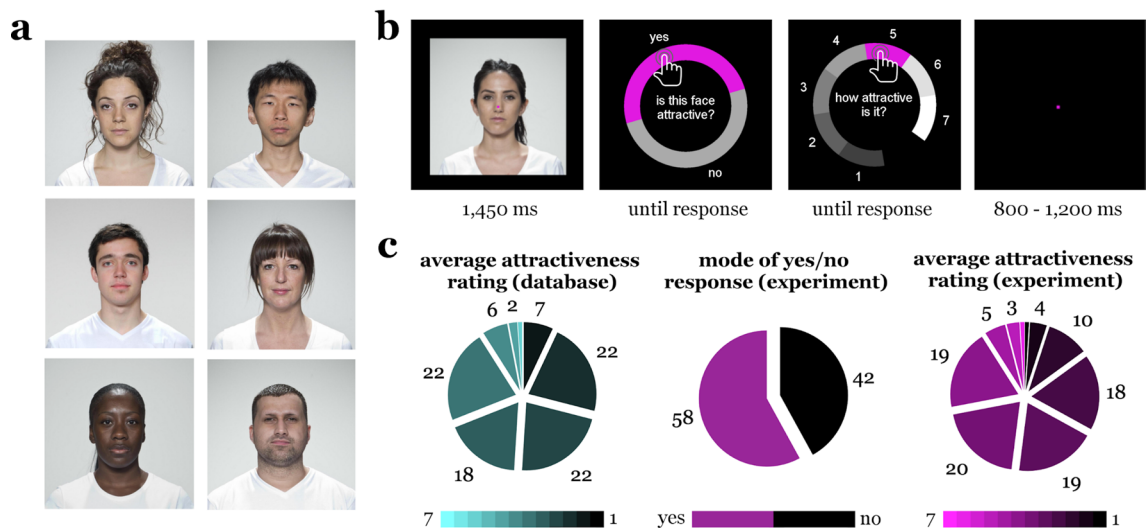
Figure 1. Stimuli and experimental approach. (**a**) Stimuli were full-front and neutral face photographs of 100 individuals, covering different sexes, ages, and ethnicities. Images are taken from Ref.[41] under a CC BY 4.0 license. (**b**) During the EEG experiment, participants viewed a single face on every trial. After seeing the face for 1450 ms, they were first asked to indicate whether they found the face attractive or not (yes/no response) and then asked to indicate how attractive they found the face on a 1–7 scale (attractiveness rating). Responses were given using the mouse. To avoid response-specific motor preparation, the positions of response options were differently arranged around a circle on every trial. (**c**) Average attractiveness ratings from the Face Research Lab London Set database (n = 2531), and yes/no responses and attractiveness ratings given by our participants (n = 23). Pie charts show histograms of rating responses (left and right charts) and the number of faces rated as attractive or unattractive by the majority of participants (middle chart). These data show that there was considerable variation in ratings across faces, with a comparable amount of faces being judged as attractive or unattractive in the current experiment. A detailed overview of participants' attractiveness judgments can be found in Supplementary Fig. S5.

ERP differences are due to variations in attractiveness or differences in task demands. Another problem is the variety in stimulus materials across studies (with different studies using faces of different genders, ages, and degrees of realism) and a lack of control for stimulus variability within individual studies, which makes it hard to assess whether ERP differences are genuinely related to differences in perceived attractiveness or other low- and high-level visual attributes of the face. Further, ERP studies can sometimes lack the sensitivity that multivariate analyses techniques offer for detecting subtle changes between conditions[38,39], thereby missing out on those temporal signatures that are weaker.

Here, we set out to resolve when cortical representations of facial attractiveness first emerge during an explicit attractiveness judgment task. By using multivariate representational similarity analysis (RSA[40]) on EEG data recorded during this task, we were able to temporally track the emergence of representations of facial attractiveness with high sensitivity, while at the same time being able to control for other sources of variability in the faces. We found that already between 150 and 200 ms of processing, brain representations reflected facial attractiveness. We uncover three key aspects of these early representations of facial attractiveness: first, we show that even such early cortical representations are partly explained by participants' idiosyncratic attractiveness ratings. Second, we demonstrate that cortical representations of facial attractiveness are not explained by other high-level face attributes, such as the person's sex, ethnicity, or age. Third, we use an artificial deep neural network (DNN) to show that cortical representations of facial attractiveness are not explained by visual features used for automated face recognition. Together, our results suggest that EEG signals carry genuine and individual representations of facial attractiveness, which emerge within the first 200 ms of cortical processing.

Results

We asked participants (n = 23) to rate the attractiveness of 100 male and female faces (see Fig. 1a for examples). Stimuli were highly controlled full-front face photographs taken from the Face Research Lab London Set[41], eliminating many sources of visual variability. On every trial, participants saw one of the faces for 1.450 ms and subsequently answered two questions (Fig. 1b). On the first question, they indicated whether they found the face attractive or not (hereinafter referred to as "yes/no response"). On the second question, they indicated how attractive they found the face on a 1–7 scale (hereinafter referred to as "attractiveness rating"). We compared participants' attractiveness ratings with ratings from a large group of observers (as provided with the Face Research Lab London Set; hereinafter referred to as "database rating"). Both the database ratings and responses collected during the experiment showed reasonable variance across the faces (Fig. 1c). However, individual-participant responses were only moderately correlated with the database ratings (r = 0.37 and r = 0.34 for the yes/no responses and attractiveness ratings, respectively), suggesting that there was substantial inter-individual variability in attractiveness judgments.
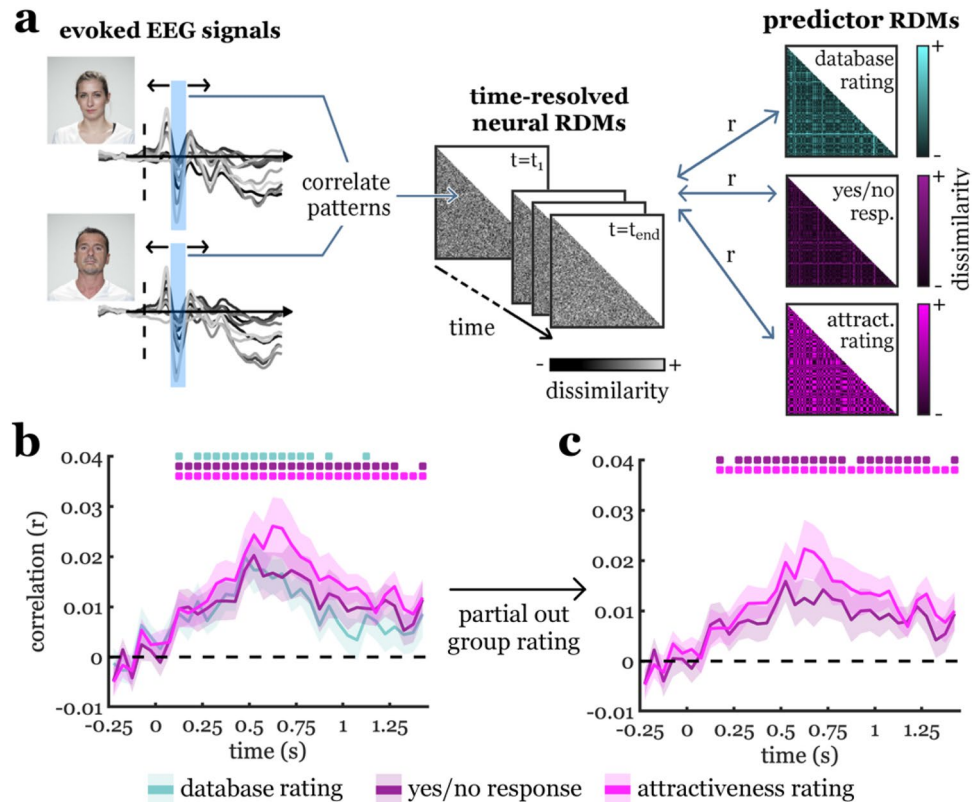
Figure 2. Analysis approach and key results. (**a**) Schematic description of the representational similarity analysis. EEG response patterns for each face were extracted separately for consecutive time bins of 50 ms (e.g., between 150 and 200 ms) across the epoch. Neural representational dissimilarity matrices (RDMs) were then constructed by correlating these response patterns (across time points and electrodes) for each pair of faces (for details, see "Materials and methods"). This yielded 100-by-100 RDMs whose entries indexed the pairwise neural dissimilarity between faces for each time bin. For each time bin separately, neural RDMs were correlated with the three predictor RDMs, which captured the faces' pairwise dissimilarity in attractiveness, based on (1) ratings from the Face Research Lab London Set database, (2) participants' individual yes/no responses, and (3) participants' individual rating responses. (**b**) All three predictor RDMs were significantly correlated with the neural RDMs, starting from 100 to 150 ms, suggesting an early neural representation of facial attractiveness. (**c**) When partialing out the average database ratings, we still found that participants' individual judgments predicted neural responses after 150–200 ms, suggesting that already during this early time window, responses partly reflect an idiosyncratic signature of facial attractiveness. Error margins represent standard errors of the mean. Significance markers denote $p < 0.05$ (corrected for multiple comparisons across time).

To measure how brain signals differed between faces of different perceived attractiveness, we recorded participants' brain activity using a 64-channel EEG system. We analyed EEG signals using time-resolved multivariate representational similarity analysis (RSA[40,42]). In RSA, the organization of brain representations is assessed by means of the pairwise (dis)similarity of each combination of stimuli, organized in neural representational dissimilarity matrices (RDMs). Neural RDMs were constructed separately for 34 discrete time bins (50 ms width) across the epoch, from 250 ms pre-stimulus to 1450 ms post-stimulus. For each time bin, response patterns (across all electrodes and across all time points within the 50 ms time bin) evoked by each one face were correlated with response patterns evoked by each other face. RDM entries were created by subtracting these correlations from 1, so that each RDM entry reflected how dissimilar a given face was represented from another face at a given time. One such 100-by-100 RDM was constructed for each time bin (Fig. 2a). Full details on the neural RDM construction can be found in the Materials and Methods section.

To investigate when cortical representations carried information about facial attractiveness, we modelled each participants' neural RDMs using three predictor RDMs (Fig. 2a): (1) a predictor RDM that reflected the faces' pairwise dissimilarity in the average attractiveness ratings taken from the Face Research Lab London Set database, (2) a predictor RDM that reflected the faces' pairwise dissimilarity in individual participants' yes/no responses given during the experiment, and (3) a predictor RDM that reflected the faces' pairwise dissimilarity in individual participants' attractiveness ratings given during the experiment. Full details on the predictor RDM construction can be found in the Materials and Methods section.

Early cortical representations of facial attractiveness. Correlating the predictor RDMs with the neural RDMs for each 50 ms time bin across the epoch, we obtained a timeseries of how well EEG response

patterns were predicted by facial attractiveness, as defined by the three predictors. Information about facial attractiveness was obtained for widespread temporal clusters across the EEG epoch and for all three types of predictors (Fig. 2b): database ratings predicted neural representations from the 100–150 ms time bin (peaking at 450–500 ms, peak t[22] = 5.08, p < 0.001, p_{corr} < 0.001, d = 1.06), individual yes/no responses predicted neural representations from the 100–150 ms time bin (peaking at 500–550 ms, peak t[22] = 4.37, p < 0.001, p_{corr} = 0.002, d = 0.91), and individual attractiveness ratings also predicted neural representations from the 100–150 ms time bin (peaking at 600–650 ms, peak t[22] = 4.53, p < 0.001, p_{corr} < 0.001, d = 0.94). As expected, the more fine-grained attractiveness ratings predicted variance in the neural data beyond the yes/no responses (see Supplementary Fig. S1a). Together, these results suggest a neural signature of facial attractiveness that already emerges during perceptual face processing.

Early representations of facial attractiveness reflect individual attractiveness judgments.   Are neural representations of facial attractiveness explicable by average preferences emerging across a large group of observers, potentially driven by a fixed set of physical face properties, with inter-individual variability reflecting merely noise? Or do they partly reflect idiosyncratic preferences, that is an individual person's unique aesthetic preference for particular faces? To resolve this question, we performed an analysis where we modelled neural RDMs as a function of individual yes/no responses and attractiveness ratings, while controlling for the database ratings using partial correlation analysis[43–45] (Fig. 2c). We found that individual attractiveness judgments still significantly predicted cortical representations, both when considering yes/no responses (from the 150–200 ms time bin; peaking at 500–550 ms, peak t[22] = 3.90, p < 0.001, p_{corr} = 0.012, d = 0.81) and when considering attractiveness ratings (from the 150–200 ms time bin; peaking at 600–650 ms, peak t[22] = 3.87, p < 0.001, p_{corr} = 0.002, d = 0.81). Similar results were obtained when partialing out the average judgments obtained across participants in the current experiment (see Supplementary Fig. S1b). Finding that already in the 150–200 ms time bin participants' individual judgments explained neural responses better than group-average ratings alone suggest that even early representations of facial attractiveness are linked to individual face preferences, rather than to physical features that similarly determine attractiveness for all observers. Further, to solidify that these early representations of facial attractiveness are indeed somewhat independent from particular physical face attributes, we performed two control analyses.

Early representations of facial attractiveness are not explained by other high-level face attributes.   In the first analysis, we tested whether other high-level face attributes (the depicted person's sex, ethnicity, or age) could explain the different cortical responses to faces of different attractiveness. To do so, we modelled neural RDMs using three predictor RDMs (Fig. 3a): (1) a predictor RDM that modelled the faces dissimilarity in sex, (2) a predictor RDM that modelled the faces dissimilarity in ethnicity, and (3) a predictor RDM that modelled the faces dissimilarity in age; for details, see Materials and Methods. Each of these high-level attributes predicted parts of the neural representation across time (Fig. 3b). Sex predicted face representations from the 100–150 ms time bin (peaking at 350–400 ms, peak t[22] = 5.31, p < 0.001, p_{corr} < 0.001, d = 1.11), ethnicity predicted representations from the 100–150 ms time bin (peaking at 100–150 ms, peak t[22] = 6.82, p < 0.001, p_{corr} < 0.001, d = 1.42), and age predicted representations from the 350–400 ms time bin (peaking at 500–550 ms, peak t[22] = 3.68, p < 0.001, p_{corr} = 0.007, d = 0.77). We then modelled neural representations as a function of individuals yes/no responses and attractiveness ratings, while controlling for all three high-level attributes using partial correlation analyses (Fig. 3c). We found that representations of facial attractiveness were not explained by the depicted persons' sex, ethnicity, and age: Database ratings still predicted neural representations from the 100–150 ms time bin (peaking at 450–500 ms, peak t[22] = 4.59, p < 0.001, p_{corr} = 0.001, d = 0.96), individual yes/no responses predicted representations from the 150–200 ms time bin (peaking at 500–550 ms, peak t[22] = 3.70, p < 0.001, p_{corr} = 0.009, d = 0.77), and individual attractiveness ratings predicted representations from the 150–200 ms time bin (peaking at 600–650 ms, peak t[22] = 4.37, p < 0.001, p_{corr} < 0.001, d = 0.91).

Early representations of facial attractiveness are not explained by deep neural network features.   In the second control analysis, we controlled for a variety of visual features analysed during face recognition by a deep neural network (DNN). DNNs are the current state-of-the-art for modelling visual representations emerging at different processing stages of biological vision[46–48]. Here, we used a DNN trained on recognizing faces (VGG-face[49]), which has previously been shown to accurately approximate features analysed during cortical face processing[50]. We extracted predictor RDMs from the 16 convolutional layers of the DNN, in which each entry reflected the pairwise dissimilarity (1-correlation) between the layer-specific activation vectors for two faces (Fig. 3d). These RDMs predicted cortical activations starting from the 50–100 ms time bin and, in line with previous reports[51–53], early representations (e.g., at 100–150 ms post-stimulus) were better captured by early DNN layers, while later representations were better captured by late DNN layers (Fig. 3e). Detailed results for all DNN layers can be found in Supplementary Fig. S2. Next, we again modelled neural representations as a function of individual yes/no responses and attractiveness ratings, but now controlling for features extracted in each of the 16 DNN layers (and additionally for the faces' sex, ethnicity, and age) using partial correlation analyses (Fig. 3f). Critically, we again found that database ratings (from the 100–150 ms time bin; peaking at 450–500 ms, peak t[22] = 4.93, p < 0.001, p_{corr} < 0.001, d = 1.03), yes/no responses (from the 100–150 ms time bin; peaking at 500–550 ms, peak t[22] = 3.54, p < 0.001, p_{corr} = 0.007, d = 0.74), and attractiveness ratings (from the 150–200 ms time bin; peaking at 600–650 ms, peak t[22] = 4.45, p < 0.001, p_{corr} < 0.001, d = 0.93) still predicted neural representations. Together, the two control analyses underscore the notion that early representations of facial attractiveness are somewhat independent from specific and fixed physical face properties.
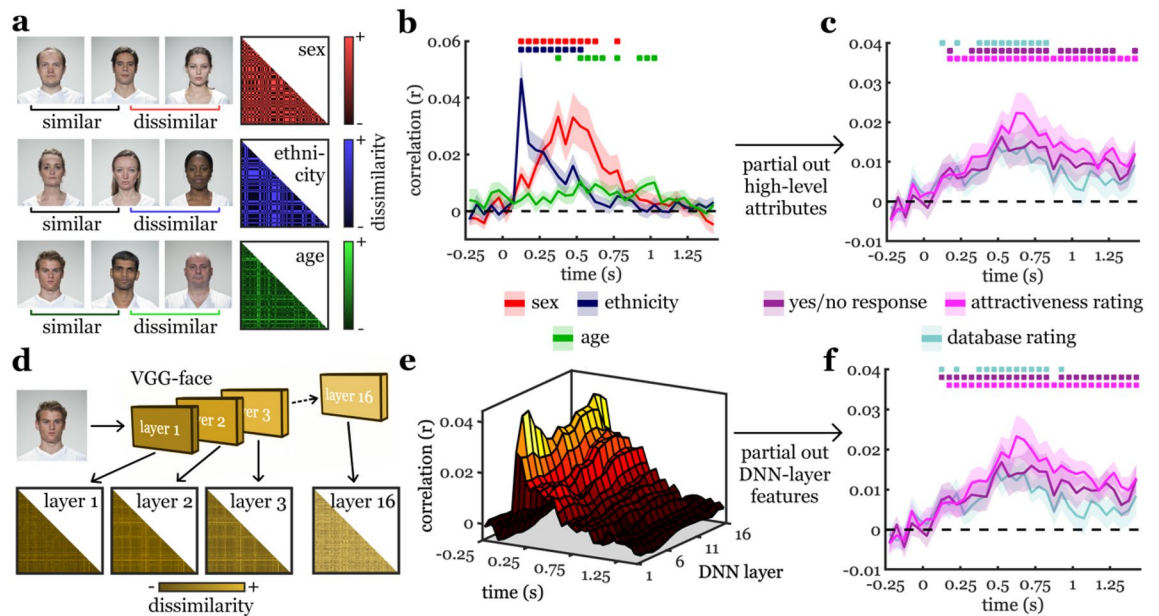
Figure 3. Controlling for high-level face attributes and DNN features. (**a**) To quantify high-level face attributes, we constructed RDMs based on the faces' dissimilarity in sex (same sex: similar, different sex: dissimilar), ethnicity (same ethnicity: similar, different ethnicity: dissimilar), and age (dissimilarity: absolute age difference between face pairs). (**b**) The three high-level attributes all explained significant proportions of the neural representation. (**c**) When controlling for the three attributes, participants' attractiveness judgments still predicted neural representations within 200 ms after onset. (**d**) To quantify DNN features, we extracted RDMs based on the dissimilarity of DNN activation patterns for each layer of a 16-layer DNN trained on face recognition (VGG-face; see "Materials and methods"). (**e**) DNN features at each layer predicted a significant proportion of the neural representation, with early layers better predicting early activations, and late layers better predicting later activations. For detailed results, see Supplementary Fig. S2. (**f**) When controlling for both the DNN features and the high-level face attributes, we again found that within 200 ms participants' attractiveness judgments still predicted neural representations. Error margins represent standard errors of the mean. Significance markers denote $p < 0.05$ (corrected for multiple comparisons across time).

In sum, our study reveals genuine and individually specific representations of face attractiveness. Across all analyses, these representations first emerged between 150 and 200 ms post-stimulus. This suggests that already during perceptual stages of face analysis, the brain computes how attractive we find a particular face.

Discussion

The current study used time-resolved representational similarity analysis on EEG data to track the dynamic emergence of cortical representations related to facial attractiveness. As the key result, we demonstrate that—across multiple analyses—robust neural representations of facial attractiveness emerge after 150–200 ms of vision. This timing suggests that differences in face attractiveness are related to processing differences in the N170 range[54–56]. Indeed, when comparing ERPs evoked by faces that were either rated as attractive or unattractive, we also found differences in N170 peak amplitudes, with stronger amplitudes for less attractive faces (see Supplementary Fig. S3). Our results thereby support earlier studies that have shown N170 modulations as a function of facial attractiveness[20–26]. Further, this timing of attractiveness-related responses is consistent with functional neuroimaging studies showing that facial attractiveness is represented in regions of the visual face processing network[57–61]. This suggests that face attractiveness is computed during perceptual processing, at the same time when basic facial configurations are analysed[55,56,62]. Our results thereby suggest that facial attractiveness is derived from perceptual face features, potentially related to favourable face configurations.

However, given that we found representations of facial attractiveness during perceptual processing, one could argue that they are not genuine representations of attractiveness, but representations of visual features that co-vary with attractiveness in our stimulus set. Our results offer two principal refutations of this argument. First, we show that early representations of facial attractiveness are neither explained by other high-level face attributes nor by features extracted by a DNN trained on face recognition. This shows that there is no straightforward mapping between the features used to determine facial attractiveness and features extracted for other categorization purposes. Second, we show that early representations of facial attractiveness are partly explained by participant's individual attractiveness judgments. This indicates that even such early processing of facial attractiveness is not strictly determined by the presence of particular physical face attributes: If representations of facial attractiveness were indeed a consequence of the presence or absence of a fixed set of visual features, one would expect that they are best predicted by the more stable average attractiveness rating across many observers. However, we find that representations of facial attractiveness are partly predicted by participants' individual attractiveness judgments, suggesting that they are not directly explained by the same physical properties for all observers.

Our results suggest that there are two types of attractiveness signals: First, there are those that can be explained by average attractiveness ratings across many observes (i.e., ratings obtained from the database). These signals show that there is a substantial amount of shared taste across participants, and that this amount of shared taste is sufficient to explain brain representations on an individual basis. As a downside, these signals are hard to discriminate from objective physical stimulus properties that may drive observers' shared taste. Second, there are signals that reflect personal taste (i.e., our participants' attractiveness ratings during the experiment), which are not explained by shared taste across individuals. Although such signals may offer evidence for inter-individual variability in attractiveness representations, their interpretation is not straightforward either: during an attractiveness judgment task like the one used here, differences between attractive and unattractive faces are not only related to differences in perceived attractiveness, but also to decision making processes.

In our experiment, the representation of such personal attractiveness preferences within the first 200 ms of processing suggests that differences in aesthetic appreciation are indeed related to perceptual computations differing between individuals. While differences in later brain representations can be attributed to task-related differences in cognitive and attentional processes as well as decision making, early perceptual representations are less sensitive to such processes[63,64]. Our findings therefore suggest that at the N170 processing stage individual features are weighted in idiosyncratic ways to give rise to an individual representation of facial attractiveness. This idea is in line with recent work showing that representations in object-selective visual cortex are best modelled by each person's idiosyncratic judgments about the objects' semantic similarities[65,66], showcasing that even perceptual brain representations can have an idiosyncratic component. Clearly, additional work needs to be done to map out this idiosyncratic component in facial attractiveness representations.

Beyond the early representation of facial attractiveness, we show that EEG responses also carried a sustained neural attractiveness signal. These sustained effects are in line with the later modulations of facial attractiveness stressed by some EEG studies[27–37]. These studies may not have found earlier differences related to facial attractiveness for varied reasons: Apart from the lower sensitivity offered by univariate ERP analyses, individual studies used largely orthogonal tasks, had small sample sizes, or did not specifically look for N170 differences. Such later EEG effects are consistent with fMRI activation differences in frontal regions as a function of facial attractiveness[57–61,67]. One possibility is that such late signals reflect neural representations related to decision making[68] and the maintenance of decisions in working memory[69]; when using an explicit attractiveness judgement task, these representations are hard to distinguish from genuine representations of face attractiveness. Alternatively, these late brain signals may reflect the cognitive processing of aesthetic quality: Some studies have suggested that such processes span frontal areas associated with stimulus valuation, such as the ventromedial prefrontal cortex and orbitofrontal cortex[61,67], while others have localized them to the default-mode network, potentially due to the self-referential character of aesthetic appreciation[70,71]. What is interesting is that such late processes seem to reflect aesthetic quality more generally[61,67,71], which can in principle arise from different visual inputs, such as human faces, visual scenes, and abstract stimuli, or even from non-visual inputs. Clearly, more research is needed to understand how neural representations of aesthetic quality are dynamically changing as they propagate across the brain, and which cortical networks they engage at different time points. By combining spatially and temporally resolved neural recordings, future studies could thereby tease apart the emergence of different perceptual, cognitive and emotional aspects of aesthetic experiences[72,73].

In sum, our study provides evidence for an early representation of facial attractiveness, that is both genuine (i.e., unrelated to other visual face attributes) and individually specific (i.e., partly explained by participants' personal attractiveness preferences). Finding that this representation emerged within 200 ms of vision provides a neural basis for rapid judgments of facial attractiveness in real-life contexts.

Materials and methods

Participants. Twenty-four adults (mean age 19.8, SD 1.6; 20 female) participated in the study. This sample size was comparable with the sample sizes of previous EEG studies on facial attractiveness[20–37], which tested a median of 20 participants, and an average of 24.6 participants. All participants had normal or corrected-to-normal vision. Participants were psychology students at the University of York and received course credits for participation. Prior to the experiment, participants provided written informed consent. All procedures were approved by the ethical committee of the Department of Psychology at the University of York and were in accordance with the Declaration of Helsinki. One participant was excluded because of technical problems that caused missing data, leaving 23 complete datasets for analysis.

Stimuli. The stimulus set consisted of 100 neutral full-front face photographs (for examples see Fig. 1a), taken from the Face Research Lab London Set[41]. These stimuli were colour photographs (1350-by-1350 pixels resolution). The standardized way of photographing limited variations in image features that are unrelated to differences in individual faces. No further adjustments were made to the faces obtained from the database. For each face, the stimulus set includes attractiveness ratings from a large set of observers (n = 2531), with considerable variance in attractiveness ratings between faces (see Fig. 1c). Each of the faces also comes with additional metadata, including the photographed person's self-reported gender, age, and ethnicity. An overview of the metadata for the 100 faces used in our Experiment is shown in Supplementary Fig. S4.

Experimental paradigm. During the EEG experiment, participants completed a single session of 700 trials, which was split into 7 blocks of 100 trials. On each trial, participants viewed one of the faces (7°-by-7° visual angle) for 1450 ms (Fig. 1b), on a uniform black background. Within each block, each face was shown exactly once, with trial order fully randomized within the block. After seeing the face, a blank screen was shown for 100 ms before participants were asked two questions. On the first question, they were asked to indicate whether

they found the face attractive or not by selecting yes or no. On the second question, they were asked how attractive they found the face by selecting a value between 1 (very unattractive) and 7 (very attractive). Answers to these questions were given using the computer mouse; participants could correct their answers as often as they wanted before proceeding by pressing the spacebar. To avoid response-specific motor preparation, the different response alternatives were placed at different positions around a circular response screen, as shown in Fig. 1b. Trials were separated by an inter-trial interval randomly varying between 800 ms and 1200 ms. Participants were instructed to keep central fixation during the inter-trial interval and the stimulus presentation; during these periods a pink fixation dot was overlaid in the centre of the screen. Eye-movements were not tracked during the experiment. Further, participants were asked to restrict eye blinks to the period when they selected their responses. The experiment was carried out in a dimly lit, quiet room. Stimuli were presented on a VIEWPixx display with a 1920-by-1020 resolution and stimulus presentation was controlled using the Psychtoolbox[74].

EEG recording and preprocessing. EEG signals were recorded using an ANT Waveguard 64-electrode system. Electrodes were arranged in accordance with the standard 10–10 system. EEG data were recorded at 250 Hz sampling rate using the ANT Neuroscan Software. Offline preprocessing was performed using FieldTrip[75]. EEG data were referenced to the Fz electrode (which was discarded after preprocessing), epoched from −500 ms to 1900 ms relative to stimulus onset, and baseline-corrected by subtracting the mean pre-stimulus signal for each electrode. A band-pass filter was applied to filter out 50 Hz line noise. Channels and trials containing excessive noise were removed based on visual inspection. On average, 5.1 channels (SD 2.2) and 73.4 trials (SD 30.3) were removed. Blinks and eye movement artifacts were removed using independent component analysis and visual inspection of the resulting components. After preprocessing, EEG epochs were cropped from −250 ms pre-stimulus to 1450 ms post-stimulus.

Measuring neural representational similarity. To track face representations across time, we used representational similarity analysis (RSA[40]). First, we created time-resolved neural representational dissimilarity matrices (RDMs), which reflected the pairwise dissimilarity of the faces' brain representations across processing time. Second, we compared the organization of the neural RDMs to a set of predictor RDMs, which captured different dimensions on which the faces' were similar or dissimilar.

Neural RDMs were constructed separately for each participant, using the CoSMoMVPA toolbox[76]. RDMs were created for 34 time bins of 50 ms width each, ranging from −250 to 1450 ms relative to stimulus onset. All following analyses were done separately for each of these 34 time bins. For each time bin, we extracted a response pattern across 12 time points (covering 50 ms at 250 Hz) and 63 electrodes (as during preprocessing electrodes were removed, electrode counts could be lower for individual participants). These data were then unfolded into a 756-element vector for further analyses. Before RDM construction, we performed principal-component analyses (PCAs) to reduce the dimensionality of these response vectors[39,77]. We split the available data into two independent subsets, with an equal number of trials per condition randomly assigned to each subset. The first subset of the data was used to perform the PCA decomposition. The PCA decomposition was then projected onto the second subset, retaining only the components needed to explain 99% of the variance in the first subset (99.1 components on average, SD across time: 13.2, SD across participants: 32.0). RDMs were constructed from the second subset. We first averaged across all available trials for each condition, and then correlated the response vectors for each pairwise combination of faces. These correlations were subtracted from 1 and entered into a 100-by-100 RDM. Each off-diagonal entry in this RDM thus reflected a measure of neural dissimilarity for a specific pair of faces; RDM diagonals were always empty. This procedure was then repeated with the two subsets swapped. Finally, the whole aforementioned analysis was repeated 50 times, with trials assigned randomly to the two subsets each time; RDMs were averaged across all repetitions, yielding a single RDM for each time bin.

Tracking neural representations of facial attractiveness. To characterize the representational organization obtained from the neural signals, we compared the neural RDMs to a set of predictor RDMs, separately for each time point and participant. Like the neural RDMs, each predictor RDM contained 100-by-100 entries, which reflected the dissimilarity of pairs of faces on a particular dimension.

To assess how strongly neural representations are determined by facial attractiveness, we created three attractiveness predictor RDMs: (1) An RDM based on individual participants' yes/no responses on the first question ("is this face attractive?"). For this RDM, pairwise matrix entries consisted of the absolute difference between yes/no-responses (coded as 1 and 2) given to the two faces. Note that participants could give different yes/no responses across repetitions of the same face. (2) An RDM based on individual participants' attractiveness ratings on question two ("how attractive is it?"). For this RDM, pairwise matrix entries consisted of the absolute difference between attractiveness ratings (on a 1–7 scale) given to the two faces. (3) An RDM based on the average attractiveness ratings (n = 2531) taken from the Face Research Lab London Set (see "Stimuli"). For this RDM, pairwise matrix entries also consisted of the absolute difference between attractiveness ratings (on a 1–7 scale) given to the two faces. Note that RDMs (1) and (2) were different for each participant, as they were based on their individual yes/no responses and attractiveness ratings, respectively, whereas RDM (3) was the same for every participant, reflecting a "ground truth" estimate of attractiveness ratings shared across a large group of people.

RDMs constructed from individual participants' yes/no responses were substantially correlated to RDMs constructed from participants' attractiveness ratings: correlations computed for each individual participant were r = 0.68 on average, and the correlation between the RDMs averaged across participants was r = 0.85. RDMs constructed from participants' individual responses in the experiment only showed a weaker correlation to the RDM constructed from the average group rating in the database (average r = 0.37 and r = 0.34 for the yes/no responses and attractiveness ratings, respectively). When averaging the RDMs across participants, we found a

much stronger correlation between RDMs reflecting the average individual-participant responses and the RDM reflecting the average ratings from the database (r = 0.80 and r = 0.73 for the yes/no responses and attractiveness ratings, respectively), suggesting that (1) there is a substantial amount of inter-individual variability in attractiveness ratings, and (2) average ratings in our experiment converged towards the average database ratings.

To quantify the correspondence between each of the predictor RDMs and the neural data, we computed Spearman-correlations between the neural RDMs and the predictor RDMs, separately for each participant; for these correlations only the lower off-diagonal elements of each RDM were used and the diagonal was always discarded. To establish correspondences between the neural RDM and a specific predictor RDM while controlling for other RDMs (see below), we used partial correlations[43–45]; these partial correlations are essentially correlations between the residual values after the control variables have been regressed out from both variables of interest. All correlations were Fisher-transformed before entering them into statistical analyses.

Controlling for high-level face attributes.   To test whether representations were uniquely attributable to face attractiveness, we explicitly controlled for a set of high-level face properties. Specifically, we controlled for a set of person attributes that may influence attractiveness ratings: a person's sex, ethnicity, and age. For each of these attributes we used the self-report metadata included in the Face Research Lab London Set (see "Stimuli"). From these data, we constructed three RDMs: (1) An RDM based on the depicted person's sex. For this RDM, pairwise matrix entries were marked as similar when both faces were of the same sex and as dissimilar when both faces were of different sexes. (2) An RDM based on the depicted person's ethnicity; as most of our participants (21/23) were Caucasian, ethnicity was binarized into Caucasian and non-Caucasian. For this RDM, pairwise matrix entries were marked as similar when both faces were of the same ethnicity and as dissimilar when both faces were of different ethnicities. (3) An RDM based on the depicted person's age. For this RDM, pairwise matrix entries reflected the absolute age difference between the two faces.

Controlling for deep neural network features.   To test whether neural representations related to face attractiveness were explicable by visual features typically extracted during face processing, we used a deep neural network (DNN) model trained on face recognition. DNNs are currently the state-of-the-art models for approximating the visual feature organization emerging during perception and have been shown to accurately approximate the organization of both low-level and high-level features in visual cortex[46–48,51–53]. Here, we used a 16-layer DNN for face recognition that was pre-trained on a huge face dataset (VGG-face[49]), as implemented in MatConvNet[78]. This DNN has previously been shown to mimic the organization of visual feature representations during cortical face processing[50]. To quantify the visual feature organization emerging in the DNN, we first ran the 100 face images through the DNN and obtained activation vectors for each of the 16 DNN convolutional DNN layers. We then constructed a model RDM for each layer by computing the representational distance (1-correlation) among each possible pair of faces.

Statistical testing.   We used one-sided t-tests against zero to identify significant correlations between the neural RDMs and predictor RDMs. False-discovery-rate (FDR) corrections were used to control for multiple comparisons across time. Exact test statistics and effect size estimates (Cohen's d) are reported for all peak effects.

Data availability
Stimuli and data are publicly available on OSF[79]. Other materials are available upon request.

References
1. Olson, I. R. & Marshuetz, C. Facial attractiveness is appraised in a glance. *Emotion* **5**, 498–502 (2005).
2. Riggio, R. E. & Woll, S. B. The role of nonverbal cues and physical attractiveness in the selection of dating partners. *J. Soc. Pers. Relat.* **1**, 347–357 (1984).
3. Walster, E., Aronson, V., Abrahams, D. & Rottman, L. Importance of physical attractiveness in dating behavior. *J. Pers. Soc. Psychol.* **4**, 508–516 (1966).
4. Benson, P. L., Karabenick, S. A. & Lerner, R. M. Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *J. Exp. Soc. Psychol.* **12**, 409–415 (1976).
5. Harrell, W. A. Physical attractiveness, self-disclosure, and helping behaviour. *J. Soc. Psychol.* **104**, 15–17 (1978).
6. Frieze, I. H., Olson, J. P. & Russell, J. Attractiveness and income for men and women in management. *J. Appl. Soc. Psychol.* **21**, 1039–1057 (1991).
7. Marlowe, C. M., Schneider, S. L. & Nelson, C. E. Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased?. *J. Appl. Psychol.* **81**, 11–21 (1996).
8. Thornhill, R. & Gangestad, S. W. Facial attractiveness. *Trends Cogn. Sci.* **3**, 452–460 (1999).
9. Little, A. C., Jones, B. C. & DeBruine, L. M. Facial attractiveness: Evolutionary based research. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **366**, 1638–1659 (2011).
10. Langlois, J. H. & Roggman, L. A. Attractive faces are only average. *Psychol. Sci.* **1**, 115–121 (1990).
11. Rhodes, G., Sumich, A. & Byatt, G. Are average facial configurations attractive only because of their symmetry?. *Psychol. Sci.* **10**, 52–58 (1999).
12. Grammer, K. & Thornhill, R. Human (homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *J. Comp. Psychol.* **108**, 233–242 (1994).
13. Rhodes, G., Proffitt, F., Grady, J. M. & Sumich, A. Facial symmetry and the perception of beauty. *Psychon. Bull. Rev.* **5**, 659–669 (1998).
14. Perrett, D. I. *et al.* Effects of sexual dimorphism on facial attractiveness. *Nature* **394**, 884–887 (1998).
15. Eisenthal, Y., Dror, G. & Ruppin, E. Facial attractiveness: Beauty and the machine. *Neural Comp.* **18**, 119–142 (2006).

16. Laurentini, A. & Bottino, A. Computer analysis of face beauty: A survey. *Comput. Vis. Image Und.* **125**, 184–199 (2014).
17. Germine, L. *et al.* Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Curr. Biol.* **25**, 2684–2689 (2015).
18. Hönekopp, J. Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 199–209 (2006).
19. Penton-Voak, I. & Perrett, D. I. Consistency and individual differences in facial attractiveness judgments: An evolutionary perspective. *Soc. Res.* **67**, 219–244 (2000).
20. Hahn, A. C. *et al.* Early and late event-related potentials are modulated by infant and adult faces of high and low attractiveness. *Soc. Neurosci.* **11**, 207–220 (2016).
21. Halit, H., de Haan, M. & Johnson, M. H. Modulation of event-related potentials by prototypical and atypical faces. *NeuroReport* **11**, 1871–1875 (2000).
22. Lu, Y., Wang, J., Wang, L., Wang, J. & Qin, J. Neural responses to cartoon facial attractiveness: An event-related potential study. *Neurosci. Bull.* **30**, 441–450 (2014).
23. Marzi, T. & Viggiano, M. P. When memory meets beauty: Insights from event-related potentials. *Biol. Psychol.* **84**, 192–205 (2010).
24. Schacht, A., Werheid, K. & Sommer, W. The appraisal of facial beauty is rapid but not mandatory. *Cogn. Affect. Behav. Neurosci.* **8**, 132–142 (2008).
25. Trujilo, L. T., Jankowitsch, J. M. & Langlois, J. H. Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. *Cogn. Affect. Behav. Neurosci.* **14**, 1061–1076 (2014).
26. Zhang, Z. & Deng, Z. Gender, facial attractiveness, and early and late event-related potential components. *J. Integr. Neurosci.* **11**, 477–487 (2013).
27. Chen, J. *et al.* Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neurosci. Lett.* **517**, 65–70 (2012).
28. Ma, Q., Hu, Y., Jiang, S. & Meng, L. The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: An ERP study. *Front. Neurosci.* **9**, 77 (2015).
29. Werheid, K., Schacht, A. & Sommer, W. Facial attractiveness modulates early and late event-related brain potentials. *Biol. Psychol.* **76**, 100–108 (2007).
30. Wiese, H., Altmann, C. S. & Schweinberger, S. R. Effects of attractiveness on face memory separated from distinctiveness: Evidence from event-related brain potentials. *Neuropsychologia* **56**, 26–36 (2014).
31. Calvo, M. H., Guierrez-Garcia, A. & Beltran, D. Neural time course and brain sources of facial attractiveness vs. trustworthiness judgment. *Cogn. Affect. Behav. Neurosci.* **18**, 1233–1247 (2018).
32. Carbon, C. C., Faerber, S. J., Augustin, M. D., Mitterer, B. & Hutzler, F. First gender, then attractiveness: Indications of gender-specific attractiveness processing via ERP onsets. *Neurosci. Lett.* **686**, 186–192 (2018).
33. Johnston, V. S. & Oliver-Rodriguez, J. C. Facial beauty and the late positive component of event-related potentials. *J. Sex. Res.* **34**, 188–198 (1997).
34. Oliver-Rodriguez, J. C., Guan, Z. & Johnston, V. S. Gender differences in late positive components evoked by human faces. *Psychophysiology* **36**, 176–185 (1999).
35. Roye, A., Höfel, L. & Jacobsen, T. Aesthetics of faces: Behavioural and electrophysiological indices of evaluative and descriptive judgment processes. *J. Psychophysiol.* **22**, 41–57 (2008).
36. Van Hooff, J. C., Crawford, H. & Van Vugt, M. The wandering mind of men: ERP evidence for gender differences in attention bias towards attractive opposite sex faces. *Soc. Cogn. Affect. Neurosci.* **6**, 477–485 (2011).
37. Zhang, Y. *et al.* Identifying cognitive preferences for attractive female faces: An event-related potential experiment using a study-test paradigm. *J. Neurosci. Res.* **89**, 1887–1893 (2011).
38. Contini, E. W., Wardle, S. G. & Carlson, T. A. Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia* **105**, 165–176 (2017).
39. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding dynamic patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* **29**, 677–697 (2017).
40. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
41. https://doi.org/10.6084/m9.figshare.5047666.v3
42. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
43. Ambrus, G. G., Kaiser, D., Cichy, R. M. & Kovács, G. The neural dynamics of familiar face recognition. *Cereb. Cortex* **29**, 4775–4785 (2019).
44. Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J. F. & Charest, I. The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. *Neuroimage* **194**, 12–24 (2019).
45. 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**, e32962 (2018).
46. Cichy, R. M. & Kaiser, D. Deep neural networks as scientific models. *Trends Cogn. Sci.* **23**, 305–317 (2019).
47. Kriegeskorte, N. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015).
48. Kietzmann, T. C., McClure, P. & Kriegeskorte, N. Deep neural networks in computational neuroscience. In *Oxford Research Encyclopedia of Neuroscience* (Oxford University Press, Oxford, 2019).
49. Parkhi, O. M., Vedaldi, A. & Zisserman, A. Deep face recognition. In *Proceedings of the British Machine Vision Conference (BMVC)* 41.1–41.12 (BMVA Press, 2015).
50. Grossman, S. *et al.* Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nat. Commun.* **10**, 4934 (2019).
51. 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**, 27755 (2016).
52. Güçlü, U. & van Gerven, M. A. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* **35**, 10005–10014 (2015).
53. Yamins, D. L. *et al.* Performance-optimized hierarchical models predict responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8619–8624 (2014).
54. Bentin, S., Allison, T., Puce, A., Perez, E. & McCarthy, G. Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* **8**, 551–565 (1996).
55. Rossion, B. & Jacques, C. The N170: Understanding the time course of face perception in the human brain. In *Oxford Library of Psychology. The Oxford Handbook of Event-Related Potential Components* (Oxford University Press, Oxford, 2012).
56. Eimer, M. (2011). The face-sensitive N170 component of the event-related brain potential. In *The Oxford Handbook of Face Perception* (Oxford University Press, Oxford, 2011).
57. Chatterjee, A., Thomas, A., Smith, S. E. & Aguirre, G. K. The neural response to facial attractiveness. *Neuropsychology* **23**, 135–143 (2009).
58. Iaria, G., Fox, C. J., Waite, C. T., Aharon, I. & Barton, J. J. S. The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: Neuropsychological and neuroimaging evidence. *Neuroscience* **155**, 409–422 (2008).

59. Kocsor, F., Feldmann, A., Bereczkei, T. & Kallai, J. Assessing facial attractiveness: Individual decisions and evolutionary constraints. *Socioaffect. Neurosci. Psychol.* **3**, 21432 (2013).
60. Pegors, T. K., Kable, J. W., Chatterjee, A. & Epstein, R. A. Common and unique representations in pFC for face and place attractiveness. *J. Cogn. Neurosci.* **27**, 959–973 (2015).
61. Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I. & Dolan, R. J. Brain systems for assessing facial attractiveness. *Neuropsychologia* **45**, 195–206 (2007).
62. Rossion, B. *et al.* Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biol. Psychol.* **50**, 173–189 (1999).
63. Cauquil, A. S., Edmonds, G. E. & Taylor, M. J. Is the face-sensitive N170 the only ERP not affected by selective attention?. *NeuroReport* **11**, 2167–2171 (2000).
64. Furey, M. L. *et al.* Dissociation of face-selective cortical responses by attention. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1065–1070 (2006).
65. Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14565–14570 (2014).
66. Charest, I. & Kriegeskorte, N. The brain of the beholder: Honoring individual representational idiosyncrasies. *Lang. Cogn. Neurosci.* **30**, 367–379 (2015).
67. Kühn, S. & Gallinat, J. The neural correlates of subjective pleasantness. *Neuroimage* **61**, 289–294 (2012).
68. Filimon, F., Philiastides, M. G., Nelson, J. D., Kloosterman, N. A. & Heekeren, H. R. How embodied is perceptual decision making? Evidence for separate processing of perceptual and motor decisions. *J. Neurosci.* **33**, 2121–2136 (2013).
69. Riley, M. R. & Constantinidis, C. Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci.* **9**, 181 (2016).
70. Vessel, E. A., Starr, G. G. & Rubin, N. Art reaches within: Aesthetic experience, the self and the default mode network. *Front. Neurosci.* **7**, 258 (2013).
71. Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L. & Starr, G. G. The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19155–19164 (2019).
72. Leder, H., Belke, B., Oeberst, A. & Augustin, D. A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* **95**, 489–508 (2004).
73. Markovic, S. Components of aesthetic experience: Aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *i-Perception* **3**, 1–17 (2012).
74. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
75. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
76. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinform.* **10**, 20 (2016).
77. Kaiser, D., Häberle, G. & Cichy, R. M. Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *J. Neurophysiol.* **124**, 145–151 (2020).
78. Vedaldi, A. & Lenc, K. MatConvNet—convolutional neural networks for Matlab. In *Proceedings of the ACM International Conference on Multimedia* (2015).
79. https://doi.org/10.17605/OSF.IO/MWQ6T

Acknowledgements

Author contributions

D.K. designed the study. D.K. and K.N. acquired data. D.K. and K.N. analyzed data. D.K. interpreted results. D.K. prepared figures. D.K. drafted the manuscript. D.K. and K.N. edited and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-74009-9.

Correspondence and requests for materials should be addressed to D.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.