Characterizing Dynamic Neural Representations of Scene Attractiveness

Daniel Kaiser

Abstract

■ Aesthetic experiences during natural vision are varied: They can arise from viewing scenic landscapes, interesting architecture, or attractive people. Recent research in the field of neuroaesthetics has taught us a lot about where in the brain such aesthetic experiences are represented. Much less is known about when such experiences arise during the cortical processing cascade. Particularly, the dynamic neural representation of perceived attractiveness for rich natural scenes is not well understood. Here, I present data from an EEG experiment, in which participants provided attractiveness judgments for a set of diverse natural scenes. Using multivariate pattern analysis, I demonstrate that scene attractiveness is mirrored in early brain signals that arise within 200 msec of vision, suggesting that the aesthetic appeal of scenes is first resolved during perceptual processing. In more detailed analyses, I show that even such early neural correlates of scene attractiveness are partly related to interindividual variation in aesthetic preferences and that they generalize across scene contents. Together, these results characterize the time-resolved neural dynamics that give rise to aesthetic experiences in complex natural environments. ■

INTRODUCTION

In our daily lives, aesthetic experiences can arise from a variety of different visual contents, such as from a scenic sunset, impressive historical architecture, or an attractive face. In the field of neuroaesthetics, one of the key questions is how the aesthetic appeal of such diverse experiences is dynamically extracted by the brain (Skov & Nadal, 2020; Pearce et al., 2016).

Over the recent decades, we have learned a lot about where in the brain aesthetic experiences are processed. Researchers have used fMRI to pinpoint the neural correlates of perceived aesthetic appeal for a range of stimuli, such as faces, landscapes, abstract patterns, and artworks (Isik & Vessel, 2021; Zhao et al., 2020; Vessel, Isik, Belfi, Stahl, & Starr, 2019; Pegors, Kable, Chatterjee, & Epstein, 2015; Winston, O'Doherty, Kilner, Perrett, & Dolan, 2007; Yue, Vessel, & Biederman, 2007; Jacobsen, Schubotz, Höfel, & Cramon, 2006; Kawabata & Zeki, 2004). Some of these studies also demonstrate that the aesthetic appeal of vastly different visual contents, such as faces and scenes (Pegors et al., 2015) or artworks and photographs (Vessel et al., 2019), is represented in similar regions across cortex, including parts of the visual cortex and the frontal cortex, as well as the default mode network. These studies suggest that aesthetic experiences can not only arise from seemingly dissimilar visual stimuli, but that these experiences yield similar cortical correlates.

Much less is known about when aesthetic experiences arise dynamically across the cortical processing cascade.

Much of the M/EEG literature addressing this question has focused on the perception of face attractiveness (Kaiser & Nyga, 2020; Carbon, Faerber, Augustin, Mitterer, & Hutzler, 2018; Zhang & Deng, 2012; Schacht, Werheid, & Sommer, 2008; Werheid, Schacht, & Sommer, 2007), with many studies highlighting that a face's attractiveness can impact early and fundamental stages of the face processing hierarchy. In our own work, we have further highlighted that such early representations of face attractiveness are partly explained by personal preferences, rather than only by attractiveness judgments that are shared among a large group of observers (Kaiser & Nyga, 2020), suggesting that even the perceptual correlates of attractiveness are shaped in personally idiosyncratic ways. Unlike the rich literature on face attractiveness, only few studies have looked at the time-resolved neural correlates of aesthetic judgments for other stimuli, such as abstract patterns (Höfel & Jacobsen, 2007; Jacobsen & Höfel, 2003) and various types of artworks (Strijbosch et al., 2022; de Tommaso et al., 2008; Cela-Conde et al., 2004). However, we currently do now know how the brain dynamically represents attractiveness for the natural scenes we typically experience during our everyday lives.

There are three open questions about the neural dynamics underlying perceived scene attractiveness: (1) Is the aesthetic appeal of natural scenes already represented during early stages of cortical processing, indicating that differences in fundamental perceptual processes are involved in the extraction of scene attractiveness? (2) Are judgments of scene attractiveness based on a general consensus of what is aesthetically pleasing or are they partly shaped by personal aesthetic preferences? (3) Is

Justus-Liebig-University Gießen, Germany

the aesthetic appeal of natural scenes extracted similarly across different scene contents?

To answer these questions, I investigated how the perceived aesthetic appeal of a set of 100 diverse natural scene images is mirrored in temporally resolved EEG signals. For this, I use a multivariate pattern analysis framework that we recently developed in an EEG study of face attractiveness (Kaiser & Nyga, 2020). Specifically, I modeled how the pairwise similarity between scenes in the EEG signals over time was related to the similarity in aesthetic appeal ratings. The resulting data demonstrate that (1) the aesthetic appeal of scenes is encoded in early brain signals emerging within 200 msec after scene presentation, (2) part of this early neural correlate of scene attractiveness is explained by interindividual variation in aesthetic perception, and (3) even early neural correlates of perceived aesthetic appeal generalize across a variety of scene contents.

METHODS

Participants

Twenty-four healthy adults took part (mean age 19.6 years, SD = 1.7; 21 women). This sample size was identical to our previous EEG study investigating face attractiveness (Kaiser & Nyga, 2020). One participant was excluded because of a technical error in the recordings, leaving a final sample of 23 participants. Participants received course credits. All participants provided written informed consent. Procedures were approved by the ethics committee of the Department of Psychology, University of York, and were in accordance with the Declaration of Helsinki.

Stimuli

The stimulus set consisted of 100 natural scene photographs. All stimuli were resized to 600 × 400 pixels. The stimuli were chosen from the top- and bottom-rated images contained in the AVA (Murray, Marchesotti, & Perronnin, 2012) and photo.net (Datta, Li, & Wang, 2008) aesthetic rating databases. To match the overall visual content across images that were expected to yield high or low aesthetic appeal ratings, stimuli were arranged into pairs (Figure 1A): In each pair, the two images depicted visually and/or conceptually similar contents (e.g., a group of people or an indoor room). To check whether there were any consistent visual differences across pairs, I computed GIST descriptors (Oliva & Torralba, 2001) for all images. I then checked whether the high attractiveness and low attractiveness images within each pair were reliably different. To this end, I trained a nearest-neighbor classifier (as implemented in CoSMoMVPA; Oosterhof, Connolly, & Haxby, 2016) to classify the images' attractiveness into high and low. This classifier was trained on the GIST descriptors for all but one pairs and then tested on the remaining pair of images (this procedure was repeated for each pair being left out). The GIST-based classifier could not successfully discriminate between the images that were taken from the top- and bottom-rated images in the database (49% correct). Similar results were obtained using the "pooling" layers of an HMAX (Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007; Riesenhuber & Poggio, 1999) filter-bank model (layer C1: 45% correct; layer C2: 49% correct). The pairing of stimuli thus ensured that there were no pronounced visual differences in basic visual characteristics between images of different aesthetic appeal (as indicated by

Figure 1. Stimuli and paradigm. (A) The stimulus set consisted of 100 natural scene photographs depicting a variety of diverse contents. Stimuli were arranged into visually similar pairs of high and low aesthetic appeal (based on visual aesthetics databases) to reduce the visual variability between scenes that were likely judged as relative attractive or unattractive. (B) During each trial of the EEG experiment, participants viewed a single scene and subsequently rated its attractiveness on two complementary response screens asking whether the image was attractive ("yes/no response") and how attractive it was on a 1–7 scale ("rating response"). Response options were arranged randomly around a circular response screen to prevent motor preparation during the image presentation. Trials were separated by an 800- to 1200-msec intertrial interval.
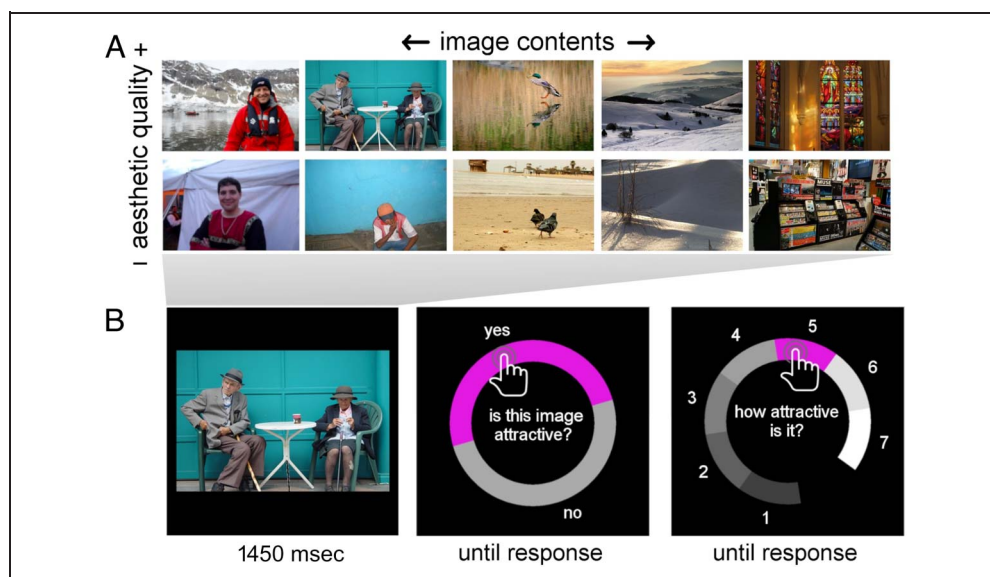
their database rankings). In all subsequent analyses, I focused on the ratings provided by participants in the current study.

Paradigm

The experimental paradigm and analysis approach was largely identical to our recent EEG study on face attractiveness perception (Kaiser & Nyga, 2020). Participants viewed a single scene on every trial (8° × 5.3° visual angle), which was presented for 1450 msec on a uniform black background (Figure 1B). After a 100-msec blank screen, participants were asked to give two complementary responses: First, they were asked to provide a binary attractiveness response, indicating whether they found the scene attractive or not (hereinafter referred to as "yes/no response"). Second, they were asked to provide a more fine-grained rating, indicating how attractive they found the scene on a 1–7 rating scale (hereinafter referred to as "rating response"). Both responses were given with the mouse and were nonspeeded. Participants were instructed to indicate how aesthetically pleasing, or attractive, or beautiful they rated each image (without any distinction between these concepts). To prevent participants from preparing a motor response, the response options for both responses were presented at random angular positions across a circular response screen (Figure 1B). Participants were further instructed to keep central fixation on a pink fixation dot during the scene presentation and to restrict eye blinks to the period when they selected their responses. Trials were separated by an intertrial interval randomly varying between 800 and 1200 msec. The experiment consisted of seven blocks, in each of which each image was shown once, in random order. Each image was therefore repeated 7 times, yielding 700 trials in total. The experiment was carried out in a dimly lit and quiet room. Stimuli were presented on a VIEWPixx display with a 1920 × 1020 resolution, and stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997).

EEG Acquisition and Preprocessing

EEG signals were recorded using an ANT Waveguard 64-electrode system and a TMSi REFA amplifier. Electrodes were arranged in accordance with the standard 10–10 system. EEG data were recorded at 250-Hz sampling rate using the ANT Neuroscan Sofware. Off-line preprocessing was performed using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). EEG data were referenced to the Fz electrode (which was discarded after preprocessing), epoched from −500 to 1900 msec relative to stimulus onset, and baseline-corrected by subtracting the mean prestimulus signal for each electrode. A band-pass filter was applied to remove 50-Hz line noise. Channels and trials containing excessive noise were removed based on visual inspection. On average, nine channels (SE = 0.6) and 69 trials (SE = 11) 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 msec prestimulus to 1450 msec poststimulus.

Extracting Neural Representational Similarity

To track representations across time, I used representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008). First, neural representational dissimilarity matrices (RDMs) were constructed separately for each participant, using the CoSMoMVPA toolbox (Oosterhof et al., 2016). RDMs were created for 34 consecutive time bins of 50 msec width, from −250 to 1450 msec relative to scene onset. The following analyses were done separately for each time bin. At each bin, response patterns were extracted across 12 time points (covering 50 msec at 250 Hz) and 63 electrodes (after preprocessing, electrode counts could be lower for individual participants). These data were then unfolded into a 756-element vector. Before RDM construction, I performed principal component analyses (PCAs) to reduce the dimensionality of the response vectors (Kaiser, Häberle, & Cichy, 2020; Grootswagers, Wardle, & Carlson, 2017). I 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 (97 components on average, SD across time: 19, SD across participants: 16). RDMs were constructed from the second subset. I first averaged across all available trials for each condition and then correlated (Spearman correlations) the response vectors for each pairwise combination of scenes. These correlations were subtracted from 1 and arranged into a 100 × 100 RDM. Each entry in this RDM reflected a measure of neural dissimilarity for a specific pair of scenes. RDM diagonals were always empty. This procedure was then repeated with the two subsets swapped. Finally, the whole 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.

Modeling Representational Similarity

Neural RDMs were modeled by a set of predictor RDMs that also spanned 100 × 100 entries and captured the scenes' pairwise similarities on a set of candidate properties. Correspondence between the predictor RDMs and the neural RDMs was assessed by correlating (Spearman correlations) all lower off-diagonal entries between the two types of RDMs, separately for each time point at which a neural RDM was available, yielding a time course of correspondence. I performed three types of RSA, using three types of predictor RDMs: First, to model how well neural

representations were predicted by attractiveness judgments, I created two predictor RDMs that captured the scenes' similarities in the yes/no responses and the rating responses. In these RDMs, each entry was computed by taking the absolute difference between the average yes/no responses (coded as 2: yes and 1: no) or attractiveness ratings between two scene images. These RDMs were constructed separately for each participant. Second, I tested whether the complex feature organization emerging in a deep neural network (DNN) model of categorization explains the correspondence between attractiveness judgments and neural representation. Although the stimulus set was matched in approximate content and low-level features across the more or less attractive images, the images may still differ in a set of complex visual features that is routinely read out during visual categorization: To quantify the extraction of such features, I employed a DNN as a model of the cortical processing hierarchy (Cichy & Kaiser, 2019; Kriegeskorte, 2015). I extracted features along the 16 layers of a VGG16 DNN (Simonyan & Zisserman, 2014) trained on ImageNet (Deng et al., 2009). For each layer, I then constructed an RDM by computing 1 minus the correlation of activation vectors for each pair of scenes. RDMs were then averaged within six layers blocks along the network (five blocks of convolutional layers and one block of fully connected layers; see Figure 4A). These RDMs were then correlated with the neural RDMs to determine how well the DNN predicted the neural data in the first place. After that, I conducted partial correlation analyses in which I reperformed the correlation between the aesthetic judgment RDMs and the neural RDMs while partialing out the RDMs for each DNN layer block. Third, to model how well average and individual attractiveness judgments predict the neural data, I created RDMs that captured the scenes' similarities in the average yes/no responses and the rating responses across a set of participants. These were created in the same way as the previous RDMs, but for each participant, I created an average RDM that was based on the average yes/no responses or rating responses from all other participants. In a partial correlation analysis, I then again correlated the predictor RDMs for the yes/no responses and ratings responses with the neural RDMs, but now partialing out the average response across all other participants. This yielded an estimate of how well individual responses account for cortical representations when average responses are controlled for (Kaiser & Nyga, 2020). Fourth, to test whether the neural correlates of scene attractiveness generalize across scene categories, I performed analyses in which the correspondence between the predictor RDMs for the yes/no responses or rating responses and the neural RDMs was only assessed for those pairwise comparisons in which the two scenes came from opposite categories. This analyses thus reveals whether attractive scenes from one category are represented similarly as attractive scenes from another category (and dissimilarly from less attractive scenes from another category), thus revealing a

relatively content-independent representation of aesthetic appeal. I tested such generalization for three categorical distinctions: scenes containing ($n = 42$) versus not containing ($n = 58$) people, natural ($n = 58$) versus man-made ($n = 42$) scenes, and scenes containing ($n = 38$) versus not containing ($n = 62$) prominent foreground objects. Scenes were assigned to these labels by the author.

Statistical Testing

Correlations between neural and predictor RDMs were compared with zero using one-sample t tests (one-sided against zero) for each time bin. The resulting p values were corrected for multiple comparisons using false discovery rate corrections. Test statistics (t values) and effect sizes (Cohen's d) are provided for all peak effects.

RESULTS

Ratings of Scene Attractiveness

The behavioral responses given during the experiment confirmed the pairing of images in our stimulus set, with stimuli obtained from the top-rated database images yielding consistently higher perceived aesthetic appeal than those obtained from the bottom-rated database images. Indeed, the expected difference within each pair was consistent across all pairs and both for the yes/no responses (Figure 2A) and the rating responses (Figure 2B). Across participants, the top-rated images yielded significantly higher yes/no responses, $t(22) = 12.6, p < .001$, and rating responses, $t(22) = 11.7, p < .001$, than the bottom-rated images.

Neural Dynamics of Perceived Scene Attractiveness

To track the emergence of neural responses related to scene attractiveness, I correlated neural RDMs with
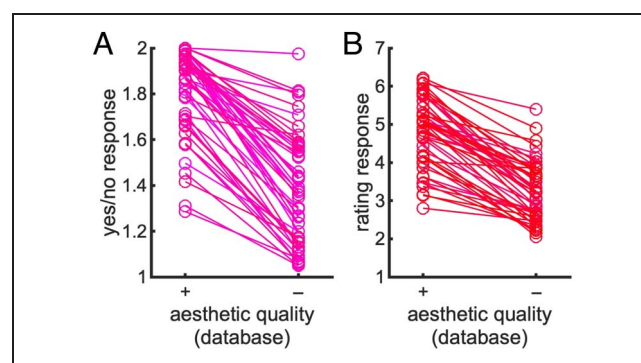

Figure 2. Analysis of attractiveness judgments. (A) Average yes/no responses across stimulus pairs (2:yes, 1:no; each line represents a pair). (B) Average rating responses across stimulus pairs (7:best, 1: worst; each line represents a pair). For both judgments and within each pair, the scene that had the higher attractiveness rating in the aesthetics databases also received the higher yes/no response and rating response in the experiment.
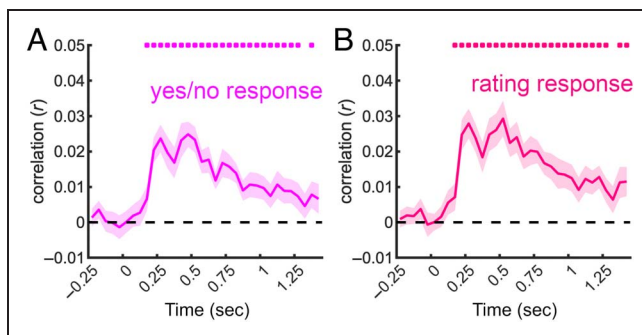
Figure 3. Neural dynamics of perceived scene attractiveness. (A) Average correlations between the neural RDMs (reflecting pairwise similarities between scenes at each time point) and a predictor RDM based on participants' yes/no responses (reflecting pairwise similarities in how similarly scenes were judged as attractive or unattractive). (B) Average correlations between the neural RDMs and a predictor RDM based on rating responses (reflecting pairwise similarities in how similarly scenes were rated). Both types of attractiveness judgments predicted neural responses from the 150- to 200-msec time bin and in a temporally sustained fashion. Error margins represent standard errors of the mean and significance markers denote $p$corr < .05.

predictor RDMs that captured the images' pairwise similarities in participants' yes/no and rating responses. Yes/no responses (Figure 3A) significantly predicted cortical responses from the 150 to 200 msec time bin (peaking at 450–500 msec, peak, $t(22) = 6.94$, $p$corr < .001, $d = 1.45$). Rating responses (Figure 3B) also predicted cortical responses from the 150- to 200-msec time bin (peaking at 500–550 msec, peak, $t(22) = 5.80$, $p$corr < .001, $d = 1.21$). This shows that the aesthetic appeal of natural scenes is rapidly and sustainably contained in EEG signals, suggesting that aesthetic experiences manifest during early stages of visual information processing and are encoded in brain responses for extended periods.

Controlling for DNN Features

Although the more and less attractive images in the stimulus set were matched for their approximate visual content and for low-level features, the correspondence between aesthetic judgments and neural representations may still be accounted for when more complex visual features are considered. To test this possibility, I compared neural representations to the representations emerging in a DNN model of visual categorization (VGG16; Simonyan & Zisserman, 2014; see Figure 4A). When correlating RDMs extracted from six layer blocks along this DNN with the neural RDMs, I found strong signatures of hierarchical correspondence (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016), with early layers predominantly predicting early brain activations, and late layers more strongly predicting later activations (Figure 4B). This result shows that the DNN was able to approximate visual feature processing in cortex. To test whether such categorization-related feature processing could account for the correspondence between attractiveness judgments and neural

representations, I correlated the RDMs constructed from the yes/no responses and rating responses with the neural RDMs while partialing out the RDMs obtained from the DNN layer blocks. Across the board and for both types of responses, correlations remained significant (Figure 4C–D), indicating that the features extracted during categorization (as approximated by the DNN) are not sufficient for explaining the cortical representation of scene attractiveness. Partialing out the RDMs from the last layer block (i.e., the fully connected layers of VGG16) produced the greatest drop in correlations, with a significant reduction compared with the original results, both for the yes/no responses and the rating responses (from 50 to 100 msec and across the whole epoch). However, correlations remained well above chance, both for the yes/no responses (from the 200 to 250 msec time bin, peaking at 450–500 msec, peak, $t(22) = 5.69$, $p$corr < .001, $d = 1.19$) and the rating responses (from the 200 to 250 msec time bin, peaking at 500–550 msec, peak, $t(22) = 4.90$, $p$corr < .001, $d = 1.02$). This indicates that aesthetic appeal is not solely a reflection of features that are read out during
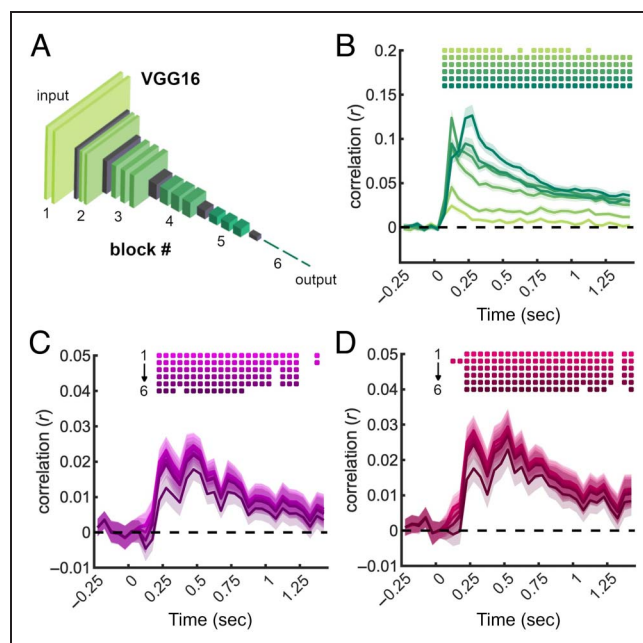

Figure 4. Controlling for DNN features. (A) To approximate the extraction of categorization-related features in the visual system, I used a VGG16 DNN trained on ImageNet. From this network, I extracted RDMs for six layer blocks along the network, as illustrated; the last three layers are the network's fully connected layers. (B) The RDMs extracted from the DNN predicted neural activations well, with early layers predominantly predicting relatively early activations and late layers offering more accurate prediction of later activations. (C) Reperforming the correlations between the RDMs based on yes/no attractiveness responses and the neural RDMs (as in Figure 3A) while partialing out the DNN RDMs obtained from different layer blocks showed that DNN features at different depths cannot account for the correspondence between attractiveness judgments and brain representations. (D) Similar results were obtained for the rating responses. Error margins represent standard errors of the mean, and significance markers denote $p$corr < .05.

categorization, but that qualitatively different features determine the attractiveness of natural visual inputs.

Personal versus Shared Aesthetic Judgments

Next, I investigated whether personal ratings of aesthetic appeal predict cortical responses beyond the average ratings provided by a group of observers (Kaiser & Nyga, 2020). I thus correlated the neural RDMs with additional predictor RDMs that for each participant were constructed from the average responses of all other participants in the experiment. For the yes/no responses (Figure 5A), averaged responses from all other participants predicted neural responses and participants' own responses, starting from the 200- to 250-msec time bin (peaking at 250–300 msec, peak, $t(22) = 6.47$, $p$corr $< .001$, $d = 1.35$). A similar pattern emerged for the rating responses (Figure 5B), starting from the 200- to 250-msec time bin (peaking at 250–300 msec, peak, $t(22) = 6.76$, $p$corr $< .001$, $d = 1.41$). For both measures, there were no significant differences between participants' own ratings and the average of others' ratings (all $p$corr $> .93$). Given this similarity, I next tested whether other people's ratings explain the correspondence between participants' own ratings and the neural responses. I thus performed a partial correlation analysis, correlating the neural RDMs with the RDMs constructed from participants' own responses, partialing out the average responses from all other participants. This analysis still revealed a significant correlation, both for the yes/no responses (from the 200 to 250 msec time bin, peaking at 400–450 msec, peak, $t(22) = 3.97$, $p$corr $= .004$, $d = 0.83$) and the rating responses (from the 150–200 msec
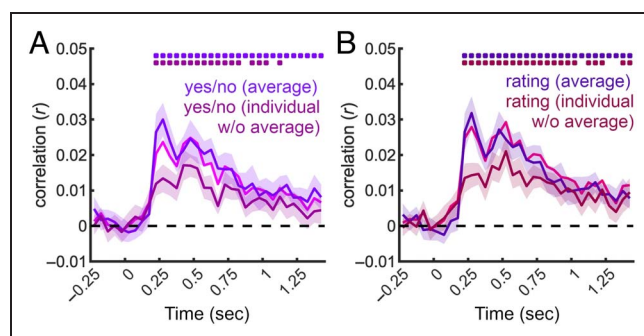

Figure 5. Neural representation of personal aesthetic judgments. (A) Correlations between the neural RDMs and a predictor RDM based on the average yes/no responses of all other participants modeled the data equally well as the predictor RDM based on each individual participant's yes/no responses. Results from the previous analysis (pink line) are shown for reference (same as in Figure 3A). The individual yes/no responses still predicted neural responses well when the average yes/no response across all people was controlled for, suggesting that interindividual variability in aesthetic judgments is reflected in (early) cortical responses. (B) A highly similar pattern of results emerged when the rating responses were analyzed. Results from the previous analysis (red line) are again shown for reference (same as in Figure 3B). Error margins represent standard errors of the mean, and significance markers denote $p$corr $< .05$.

time bin, peaking at 500–550 msec, peak, $t(22) = 5.16$, $p$corr $< .001$, $d = 1.08$). This finding suggests that although both personal and shared ratings can predict (early) neural representations, these ratings explain complimentary shares of the neural dynamics: If predictions based on average judgments captured essentially identical variance in the neural data as predictions based on individual judgments, partialing out the RDM based on average judgments should have eliminated the correlation between the RDM based on individual judgments and the neural data. That individual judgments continued to predict neural representations in this analysis thus suggests that brain representations of aesthetic appeal cannot be fully accounted for by a form of "average taste," but that individual ratings of aesthetic appeal are needed for more accurate prediction.

Extracting Aesthetic Appeal across Scene Contents

A key question in the literature is whether the aesthetic appeal of different content types is extracted through similar or different cortical mechanisms. However, a particular focus in the literature has thus far been placed on the attractiveness of humans. In a first analysis, I therefore asked whether aesthetic appeal is extracted differently for scenes that depict humans and scenes that do not. For this purpose, I split the stimulus set into scenes that did or did not depict humans. I then correlated the neural RDMs and predictor RDMs, only using the pairwise comparisons across scenes with and without humans (Figure 6A). This analysis revealed a significant correspondence between perceived aesthetic appeal and brain responses, both for the yes/no response (from the 200 to 250 msec time bin, peaking at 450–500 msec, peak, $t(22) = 6.22$, $p$corr $< .001$, $d = 1.30$) and the rating response (from the 200- to 250-msec time bin, peaking at 500–550 msec, peak, $t(22) = 5.57$, $p$corr $< .001$, $d = 1.16$). Second, I asked whether aesthetic appeal is extracted similarly for natural and man-made scenes, as suggested by previous fMRI findings (Vessel et al., 2019). I then again performed the RSA only using pairwise comparisons across the natural and man-made scenes (Figure 6B). This analysis also revealed a significant correspondence between perceived attractiveness and brain responses, both for the yes/no response (from the 200- to 250-msec time bin, peaking at 200–250 msec, peak, $t(22) = 4.30$, $p$corr $= .005$, $d = 0.90$) and the rating response (from the 200- to 250-msec time bin, peaking at 500–550 msec, peak, $t(22) = 4.28$, $p$corr $= .002$, $d = 0.89$). Third, I asked whether scene attractiveness is represented differently when it is potentially derived primarily from prominent foreground objects (e.g., a person or an animal occupying the scene foreground), rather than more global scene information. I thus tested whether aesthetic appeal is extracted differently for scenes that had a dominant foreground object, which could allow for a rapid aesthetic appraisal solely based on a single object,
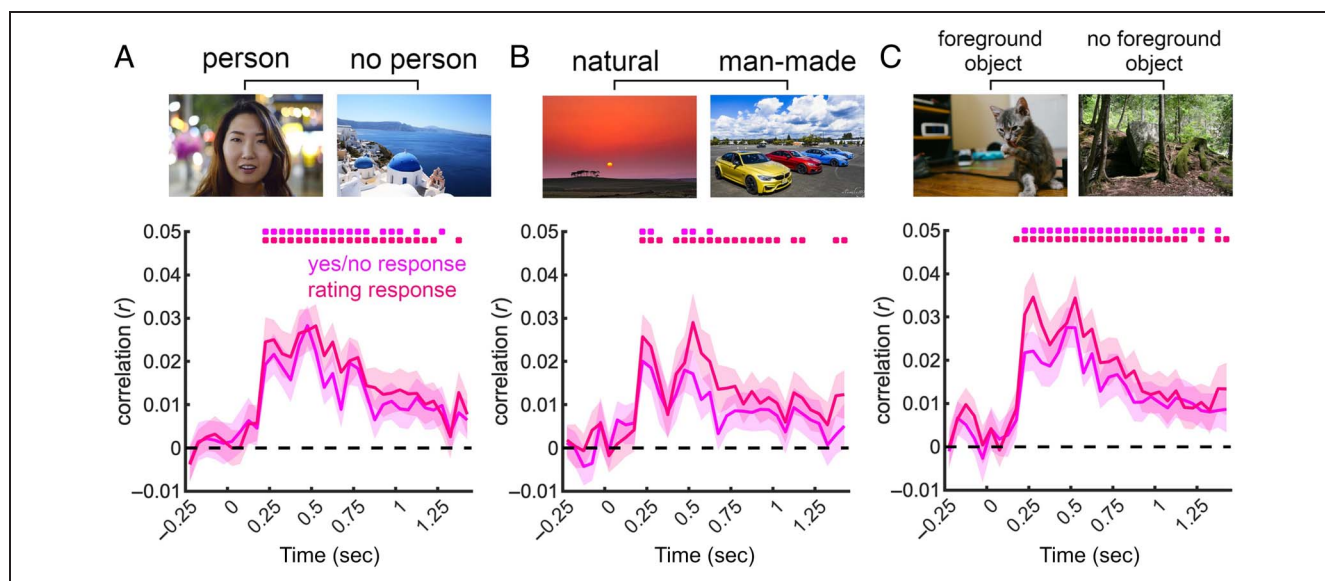
Figure 6. Neural representation of attractiveness across scene contents. (A) Correlations between the neural RDMs and the predictor RDMs based on yes/no responses or ratings responses when only pairwise comparisons between scenes that did or did not contain a person were considered. (B) Correlations between the neural RDMs and the predictor RDMs when only pairwise comparisons between natural and man-made scenes were considered. (C) Correlations between the neural RDMs and the predictor RDMs when only pairwise comparisons between scenes that did or did not contain prominent foreground objects were considered. Together, these three analyses suggest that dynamic neural representations of scene attractiveness are relatively content-independent. Error margins represent standard errors of the mean, and significance markers denote $p$corr < .05.

and scenes that did not. I then performed the RSA only using pairwise comparisons across the foreground-object and no-foreground-object scenes (Figure 6C). Yet again, this analysis revealed a significant correspondence between perceived aesthetic appeal and brain responses, both for the yes/no response (from the 200- to 250-msec time bin, peaking at 450–500 msec, peak, $t(22) = 5.92$, $p$corr < .001, $d = 1.24$) and the rating response (from the 150- to 200-msec time bin, peaking at 250–300 msec, peak, $t(22) = 5.96$, $p$corr < .001, $d = 1.24$). Together, these analyses show that the early extraction of aesthetic appeal for natural scenes is broadly similar across contents, suggesting a common neural correlate for aesthetic perception that is already evident at early stages of visual processing.

DISCUSSION

In this study, I used multivariate pattern analysis on time-resolved EEG data to unveil the neural dynamics of perceived scene attractiveness. I report three key results: (1) Neural representations of scene attractiveness emerge rapidly in brain responses, starting within 200 msec of stimulus onset. (2) These early representations are partly explained by interindividual variation in attractiveness judgments. (3) Neural responses similarly track aesthetic appeal across different scene contents. I will, in turn, discuss the implications of these three key results.

My findings show that the aesthetic appeal of natural scenes is extracted early on during the visual processing cascade, suggesting that scene attractiveness is partly

resolved during perceptual analysis. This is in line with multistage models of aesthetic perception that posit an initial sensory processing stage at which aesthetically pleasing features are extracted (Redies, 2015; Leder, Belke, Oeberst, & Augustin, 2004), as well as with models that explain aesthetic experiences via fluency in sensory processing (Reber, Schwarz, & Winkielman, 2004). My results further complement EEG work using other visual inputs such as faces (Kaiser & Nyga, 2020; Zhang & Deng, 2012; Schacht et al., 2008; Werheid et al., 2007) or abstract patterns (Jacobsen & Höfel, 2003), which show that aesthetic appeal impacts early visual processing stages. So far, only one other magnetoencephalography study also included scene photographs as stimuli (Cela-Conde et al., 2004), but primarily found late modulations of cortical responses starting after 400 msec. However, most stimuli in this study were artworks, and no separate analyses for artworks and scene photographs were conducted. Together, our study suggests that scene attractiveness is first resolved during fundamental stages of scene analysis, around the time when other scene attributes such as scene category or geometry are analyzed (Kaiser, Inciuraite, & Cichy, 2020; Cichy, Khosla, Pantazis, & Oliva, 2017; Harel, Groen, Kravitz, Deouell, & Baker, 2016). Which sensory properties allow scene attractiveness to be resolved during early cortical processing? The matching across relatively attractive and unattractive scenes in the current study suggests that scene attractiveness is not resolved based on a confined set of simple visual features. Furthermore, the failure of categorization-related DNN features in explaining representations of scene attractiveness suggests that the features routinely extracted for the purpose of image

categorization are not the same as the features used for determining an images aesthetic appeal. Which features are evaluated for this purpose is a key question for future research. Beyond these early responses, aesthetic appeal was represented in a sustained way. This sustained neural reflection of aesthetic perception may reflect a transitioning from earlier, sensory-driven responses to later, cognitive appraisal processes (Redies, 2015; Leder et al., 2004). The precise nature of this transition, however, needs to be mapped out in future studies that combine temporally resolved EEG recordings with spatially resolved neural recordings.

Interestingly, the current data suggest that the early neural correlates of perceived scene attractiveness are not fully explained by average ratings across participants. When controlling for average attractiveness ratings across participants, brain responses were still predicted by individual participants' personal ratings. This finding is in line with our previous study on face attractiveness perception (Kaiser & Nyga, 2020), where we could show that early cortical responses are partly predicted by personal attractiveness ratings. Together, these findings suggest that, for multiple visual stimulus categories, even early cortical correlates of aesthetic perception are inherently personal. This notion is not only in line with behavioral studies highlighting substantial interindividual variation in aesthetic judgment (Leder, Goller, Rigotti, & Forster, 2016; Hönekopp, 2006), but also with recent neuroimaging results that show that even basic perceptual representations vary across observers in idiosyncratic ways (Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014). An intriguing possibility that warrants further investigation is that individual differences in the functional architecture of the visual system give rise to different levels of processing fluency (Oppenheimer, 2008; Reber et al., 2004) across stimuli, which ultimately cause idiosyncrasies in aesthetic perception.

My findings further highlight that aesthetic appeal is extracted similarly across different scene contents, such as across (1) scenes that do or do not contain people, (2) natural or man-made scenes, and (3) scene that do or do not contain prominent foreground objects. These results are an important demonstration that even early cortical correlates of perceived aesthetic appeal do not need to be content-specific. They thereby support recent fMRI work that shows a generalization across aesthetic perception for faces, scenes, and artworks (Vessel et al., 2019; Pegors et al., 2015; but see Hu, Huang, Eickhoff, Peng, & Sui, 2020). However, such generalization across contents may ultimately depend on the intrinsic variability in stimulus sets (our stimuli were all rich natural scenes) and the sensitivity of the employed analyses (multivariate pattern analyses may yield higher sensitivity than univariate analyses). Further studies can use the multivariate EEG analysis framework developed here to test generalization across more vastly different visual contents or across modalities.

Together, the current study provides a demonstration for an early cortical correlate of perceived scene attractiveness, which may form the basis of rapid aesthetic evaluation in the wild.

Data Availability Statement

All data are available on the Open Science Framework (OSF; doi.org/10.17605/OSF.IO/SKHF7). Other materials will be made available upon request.

Author Contribution

Daniel Kaiser: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing—Original draft; Writing—Review & editing.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357, PubMed: 9176952

Carbon, C. C., Faerber, S. J., Augustin, M. D., Mitterer, B., & Hutzler, F. (2018). First gender, then attractiveness: Indications of gender-specific attractiveness processing via

ERP onsets. *Neuroscience Letters*, *686*, 186–192. https://doi
.org/10.1016/j.neulet.2018.09.009, PubMed: 30217503

Cela-Conde, C. J., Marty, G., Maestú, F., Ortiz, T., Munar, E.,
Fernández, A., et al. (2004). Activation of the prefrontal
cortex in the human visual aesthetic perception.
Proceedings of the National Academy of Sciences, U.S.A.,
101, 6321–6325. https://doi.org/10.1073/pnas.0401427101,
PubMed: 15079079

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., &
Kriegeskorte, N. (2014). Unique semantic space in the brain
of each beholder predicts perceived similarity. *Proceedings
of the National Academy of Sciences, U.S.A.*, *111*,
14565–14570. https://doi.org/10.1073/pnas.1402594111,
PubMed: 25246586

Cichy, R. M., & Kaiser, D. (2019). Deep neural networks
as scientific models. *Trends in Cognitive Sciences*, *23*,
305–317. https://doi.org/10.1016/j.tics.2019.01.009, PubMed:
30795896

Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017).
Dynamics of scene representations in the human brain
revealed by magnetoencephalography and deep
neural networks. *Neuroimage*, *153*, 346–358.
https://doi.org/10.1016/j.neuroimage.2016.03.063,
PubMed: 27039703

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A.
(2016). Comparison of deep neural networks to
spatio-temporal cortical dynamics of human visual object
recognition reveals hierarchical correspondence. *Scientific
Reports*, *6*, 27755. https://doi.org/10.1038/srep27755,
PubMed: 27282108

Datta, R., Li, J., & Wang, J. Z. (2008). Algorithmic inferencing of
aesthetics and emotion in natural images: An exposition. In
*2008 15th IEEE international conference on image
processing* (pp. 105–108). IEEE.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L.
(2009). Imagenet: A large-scale hierarchical image database.
In *2009 IEEE conference on computer vision and pattern
recognition* (pp. 248–255). IEEE.

de Tommaso, M., Pecoraro, C., Sardaro, M., Serpino, C.,
Lancioni, G., & Livrea, P. (2008). Influence of aesthetic
perception on visual event-related potentials. *Consciousness
and Cognition*, *17*, 933–945. https://doi.org/10.1016/j.concog
.2007.09.003, PubMed: 17977747

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017).
Decoding dynamic brain patterns from evoked responses: A
tutorial on multivariate pattern analysis applied to time series
neuroimaging data. *Journal of Cognitive Neuroscience*, *29*,
677–697. https://doi.org/10.1162/jocn_a_01068, PubMed:
27779910

Harel, A., Groen, I. I., Kravitz, D. J., Deouell, L. Y., & Baker, C. I.
(2016). The temporal dynamics of scene processing: A
multifaceted EEG investigation. *eNeuro*, *3*. https://doi.org/10
.1523/ENEURO.0139-16.2016, PubMed: 27699208

Höfel, L., & Jacobsen, T. (2007). Electrophysiological indices of
processing aesthetics: Spontaneous or intentional processes?
International Journal of Psychophysiology, *65*, 20–31.
https://doi.org/10.1016/j.ijpsycho.2007.02.007, PubMed:
17400317

Hönekopp, J. (2006). Once more: Is beauty in the eye of the
beholder? Relative contributions of private and shared taste
to judgments of facial attractiveness. *Journal of
Experimental Psychology: Human Perception and
Performance*, *32*, 199–209. https://doi.org/10.1037/0096-1523
.32.2.199, PubMed: 16634665

Hu, C.-P., Huang, Y., Eickhoff, S. B., Peng, K., & Sui, J. (2020).
Seeking the "beauty center" in the brain: A meta-analysis of
fMRI studies of beautiful human faces and visual art.
Cognitive, Affective, & Behavioral Neuroscience, *20*,
1200–1215. https://doi.org/10.3758/s13415-020-00827-z,
PubMed: 33089442

Isik, A. I., & Vessel, E. A. (2021). From visual perception to
aesthetic appeal: Brain responses to aesthetically appealing
natural landscape movies. *Frontiers in Human
Neuroscience*, *15*, 676032. https://doi.org/10.3389/fnhum
.2021.676032, PubMed: 34366810

Jacobsen, T., & Höfel, L. (2003). Descriptive and evaluative
judgment processes: Behavioral and electrophysiological
indices of processing symmetry and aesthetics. *Cognitive,
Affective, & Behavioral Neuroscience*, *3*, 289–299.
https://doi.org/10.3758/CABN.3.4.289, PubMed:
15040549

Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V.
(2006). Brain correlates of aesthetic judgment of beauty.
Neuroimage, *29*, 276–285. https://doi.org/10.1016/j
.neuroimage.2005.07.010, PubMed: 16087351

Kaiser, D., Häberle, G., & Cichy, R. M. (2020). Real-world
structure facilitates the rapid emergence of scene category
information in visual brain signals. *Journal of
Neurophysiology*, *124*, 145–151. https://doi.org/10.1152/jn
.00164.2020, PubMed: 32519577

Kaiser, D., Inciuraite, G., & Cichy, R. M. (2020). Rapid
contextualization of fragmented scene information in
the human visual system. *Neuroimage*, *219*, 117045. https://
doi.org/10.1016/j.neuroimage.2020.117045, PubMed:
32540354

Kaiser, D., & Nyga, K. (2020). Tracking cortical representations
of facial attractiveness using time-resolved representational
similarity analysis. *Scientific Reports*, *10*, 16852. https://doi
.org/10.1038/s41598-020-74009-9, PubMed: 33033356

Kawabata, H., & Zeki, S. (2004). Neural correlates of beauty.
Journal of Neurophysiology, *91*, 1699–1705. https://doi.org
/10.1152/jn.00696.2003, PubMed: 15010496

Kriegeskorte, N. (2015). Deep neural networks: A new
framework for modeling biological vision and brain
information processing. *Annual Review of Vision Science*, *1*,
417–446. https://doi.org/10.1146/annurev-vision-082114
-035447, PubMed: 28532370

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008).
Representational similarity analysis-connecting the branches
of systems neuroscience. *Frontiers in Systems Neuroscience*,
2, 4. https://doi.org/10.3389/neuro.06.004.2008, PubMed:
19104670

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A
model of aesthetic appreciation and aesthetic judgments.
British Journal of Psychology, *95*, 489–508. https://doi.org/10
.1348/0007126042369811, PubMed: 15527534

Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and
shared taste in art and face appreciation. *Frontiers in
Human Neuroscience*, *10*, 155. https://doi.org/10.3389
/fnhum.2016.00155, PubMed: 27148012

Murray, N., Marchesotti, L., & Perronnin, F. (2012). AVA: A
large-scale database for aesthetic visual analysis. In *2012 IEEE
conference on computer vision and pattern recognition*
(pp. 2408–2415). IEEE.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the
scene: A holistic representation of the spatial envelope.
International Journal of Computer Vision, *42*, 145–175.
https://doi.org/10.1023/A:1011139631724

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011).
FieldTrip: Open source software for advanced analysis of
MEG, EEG, and invasive electrophysiological data.
Computational Intelligence and Neuroscience, *2011*,
156869. https://doi.org/10.1155/2011/156869, PubMed:
21253357

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016).
CoSMoMVPA: Multi-modal multivariate pattern analysis of

neuroimaging data in Matlab/GNU octave. *Frontiers in Neuroinformatics*, *10*, 27. https://doi.org/10.3389/fninf.2016.00027, PubMed: 27499741

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*, 237–241. https://doi.org/10.1016/j.tics.2008.02.014, PubMed: 18468944

Pearce, M. T., Zaidel, D. W., Vartanian, O., Skov, M., Leder, H., Chatterjee, A., et al. (2016). Neuroaesthetics: The cognitive neuroscience of aesthetic experience. *Perspectives on Psychological Science*, *11*, 265–279. https://doi.org/10.1177/1745691615621274, PubMed: 26993278

Pegors, T. K., Kable, J. W., Chatterjee, A., & Epstein, R. A. (2015). Common and unique representations in pFC for face and place attractiveness. *Journal of Cognitive Neuroscience*, *27*, 959–973. https://doi.org/10.1162/jocn_a_00777, PubMed: 25539044

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364–382. https://doi.org/10.1207/s15327957pspr0804_3, PubMed: 15582859

Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in Human Neuroscience*, *9*, 218. https://doi.org/10.3389/fnhum.2015.00218, PubMed: 25972799

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025. https://doi.org/10.1038/14819, PubMed: 10526343

Schacht, A., Werheid, K., & Sommer, W. (2008). The appraisal of facial beauty is rapid but not mandatory. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 132–142. https://doi.org/10.3758/CABN.8.2.132, PubMed: 18589504

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 411–426. https://doi.org/10.1109/TPAMI.2007.56, PubMed: 17224612

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Skov, M., & Nadal, M. (2020). A farewell to art: Aesthetics as a topic in psychology and neuroscience. *Perspectives on Psychological Science*, *15*, 630–642. https://doi.org/10.1177/1745691619897963, PubMed: 32027577

Strijbosch, W., Vessel, E. A., Welke, D., Mitas, O., Gelissen, J., & Bastiaansen, M. (2022). On the neuronal dynamics of aesthetic experience: Evidence from electroencephalographic oscillatory dynamics. *Journal of Cognitive Neuroscience*, *34*, 461–479. https://doi.org/10.1162/jocn_a_01812, PubMed: 35015884

Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L., & Starr, G. G. (2019). The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proceedings of the National Academy of Sciences, U.S.A.*, *116*, 19155–19164. https://doi.org/10.1073/pnas.1902650116, PubMed: 31484756

Werheid, K., Schacht, A., & Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biological Psychology*, *76*, 100–108. https://doi.org/10.1016/j.biopsycho.2007.06.008, PubMed: 17681418

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*, 195–206. https://doi.org/10.1016/j.neuropsychologia.2006.05.009, PubMed: 16828125

Yue, X., Vessel, E. A., & Biederman, I. (2007). The neural basis of scene preferences. *NeuroReport*, *18*, 525–529. https://doi.org/10.1097/WNR.0b013e328091c1f9, PubMed: 17413651

Zhang, Z., & Deng, Z. (2012). Gender, facial attractiveness, and early and late event-related potential components. *Journal of Integrative Neuroscience*, *11*, 477–487. https://doi.org/10.1142/S0219635212500306, PubMed: 23351053

Zhao, X., Wang, J., Li, J., Luo, G., Li, T., Chatterjee, A., et al. (2020). The neural mechanism of aesthetic judgments of dynamic landscapes: An fMRI study. *Scientific Reports*, *10*, 20774. https://doi.org/10.1038/s41598-020-77658-y, PubMed: 33247221