COGNITIVE NEUROSCIENCE

Integrative processing in artificial and biological vision predicts the perceived beauty of natural images

Sanjeev Nara[1] and Daniel Kaiser[1,2]*

Previous research shows that the beauty of natural images is already determined during perceptual analysis. However, it is unclear which perceptual computations give rise to the perception of beauty. Here, we tested whether perceived beauty is predicted by spatial integration across an image, a perceptual computation that reduces processing demands by aggregating image parts into more efficient representations of the whole. We quantified integrative processing in an artificial deep neural network model, where the degree of integration was determined by the amount of deviation between activations for the whole image and its constituent parts. This quantification of integration predicted beauty ratings for natural images across four studies with different stimuli and designs. In a complementary functional magnetic resonance imaging study, we show that integrative processing in human visual cortex similarly predicts perceived beauty. Together, our results establish integration as a computational principle that facilitates perceptual analysis and thereby mediates the perception of beauty.

INTRODUCTION

During our daily lives, some visual images reliably evoke a feeling of beauty while others do not. However, why do images differ in their ability to evoke beauty? Models of aesthetic appreciation assume a succession of two broadly different processing stages (*1*–*3*): (i) a rapid and automatic appraisal, mainly driven by objectifiable physical stimulus attributes, and (ii) a slower cognitive evaluation that is more strongly influenced by personal experience and context. Focusing on the physical attributes that make a stimulus beautiful to human observers, studies on low-level visual features revealed a set of properties that are associated with perceived beauty, such as an image's color, curvature, or symmetry (*4*–*7*). The perception of beauty may thus arise from the presence of relatively basic features, as well as their spatial configuration in the image (*5*, *8*). The important role of visual image properties in evoking beauty is consistent with neuroscientific studies that show that beauty ratings for natural images are predicted by activations in cortical regions responsible for perceptual analysis (*9*–*11*) and early neural responses associated with perceptual processing (*12*).

Despite the realization that the beauty of an image is—at least partially—determined during perceptual analysis, it is still largely unclear which perceptual mechanisms govern the formation of the phenomenological experience of beauty. An intriguing proposal is that processing fluency, that is, the ease with which is stimulus can be analyzed in the visual system, plays a critical role for perceived beauty (*13*, *14*). This idea is further refined in the pleasure-interest model of aesthetic liking (PIA model), which stresses the discrepancy between expected and experienced processing fluency (*15*). Critical evidence for fluency-based proposals comes from studies showing that the degree of structure or organization in a stimulus (such as whether multiple image elements can be organized according to Gestalt principles) affects both perceptual processing efficiency and perceived beauty (*5*, *8*).

One critical limitation of the human visual system is its confined resources for representing multiple objects at once (*16*–*18*). On the neural level, simultaneous objects compete for these confined resources, leading to marked reductions in neural activity when multiple individual objects are presented (*19*–*21*). Typical compositions were previously associated with a release from neural competition, alleviating the detrimental neural effects of representing multiple simultaneous objects (*22*, *23*). This decrease in competition has been linked to an increasing capacity for integration: When image elements (such as multiple objects) can be represented as a meaningful whole, rather than multiple unrelated entities, the visual system can efficiently reduce the complexity of representations, yielding a more efficient—or fluent—neural code (*22*, *24*). Such effects may prominently facilitate the processing of natural scenes, where usually dozens to hundreds of objects (*25*) are competing for representation. Although we know that efficient neural integration facilitates perceptual analysis and thus increases perceptual fluency, it is unclear whether efficient integration similarly determines perceived beauty.

Here, we thus tested whether the degree of visual integration across an image can reliably predict whether natural images are perceived as beautiful. To quantify integration, we used a deep neural network (DNN) as a model of the biological visual processing cascade (*26*, *27*). Within this neural network, we quantified integration by computing how well whole images were predicted by a combination of their individual parts. This integration measure predicted perceived beauty across four studies with different natural images and under different task demands. In the light of recent reports of critical differences between DNN models and human brain and behavior (*28*–*30*), we, however, cannot assert with highest confidence that integrative processing in DNNs faithfully resembles integrative processing in the human visual system. We thus applied the same analysis logic to human functional magnetic resonance imaging (fMRI) data recorded for a subset of images. We show that integration in scene-selective visual cortex predicts perceived beauty in a similar way as our DNN-derived quantification of integration. Together, these results highlight that integration is a critical computation for the evaluation of beauty in natural images.

[1]Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Gießen, Gießen Germany. [2]Center for Mind, Brain and Behavior (CMBB), Philipps-University Marburg and Justus Liebig University Gießen, Marburg, Germany.
*Corresponding author. Email: danielkaiser.net@gmail.com

RESULTS

Quantifying integration in a DNN

Here, we tested whether the degree of visual integration across an image can reliably predict whether the image is rated as beautiful. To quantify integration, we used a DNN as a model of the biological visual system (*26*, *27*). Specifically, we used a VGG16 network architecture (*31*) pretrained on scene categorization using the Places365 image set (*32*) (Fig. 1A).

Within this neural network, we quantified integration by computing how well whole images were predicted by a combination of their individual parts, where accurate prediction indexes largely parallel processing, and thus a low degree of integration, while less accurate prediction indexes interactive processing across image parts, and thus a higher degree of integration. This logic is derived from human neuroimaging studies, which have shown that the average response to multiple image elements accurately predicts the response to the full image (*33*–*37*), and that a relative decrease in this prediction indicates integrative processing that renders responses to the whole image dissimilar from the responses to its parts (*24*, *34*, *38*). Recent computational investigations suggest that DNNs show a similar averaging of responses for multiple image elements (*39*), although average responses in DNNs do not always resemble representations to the whole as faithfully as in the human brain (*40*). Critically, a recent study suggests that DNNs integrate information in similar ways as the human brain, evidenced by nonlinearities in responses for multiple typically configured objects (*41*), similar to those observed in visual cortex (*34*).

Here, we computed integrative processing for individual images, by feeding the network two halves of an image (e.g., the bottom-left and top-right quadrants versus the bottom-right and top-left quadrants), as well as the full image (Fig. 1B). For each image, and separately for each network layer, we then computed how much the activation pattern to the full image was correlated to the average activation pattern to the two halves. The strength of this correlation was taken as a measure of integrative processing, where lower values (i.e., a higher dissimilarity between the whole and its parts) indicates a higher degree of integration (Fig. 1C). The integration measure was computed separately at five different spatial granularities, where halves were created by dividing the image into 2×2, 4×4 (as in Fig. 1B), 8×8, 16×16, or 32×32 identical squares. Each image half contained all odd or all even squares (i.e., corresponding to either all white or black squares on a checkerboard). This procedure allowed us to probe integrative processing at different spatial scales.

We then tested whether our DNN-derived integration measure could successfully predict perceived beauty. To this end, we use the image-specific integration measure to predict beauty ratings in a series of four studies with varying image sets and task demands.

Predicting perceived beauty from integration in a DNN

In study 1, we collected beauty ratings for 250 natural scene images from 25 online participants (Fig. 2A; see Materials and Methods for details). During our experiment, we only showed the images briefly (50-ms exposure time). Previous work has shown that observers can judge the beauty of an image even under such brief presentation regimes (*42*). We reasoned that brief exposure would lead to more successful prediction of perceived beauty from our integration measure, as observers do not have time for extensive cognitive evaluation of the image. Correlating the integration measure with beauty ratings revealed a strong relationship between integration and perceived beauty, with correlations of up to $r = 0.6$ (Fig. 2B), showing that a higher degree of integrative processing predisposes a higher beauty rating. There were two notable patterns in these correlations: First, correlations were strongest in intermediate to late network layers, suggesting that integration over mid- and high-level features determines perceived beauty. Second, correlations were apparent across all spatial scales but strongest for the coarser scales, with a decrease for the 16×16 and 32×32 scales. This suggests that integration across larger parts of the images is a stronger predictor of perceived beauty than integration across fine details. To test whether the integration measure could also accurately predict beauty ratings

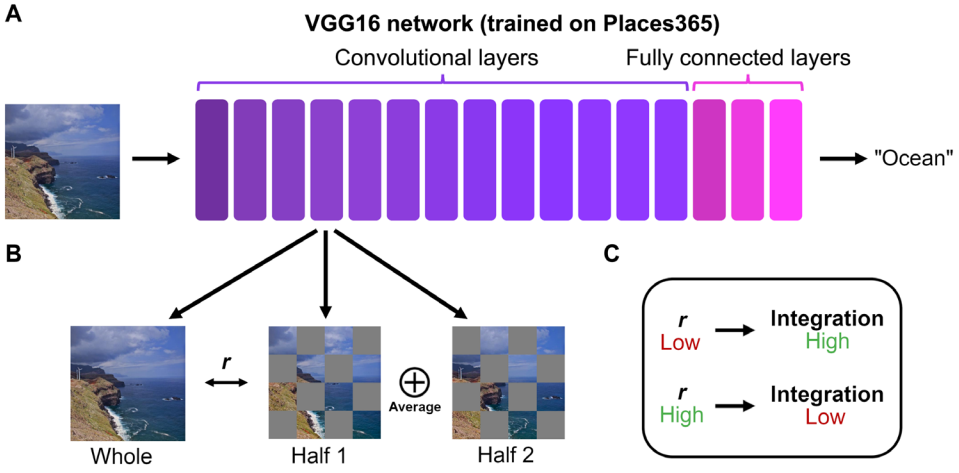

Fig. 1. DNN-based quantification of integrative processing. (**A**) We used a VGG16 DNN as an in silico model of cortical scene processing. The DNN was trained on scene categorization using the Places365 dataset. (**B**) We fed the DNN with each full image, as well as with two halves of the images. Halves were generated by obscuring 50% of the image in a checkerboard-like fashion, with different spatial scales (i.e., number of squares on the checkerboard). The example shows the 4×4 spatial scale. To quantify integration, we correlated the layer-specific activation pattern to each whole image with the average of the activation patterns to the two halves. (**C**) When the resulting correlation is low, one can infer more integration (as integrative processes are not captured by the activation patterns to the parts), whereas when the correlation is high, one can infer less integration (as the average of the parts accurately captures the whole).

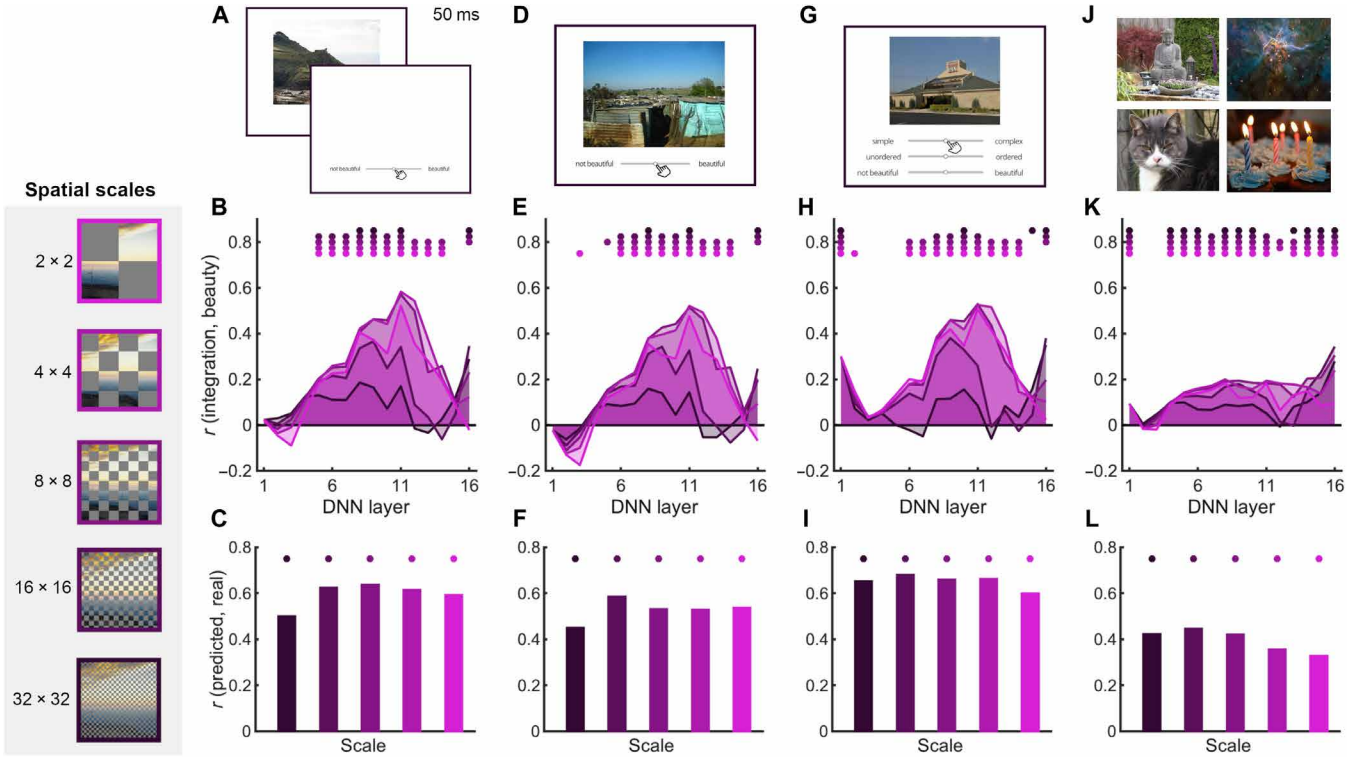

Fig. 2. The degree of integrative processing in a DNN predicts perceived beauty. (**A**) In study 1, participants briefly viewed 250 natural scene images and rated their beauty. (**B**) We then correlated the DNN-derived integration measure (see Fig. 1) with beauty ratings across images, separately for each network layer and each spatial scale (see left). Integration significantly predicted beauty ratings across images, with highest correlations in intermediate network layers and coarser spatial scales. (**C**) To assess predictions for novel images, we further estimated linear models with the beauty ratings as the criterion and the integration measure in each layer as predictors, for all but one image. We then generated predictions for the left-out images with these linear models and correlated the predicted ratings of the left-out images with the real ratings for these images. The linear model could predict ratings for novel images across all spatial scales. (**D**) In study 2, participants viewed the same set of images as in study 1, now with unlimited viewing time. (**E** and **F**) Results strongly resembled the results from study 1. (**G**) In study 3, participants viewed a disjoint set of 250 natural scenes images, again with unlimited viewing time. Here, participants additionally rated the images' complexity and order (see text). (**H** and **I**) Results were again similar to studies 1 and 2. (**J**) In study 4, we used beauty ratings collected for 900 vastly different natural images from the OASIS database. (**K** and **L**) Even for such diverse images, our integration measures significantly predicted beauty ratings, though to a lesser extent than for the more homogeneous natural scene images. Dots indicate $P < 0.05$ (corrected for multiple comparisons).

for novel images, we next fit a linear model on integration values derived from all 16 network layers, using all but one image. We then predicted the values for the left-out image using the fit model. After each image was left out once, we correlated the model predictions with the real beauty ratings. We found that the linear model could successfully predict beauty ratings for novel images at all spatial scales (Fig. 2C), with numerically strongest predictions for an intermediate 8 × 8 spatial scale.

In study 2, we asked whether the same pattern of results could be replicated when observers are able to study the image as long as they want. We thus tested another 25 online participants, who rated the same 250 natural scene images. Here, they had unlimited time to provide their beauty rating while the image stayed on the screen, allowing them to also cognitively evaluate the image in detail (Fig. 2D). Perhaps unexpectedly, the integration measure essentially predicted perceived beauty equally well as for briefly presented images (Fig. 2, E and F), replicating the pattern obtained in study 1.

In study 3, we sought to replicate the findings from study 2 with a completely different set of 250 natural scenes, rated for their beauty by 26 online observers. Here, observers additionally rated the complexity and order of each image (Fig. 2G; see Materials and Methods for details), which allowed us to gauge how the integration measure relates to human ratings of how complex or ordered an image is (see below). Results again replicated the pattern from studies 1 and 2, showing that integrative processing is a strong predictor for perceived beauty (Fig. 2, H and I). We further tested whether human-rated image complexity and order could explain integration within the DNN. Complexity and order both linearly predicted beauty ratings ($r = 0.11$, $P = 0.09$ for complexity, $r = 0.31$, $P < 0.001$ for order). In a variance partitioning analysis (*43*, *44*), we found that beauty predictions in a linear model were mainly driven by order ($R^2 = 0.14$), with complexity ($R^2 < 0.01$) and a combined model ($R^2 = 0.01$) not predicting additional unique variance. Further, complexity and order were only moderately correlated to the integration measure derived from the DNN (all $r < 0.22$ for complexity and $r < 0.22$ for order). When complexity and order ratings were partialed out, integration in the DNN could still predict beauty ratings well (see fig. S1), suggesting that human ratings of complexity and order do not fully capture the visual features that predispose integrative processing in the DNN.

Last, in study 4, we tested whether our integrative processing measure could predict perceived beauty not only for natural scenes but also for a wide range of photographs that depict objects, people, and everyday situations. Here, we used beauty ratings obtained for a set of 900 diverse natural images contained in the Open Affective Standardized Image Set (OASIS) database (*45*, *46*) (Fig. 2J). For these images, given their large variability in content and emotional valence, we expected that predictions derived from our integrative processing measure would be reduced. If the integration measure still predicted perceived beauty in these images, however, it suggests that integrative processing is a computation that predisposes beauty across natural images from various domains. As hypothesized, correlations were indeed lower, but integrative processing still predicted beauty ratings (Fig. 2K). Similar to the previous studies, perceived beauty was again better predicted from intermediate layer activations and from coarser spatial scales. Despite these reduced correlations, the DNN-derived integration measure still successfully predicted beauty ratings for novel images (Fig. 2L).

In supplementary analyses, we show that the same prediction of perceived beauty can be achieved with a VGG16 network trained on object categorization instead of scene categorization (fig. S2) and that predictions are stable across categorical image clusters identified in a data-driven way (fig. S3). We further evaluated a second possible predictor of perceived beauty: We evaluated part-based similarity, that is, the degree to which parts of an image show visual similarity to other parts of the image. Previous research has suggested that similarity between multiple parts can, like integration across parts, alleviate neural competition in visual cortex (*47*). To quantify part-based similarity, we computed the similarity in DNN activations to one half of the image and the other half of the image. Our data show that part-based similarity is also capable of predicting perceived beauty, albeit to a lesser extent than integration (fig. S4, A and B), and that the association between integration and beauty cannot be explained by an image's self-similarity (fig. S4C). Last, we show that the degree to which whole images activate the network (operationalized through the L2 norm of the activation pattern in each layer) also predicts beauty ratings to some extent but that this prediction cannot account for the stronger predictions provided by our integration measure (fig. S5). When adding all three predictors in a linear model, a variance partitioning analysis showed that the integration measure predicts a substantial share of unique variance in the beauty ratings (fig. S6).

Together, our four studies demonstrate that a DNN-derived measure of integrative processing predicts perceived beauty across different natural images and under different task demands. Using a DNN provides a powerful way of estimating integrative processing in an objective and scalable way, which can be used to derive a computational prediction for perceived beauty across large sets of images.

Charting visual properties that drive the prediction of perceived beauty

While the results thus far demonstrate that the degree of integrative processing in a DNN model predicts perceived beauty, they do not directly reveal which visual properties of the images enable this prediction. To chart how different visual properties contribute to the prediction of perceived beauty, we conducted an additional analysis, in which we manipulated the images supplied to the DNN and assessed how the prediction of beauty changes as a function of visual properties being removed from the images in a targeted way. For this analysis, we focused on the 8×8 spatial scale, which offered the numerically highest correlations in the original analysis (see Fig. 2). Analysis on the other spatial scales yielded qualitatively similar results (see fig. S7).

Specifically, we manipulated the images in three different ways (Fig. 3A; see Materials and Methods for details): First, to test how color, luminance, and contrast contribute to predictions, we grayscaled the images and additionally either equated their luminance or their luminance and contrast (*48*). Stripping away color, luminance, and contrast did not systematically impair the prediction of perceived beauty across all studies (see Fig. 3B for correlations between integration and beauty and Fig. 3C for predictions by a linear model trained on the integration measure across all layers). While color and luminance did not alter the predictions of perceived beauty (all $P > 0.06$, comparison to original analysis), predictions were reduced when also the contrast was matched for studies 3 ($P < 0.001$) and 4 ($P = 0.001$) but not studies 1 and 2 (both $P > 0.54$). Together, these results suggest that simple visual features like color, luminance, and contrast cannot account for our effects. As a proof of concept, we additionally "pixelated" the images by randomly shuffling all pixels. As expected, this extreme manipulation abolished the correlation between integration and beauty in all studies (all $P < 0.001$).

Second, to test how spatial frequency content contributes to predictions, we low-pass or high-pass filtered the images (*49*, *50*). The high-pass–filtered images yielded predictions similar to the original analysis (all $P > 0.06$). Although low-pass–filtered images still allowed for successful prediction (Fig. 3C), predictions were reduced compared to the original images (all $P < 0.001$) and compared to the high-pass filtered images (studies 1 to 3: all $P < 0.001$; study 4: $P = 0.084$). This suggests that the prediction of beauty hinges more strongly on integration over details conveyed by high spatial frequencies than on integration across global layout conveyed by low spatial frequencies.

Last, to test how the images' global spatial configuration contributes to predictions, we rotated the images by 90° or 180° (*51*–*53*) or jumbled them across space (*52*, *54*). Perhaps unexpectedly, these manipulations of high-level image configurations did not affect the prediction of beauty systematically across studies. Image rotation did not have any significant effect on predictions (all $P > 0.10$, comparison to original analysis; Fig. 3C). While jumbling reduced predictions in studies 1 ($P < 0.001$) and 2 ($P = 0.048$), it did not alter predictions in studies 3 and 4 (both $P > 0.16$). This suggests that the integration effects that mediate the perception of beauty are not contingent on the typical global structure of natural scenes. This notion is consistent with the highest correlations observed in intermediate-to-late network layers, with a peak in layer 11, suggesting that effective integration of features of intermediate complexity drives the predictions of perceived beauty. Together, the manipulation of visual properties did not unequivocally distill the features whose integration is ultimately critical for predicting beauty. The analysis, however, suggests that neither very basic visual features like color or luminance nor high-level configural properties exclusively drive predictions of beauty.

In another supplementary analysis, we additionally assessed whether image symmetry explains the prediction of perceived beauty. To this end, we computed each image's symmetry using a method that identifies the most prominent symmetry axes in a data-driven

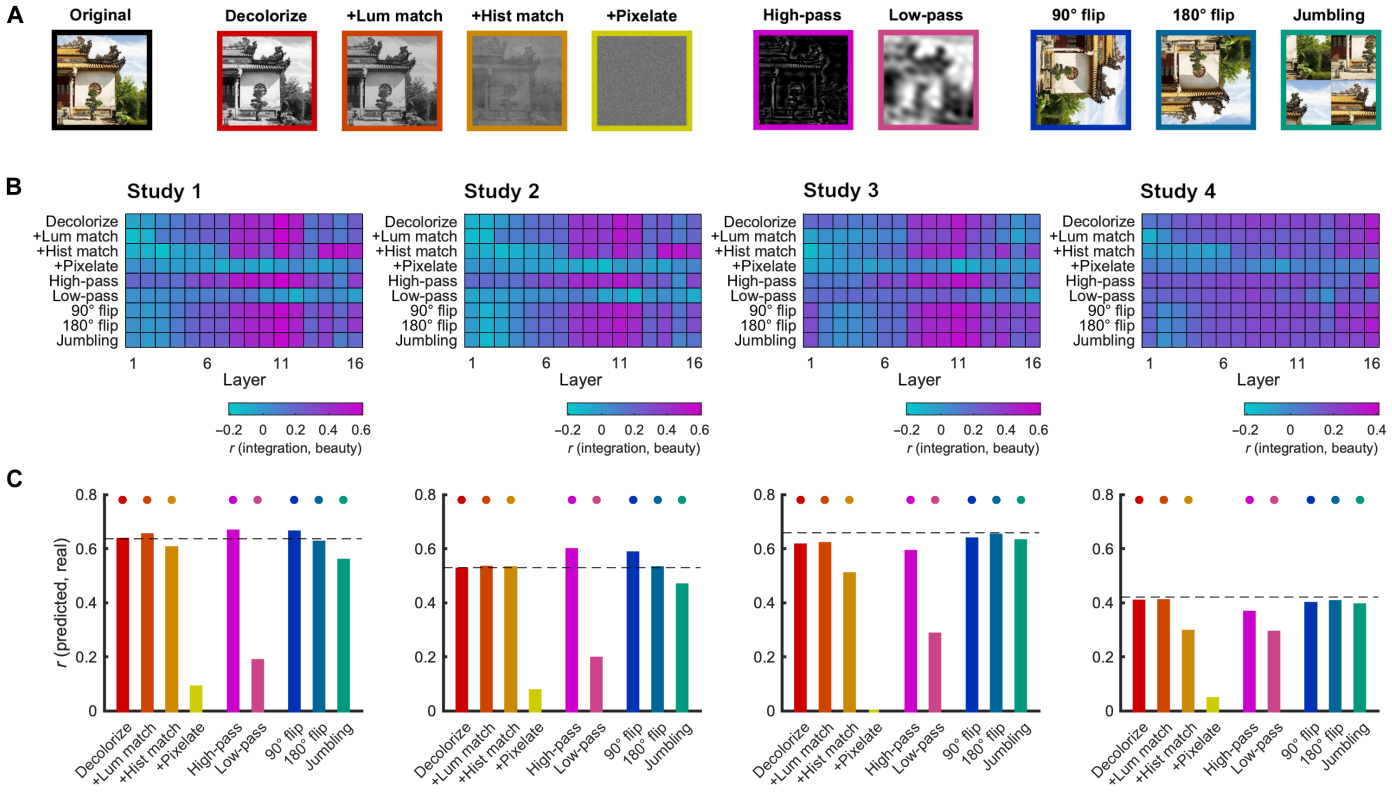

Fig. 3. Charting integration across visual image properties. (**A**) Here, we manipulated the original input images supplied to the DNN to explore how the relationship of integration and perceived beauty changes when visual properties are stripped away in targeted ways. We performed three complementary types of image manipulations: First, we gray-scaled and additionally luminance- and contrast-matched the images. In addition, as a proof of concept, we additionally removed all spatial information by shuffling the image pixels. Second, we high-pass or low-pass filtered the images. Third, we interfered with the global configuration of the image by rotating the image or jumbling image parts across space. (**B**) Correlations between the integration measure and perceived beauty (on the 8 × 8 spatial scale) remained similar to the original analysis (see Fig. 2) for most manipulations, with the exception of the pixelated images and the low-pass filtered images. (**C**) Statistically assessing these differences in a linear modeling analysis showed that integration indeed significantly predicted perceived beauty across most of the manipulations. Only the pixelation and low-pass filtering reduced predictions compared to the original analysis (dashed line) across all the experiments. Further, contrast matching reduced predictions in studies 3 and 4 but not studies 1 and 2. Both very basic visual features like color and luminance and more high-level configural properties thus cannot directly account for the relationship of integration and perceived beauty, suggesting that features of intermediate complexity are critical for the prediction of perceived beauty.

way (*55*). Symmetry across the vertical and horizontal images axes only provided weak predictions of perceived beauty and could not account for the relationship between integrative processing and beauty (see fig. S8).

Predicting perceived beauty from integration in the human visual system

However, we cannot be fully sure whether our DNN indeed replicates the way in which the human brain integrates information across images: After all, DNNs have been shown to diverge from human visual processing in potentially critical ways (*28*, *30*, *40*). Given this backdrop, we wanted to investigate whether an integrative processing measure derived from human fMRI data is similarly capable of predicting perceived beauty. We thus ran an fMRI study, in which 21 participants viewed a set of 32 natural scene images (which were rated for beauty in study 3). During this study, participants viewed the whole images as well as their two complimentary halves (see Materials and Methods for details). Halves were the bottom-left and top-right versus the bottom-right and top-left quadrants, thus resembling the 2 × 2 spatial scale in the DNN analysis. We extracted multivoxel fMRI patterns from a set of regions in retinotopic early visual cortex (V1, V2, V3, and V4) and scene-selective cortex [occipital place area (OPA), medial place area (MPA), and parahippocampal place area (PPA)]. We then performed an analysis analogous to our DNN analysis: For each scene, we computed the correlation between the multivoxel response pattern to the whole image and the average of the multivoxel response pattern to the two halves (Fig. 4A; see Materials and Methods for details). This correlation was again used as a measure of integrative processing, where lower correlations indicate a higher degree of integration (*24*, *34*, *38*).

Correlating the integration measure derived from the fMRI data with the beauty ratings for the 32 images used in the fMRI experiment revealed a significant correlation in V2 and scene-selective PPA (Fig. 4B). However, there is a possibility that different degrees of image-specific integration simply reflect differences in the reliability of responses across images: If an image yields less reliable responses, then the response cannot be approximated well to begin with. To address this concern, we additionally computed a measure of reliability for each whole scene (see Materials and Methods),

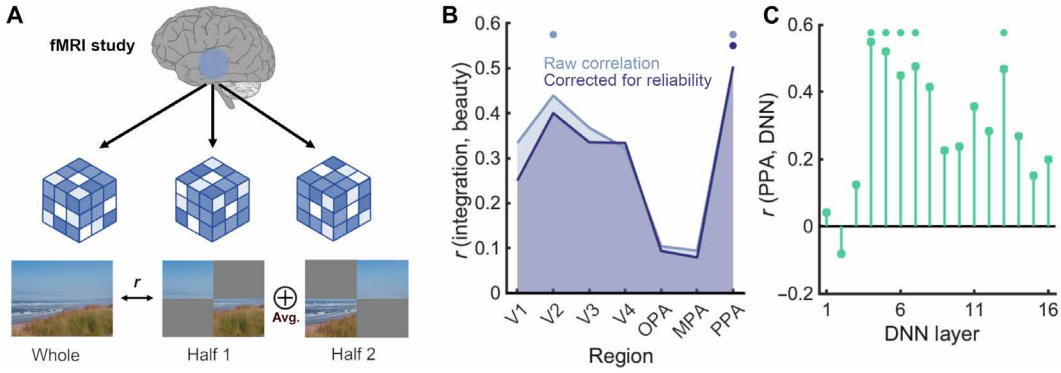

Fig. 4. Integrative processing in scene-selective PPA predicts perceived beauty. (**A**) To test whether integrative processing in the human visual system similarity predicts perceived beauty as integrative processing in an artificial DNN, we conducted an fMRI study in which participants viewed 32 whole natural scene images and their corresponding halves (on the 2 × 2 spatial scale). Integration was quantified by computing the correlation between the response pattern evoked by each whole image (from half of the runs) and the average of the response patterns evoked by its two halves (from the other half of runs). Integration was assessed for four regions in early visual cortex (V1 to V4) and four regions in scene-selective cortex (OPA, MPA, and PPA). (**B**) The degree of integrative processing in V2 and PPA significantly correlated with beauty ratings across images, suggesting that integrative processing in the biological visual system is similarity related to perceived beauty as integrative processing in an artificial DNN. The correlation between integration and beauty ratings in PPA remained significant when controlling for the reliability of fMRI responses to the whole images (by assessing correlations in response patterns evoked by the whole images across experimental runs). (**C**) Across images, the degree of integrative processing in the DNN (on the 2 × 2 spatial scale) correlated with the degree of integration in the PPA, specifically in middle layers of the DNN, revealing a correspondence between integration in biological and artificial vision. Dots indicate $P < 0.05$ (corrected for multiple comparisons).

which we partialed out when correlating the integration measure and the beauty ratings. In this analysis, integration in PPA still significantly predicted perceived beauty. We also correlated the integration measure obtained from the PPA with the integration measure obtained from the DNN on 2 × 2 spatial scale (Fig. 4C). We found significant correlations between the degree of integrative processing in the PPA and the DNN, again peaking at the intermediate layers, where integration was most predictive of the beauty ratings (see Fig. 2). This suggests that integration varies across images in similar ways in human visual cortex and in our DNN model.

Together, our fMRI results show that integrative processing in scene-selective cortex and, specifically in the PPA, predicts the aesthetic appeal of natural images. Together with the DNN analyses, our fMRI data thus provide converging evidence that spatial integration is a critical computation that predisposes perceived beauty.

DISCUSSION

Our study unveils that integrative processing, as a computational principle in the visual system, is predictive of the perceived beauty of natural images. We show that a quantification of integration in an artificial neural network can reliably predict beauty ratings for natural scene images across different scene image sets and across different stimulus presentation times. Even when extrapolating the analysis to a widely varying set of natural images (containing objects, people, and scenes), integration in the network could predict perceived beauty. In an fMRI analysis, we show that an analogous quantification of integration in the human visual system can equally predict beauty ratings for natural scenes. Together, these results provide critical evidence for the notion that efficient processing in sensory systems, as enabled by efficient spatial integration, is linked to the aesthetic evaluation of sensory inputs.

This notion fits well with theories that explain perceived beauty through processing fluency in brain systems (*13*, *14*, *56*). Efficient integration has been previously related to a decrease in neural competition between image elements, both in simple (*23*, *38*) and naturalistic (*22*, *34*) visual stimuli. On this note, integration reduces interference between image elements by aggregating them into fewer compound representations (*22*, *24*) and thereby increases the ease with which these fewer representations can be processed in parallel. Integrative processing may of course not constitute the only computational principle that determines processing fluency, and thereby perceived beauty. Uncovering other complementary principles may further increase the amount of variance in beauty ratings that can be predicted from sensory-derived measures of stimulus processing. Our findings support the idea of processing fluency theory that higher fluency relates to increased beauty. However, how do they relate to the PIA model (*15*), where the discrepancy between expected and experienced fluency is the critical determinant of perceived beauty? Our pretrained DNN models also form statistical "expectations" about inputs, based on the distribution of visual features in the training set. A discrepancy between expected and experienced fluency may simply result from how efficiently information is integrated for a given image, compared to images experienced during training. The PIA model also stresses the cognitive control factors that drive human interest in an image. While these cognitive factors are not captured by our modeling approach, they could be implemented in the future by devising models which explicitly include top-down control processes.

The finding that integrative processing predicts perceived beauty is also interesting because integrative processing constitutes a computational principle that can operate across various features. Identifying these computational principles has the potential to reconcile the somewhat fragmented literature on visual feature preferences. It will be interesting to see whether favorable levels of stimulus complexity or density (*8*, *57*, *58*), the spatial regularity of visual arrangements (*8*, *59*), or even the typical compositions of artworks (*5*, *60*, *61*) can be related to differences in integrative processing. If they

can, differing degrees of integrative processing will provide explanations for why particular types of stimulus configurations are perceived as beautiful—because they are more readily integrated into a meaningful whole in the visual system. In this context, our DNN-based integration measure can be readily used to evaluate various types of stimuli, from arrays of basic visual shapes to real-world objects, and even visual art.

Our approach shows that integration predicts beauty ratings across groups of observers (i.e., the shared taste across a group of people). However, research on individual differences indicates that ratings of aesthetic appeal can be highly personal and vary substantially across observers in meaningful ways (*62*–*64*). For example, Vessel and colleagues (*64*) showed that only between 66% (faces), 29% (landscapes), and 11% (architecture) of the variance in beauty ratings is explained by shared taste. Recent electroencephalography results show that such idiosyncrasies in perceived beauty are already reflected in early cortical responses linked to the perceptual analysis of faces (*65*) and scenes (*12*). While the amount of shared and personal taste varies across stimulus domains and image sets, honoring the interindividual variability in beauty ratings clearly promises to enhance image-based predictions of perceived beauty. In future studies, such individual differences could directly be assessed by obtaining integrative processing measures in the fMRI and beauty ratings for the same participants: Such data would allow for linking individual differences in integrative processing with individual differences in perceived beauty.

Our DNN results further highlight that integration effects at different spatial scales can predict perceived beauty. This suggests that the efficiency of sensory processing that is critical for rating aesthetic appeal is computed at different levels of detail, from the integration of local spatial features that do not necessarily transport semantic information to the integration of larger regions that unites semantically meaningful image elements. When systematically manipulating the input to the DNN by changing image properties, we found that neither basic characteristics like color and luminance of an image nor configural high-level properties like scene orientation (*51*–*53*) or coarse spatial structure (*52*, *54*) reduced the predictions of beauty from our integration measure. This is interesting, because it suggests that the integration of features of intermediate complexity drives the prediction of beauty. What these features could be is an interesting question for future research. We further discuss this issue below, based on our understanding of representations in scene-selective PPA. Beyond these considerations, it will be interesting to see whether integration effects at different scales relate to different processing steps in the visual hierarchy, with varying sensitivity to visual and semantic information. An interesting future avenue would be to not only compute the integration of purely visual information: For instance, future studies could also look at conceptual descriptors, for example using language models (*66*, *67*), to capture information integration in conceptual space for visual images.

Our fMRI results show that integrative processing in the biological brain can predict perceived beauty in a similar way as integrative processing in DNNs. Given recent reports of critical differences between visual processing in humans and DNNs (*28*–*30*), the similarity in beauty predictions derived from a DNN and human fMRI data, as well as the similarity between the integrative processing measures in PPA and a scene-trained DNN, provides a critical test of the alignment between artificial and biological vision. Our findings further stress the mechanistic similarity between DNNs and the human visual system (*27*, *68*, *69*) and showcases the possibility to use DNNs as high-throughput tools for uncovering the computational principles that guide visual processing (*26*, *70*). Our focus is different from previous computational approaches that probe and optimize the prediction of beauty from activation patterns in DNNs (*71*–*74*) (also see fig. S9 for a prediction-based analysis on all DNN features in the current study). The critical difference to approaches that aim to maximize predictive accuracy is that these approaches typically do not yield deep mechanistic insights into why DNN activations allow for predicting beauty (beyond charting which training regimes or network layers contribute to accurate prediction). Our approach, by contrast, focuses on a single, interpretable predictor computed from network activations. We believe that such an approach opens the door to exploring, and subsequently validating, computational principles in the biological brain, including those that govern the sensory stages of aesthetic appreciation.

In the human visual system, the beauty of natural scene images was only robustly predicted from responses in scene-selective PPA. Previous studies indicate that PPA is capable of integrating scene elements distributed across visual space, specifically when the scenes form a meaningful perceptual whole (*52*, *75*, *76*). Here, we show that such integration processes systematically vary across scenes and that this variation is a determinant for a scene's aesthetic appeal. An open question concerns what exactly is integrated in PPA. Processing in this region has previously been linked to both categorical and semantic attributes of a scene (*53*, *77*, *78*), as well as to low- and mid-level visual properties that are reliably associated with scenes (*79*–*82*). Our DNN analysis suggest that both mid- and high-level properties computed at intermediate to deep layers may be critical for determining perceived beauty, and PPA activations have indeed been linked to activations in intermediate to deep DNN layers before (*43*, *83*). The results from our image manipulation analysis suggest that predictions of beauty cannot be pinpointed to low- or high-level scene features in a straightforward way. Integration across high spatial frequencies was a more powerful predictor of beauty than integration across low spatial frequencies. This is consistent with a recent report that high spatial frequencies drive scene-selective cortex more strongly than low spatial frequencies under comparable contrast statistics (*84*). That being said, integration across low spatial frequencies still significantly predicted perceived beauty (see Fig. 3C). Delineating the types of mid-level features whose integration determined beauty currently remains an open issue: While we know that mid-level features are a critical driver of high-level visual cortex responses (*82*, *85*), what exactly these mid-level features entail is largely unclear. At this point, more research is needed to uncover the critical features across which the PPA integrates information. One notable limitation of the current fMRI study is that we could only evaluate a small subset of images, which does not allow us to systematically scrutinize the features that drive integration effects in human visual cortex.

Besides the PPA, our data show substantial correlations between integration and beauty ratings in V1 to V4, which perhaps did not reach statistical significance given the limited number of stimuli. It would thus be premature to dismiss a link between integrative processing in early visual cortex (i.e., the integration of simple visual features) and perceived beauty. The relatively weak and partly inconsistent effects in early DNN layers as well as the limited influence of color, luminance, and contrast on predictions, however, argue against an extensive influence of simple visual feature integration

on beauty ratings. Integration in the scene-selective OPA and MPA, in contrast to PPA, did not predict perceived beauty. This highlights processing differences across the scene-selective network in visual cortex (*77*). The MPA is primarily associated with environmental analysis relevant for navigation (*77*) and may thus not be driven very strongly by our stimuli and task (see also the low univariate response in this region; fig. S10). The OPA is often associated with the analysis of local scene elements (*83*, *86*) or the view-specific representation of scenes (*87*–*89*). The finding that integrative processing in PPA, but not OPA, predicts perceived beauty may thus indicate that the integration of more complex features than those coded in the OPA is most critically related to an image's aesthetic appeal.

Last, our study quantified integration across visual space. Yet, space is not the only dimension across which efficient integration can mediate perceived beauty. For example, a recent study used DNN models to quantify the integration of visual features across hierarchical levels to model aesthetic perception and has, in turn, linked such hierarchical integration processes to parietal and frontal brain systems (*90*). Future studies could also link perceived beauty to integration across time: Recent studies in neuroaesthetics increasingly focus on more naturalistic and dynamic stimuli (*11*, *91*–*93*), and it will be interesting to see whether efficient information integration in the time domain can also predict the aesthetic appeal of dynamically evolving visual scenes.

In sum, our study establishes integrative processing as a computational principle in the visual system that is capable of explaining perceived beauty. When images are more strongly integrated across visual space—and consequentially the representation of the whole gets more dissimilar to the representation of its parts—they are perceived as more beautiful. Gestalt psychologists famously noted that "the whole is something else than the sum of its parts" (*94*). Here, we show that not only the degree to which the whole is different from its parts has an impact on the formation of efficient representations the whole but also that this process is linked to whether or not we assign beauty to it. Our discovery provides a fresh impulse for research in neuroaesthetics: Moving from the study of individual visual features and their impact on aesthetic appeal toward the study of overarching computational principles has the potential to reconcile research on different visual features and stimulus domains.

MATERIALS AND METHODS

Participants

We ran three online studies in which participants rated the beauty of natural scene images. Study 1 was completed by 25 participants (mean age, 23.9; SD = 5.2; 18 male, 7 female), study 2 was completed by 25 participants (mean age, 24.0; SD = 4.3; 10 male, 15 female), and study 3 was completed by 26 participants (mean age, 24.9; SD = 5.2; 12 male, 13 female, 1 nonbinary). Participants were recruited through Prolific (www.prolific.co) and received monetary reimbursement. Informed consent was provided through an online form. The studies were approved by the Ethical Committee of the Institute of Psychology at the University of York (approval reference 20228).

We additionally ran an fMRI study, where participants viewed whole and partial natural images. The fMRI study was completed by 22 participants (mean age, 28.8; SD = 4.5; 10 male, 12 female). One participant did not complete all experimental runs and was excluded from further analyses. Participants received monetary reimbursement and provided written informed consent. The study was approved by the General Ethical Committee of the Justus Liebig University Gießen (approval reference AZ 25/22).

All participants were healthy adults with normal or corrected-to-normal vision. Sample sizes were determined through convenience sampling, as data were subsequently averaged across participants for our analyses. Reliability measures showed good agreement between participants (see below).

Rating study design

Three behavioral studies were conducted online, using the Gorilla testing platform (*95*). In study 1, participants viewed 250 images of natural scene photographs in random order, sampled from the validation set of the Places365 dataset (*32*) to include a large variety of contents. All images depicted outdoor scenes. On every trial, they viewed a single scene, presented for 50 ms. After the scene presentation, they were presented with a slider, operated by the mouse, on which they adjusted how beautiful they rated the scene. Slider values were coded between 0 and 100. Before the experiment, beauty was defined to participants as how beautiful or aesthetically pleasing the scene is. After participants provided their ratings, participants could advance to the next trial by pressing a mouse button.

In study 2, participants viewed the same 250 images as in study 1, in random order. The study design was identical to study 1, but here, the slides was shown at the same time as the scene (below the image). The scene was visible until participants provided their rating. Participants were not instructed to respond fast and were given as much time as needed to adjust the slider.

In study 3, participants viewed another set of 250 images in random order, sampled similarly to studies 1 and 2, but the image set was completely disjunct from the previous set. The design was identical to study 2, but on every trial, the scene was accompanied by three sliders, on which they could adjust: (i) how complex they rated the scene, (ii) how ordered they rated the scene, and (iii) how beautiful they rated the scene. Complexity was defined as the number of distinct elements (such as objects, shapes, or colors) present in the scene, compared to how many such items are expected in a typical scene. Order was defined as the degree to which the different scene elements (such as objects) are positioned across the scene and relative to each other in a typically structured manner. Both complexity and order predicted beauty ratings linearly (see Results).

Database ratings

In addition to the three behavioral studies, we used beauty ratings collected for the images in the OASIS database (*46*). This database contains 900 photographs depicting a large variety of contents, from people to objects and scenes. Beauty ratings for these images from a total of 757 observers were compiled by Brielmann and Pelli (*45*). In their study, each image was rated by at least 104 observers.

Behavioral data analysis

For studies 1 to 3, as well as the OASIS beauty ratings, a mean score for each image was computed from the ratings of all observers. Beauty ratings were reliable across people in studies 1 to 3, as shown by split-half correlations ($r > 0.89$ for all studies). Beauty ratings were also highly correlated between studies 1 and 2, which used the same stimuli under brief and unlimited exposure times ($r = 0.88$).

Detailed reliability measures for the OASIS ratings are reported elsewhere (*45*), with split-half reliability $r > 0.95$.

DNN analysis

We used a VGG16 (*31*) DNN trained on scene categorization using the Places365 image set (*32*). The VGG architecture was chosen as it has been shown to provide a good computational approximation of the ventral visual pathway (*96*). The pretrained DNN was obtained from https://github.com/CSAILVision/places365. The network was originally deployed in Caffe (*97*) and imported to MATLAB using the Caffe Importer for the MATLAB Deep Learning Toolbox (https://de.mathworks.com/matlabcentral/fileexchange/61735-deep-learning-toolbox-importer-for-caffe-models). Results from a VGG16 network trained on Imagenet (*98*) and implemented in MATLAB, are reported in fig. S2.

We fed the network with either the full scene or two halves of the scene. Halves were created by slicing the scene into 2 × 2, 4 × 4, 8 × 8, 16 × 16, or 32 × 32 identical squares. Each image half contained all odd or all even squares (i.e., corresponding to either all white or black squares on a checkerboard). This slicing yielded two image halves for each of five different spatial scales. The full image and all possible halves were fed to the network, and layer-specific activation patterns for each image were obtained separately for all 16 layers of the network.

To quantify integration, we correlated the layer-specific activation patterns to the whole scene with the average response pattern to the scene halves, separately for each spatial scale (e.g., by slicing into 2 × 2 pieces). Note that any linear combination with equal weights (e.g., the sum) would yield equivalent correlations.

The sign of the resulting correlations was flipped to yield a measure of integration: When processing is largely parallel, activation patterns to the full image should be accurately predicted by the average of the activation patterns to the two halves (resulting in a low integration measure), as previously shown in human cortex (*33*, *35*–*37*) and for DNNs (*39*–*41*). Unlike in human cortex, averaging the response patterns to constituent objects does not perfectly predict the response to multiple objects (*40*); in our analysis, however, the overall quality of the fit is not critical, as we only examine relative differences in the fit across images. When processing is more integrative, activation patterns to the full image should be less accurately predicted by the average of the activation patterns to the two halves (resulting in a higher integration measure). Measuring integration through such multivariate pattern combination analysis has successfully been used in fMRI work (*24*, *34*, *38*, *99*). In sum, our procedure thus yielded a quantification of integration for each image, at each spatial scale, and in each network layer.

We also tested whether different image partitioning for quantifying integration yielded similar results as our checkerboard split. Performing the analysis with image halves based on a random selection of parts on the 4 × 4 and 8 × 8 spatial scales yielded similar results as the checkerboard split (figs. S11 and S12).

We additionally computed a part-based similarity measure as an alternative predictor for perceived beauty. To quantify part-based similarity, we correlated the activation pattern to one image half with the activation pattern to the other half, separately for each spatial scale and each network layer. This correlation can be directly interpreted as a measure of part-based similarity, where higher correlations signal greater visual correspondence between the image halves.

Last, we computed how much each whole image drives the DNN, as another alternative predictor for perceived beauty. Here, we computed the L2 norm for each network layer as a measure of activation strength.

Raw values of the integration measure, the part-based similarity measure, and the L2 activation measure across DNN layers are reported in fig. S13. Intercorrelations between the measures are reported in fig. S14.

Predicting beauty ratings from DNN integration

To assess how beauty rating were predicted by integration in the DNN, we correlated (Spearman correlations) the image-specific mean beauty ratings for each of the four studies with the image-specific quantifications of integration, separately for each network layer. The same analysis was performed for the part-based similarity measure, as reported in fig. S3. All *P* values corresponding to these correlations were false-discovery-rate (FDR) corrected across network layers and spatial scales.

To assess the unique contribution of integration in predicting beauty ratings, we also performed partial correlation analyses, where either the part-based similarity measure or the activation strength for the whole image was partialed out when correlating the beauty ratings with the integration measure. Partial correlations provide a way of assessing the association between two variables while removing the contribution of a third variable to this association. In our context, the resulting partial correlations index how well beauty ratings are predicted by the integration measure if we account for the effect of either the part-based similarity or the activation strength for the whole image. These analyses are reported in figs. S3 and S4. A similar partial correlation analysis was conducted using the complexity and order ratings in study 3. Here, the association between the integration measure and the beauty ratings was assessed while controlling for the complexity and order ratings (fig. S1).

To better understand how the measures jointly predict beauty ratings, we also performed a variance partitioning analysis, based on an implementation in the Net2Brain toolbox (*100*), where the unique variance accounted for by each predictor as well as each combination of predictors was examined in a set of linear models (*43*, *44*). Specifically, we constructed linear models that contained one of the predictors, each possible combination of two predictors, or all three predictors. By subtracting the variance explained (adjusted R^2) by the full model from the variance explained by reduced models, we could gauge the unique variance explained by the reduced model. The resulting decomposition of the variance is illustrated in fig. S6.

We also assessed whether the DNN integration measure could successfully predict beauty ratings for novel images. To this end, we fit a linear model with all layers included as predictors for the beauty ratings for all images but one. Both the criterion and predictors were *z*-scored before estimating the regression weights. We then derived a predicted beauty rating for the left-out image from the estimated linear model. Repeating this procedure for all images being left out once yielded a predicted value for each image. To assess how well the model could predict the beauty ratings for the held-out images, we correlated the predicted beauty ratings with the real beauty ratings obtained from our human observers.

Manipulating visual image properties

To assess which features contribute to successful prediction of perceived beauty by our integration measure, we systematically manipulated the visual properties of the images supplied to the DNN (see Fig. 3A). In a first series of manipulations, we (i) stripped away color of the image by gray-scaling each image, (ii) additionally equated the luminance of each image using the luminance equation algorithm of the Spectrum, Histogram, and Intensity Normalization and Equalization toolbox (*48*), (iii) additionally equated the contrast of the images using the histogram matching algorithm of the SHINE toolbox (*48*), and (iv) additionally pixelated the image by shuffling all pixels in the 640 × 640 pixel images. Second, we assessed the contribution of spatial frequency content by frequency filtering the images. High-pass filtering was performed above a cutoff of 80 cycles per image, and low-pass filtering was performed below the same cutoff. Last, we manipulated the spatial configuration of the image by rotating the image by 90° (clockwise) or 180° or by jumbling the image in a crisscrossed way, exchanging the upper-left and lower-right and well as the upper-right and lower-left quadrants.

We then computed the same analysis as before, first correlating the integration measure across layers with the beauty ratings and then assessing predictions of a linear model trained on the integration measure across all layers (see Fig. 2). The impact of the image manipulations was assessed on the linear model predictions to reduce the complexity of the result patterns (avoiding excessive comparisons across layers). Differences between correlations were assessed using z-tests (*101*, *102*). The resulting *P* values were FDR corrected across the different image manipulations.

fMRI study

During the fMRI experiment, participants viewed 32 scene images, which were a subset of the stimulus set used in study 3. These scenes were chosen to come from four broadly defined categories (beaches, buildings, highways, and mountains) and to cover a range of beauty ratings (mean average rating: 66/100; minimum: 35; maximum: 87). The scenes were cropped so that all images had square aspect ratio and resized to 512 × 512 pixels. In each of 10 fMRI runs (4.5 min each), participants completed 128 trials (1280 in total). In each trial, they saw a scene image (7.5° × 7.5° visual angle) for 500 ms, followed by an intertrial interval of 1500 ms, during which a black fixation cross was shown. During each run, each scene was shown in three possible conditions: whole image (32 trials per run), top-left and bottom-right quadrants only (32 trials per run), or top-right and bottom-left quadrants only (32 trials per run). These conditions thus corresponded to the 2 × 2 spatial scale in the DNN analysis. Each run additionally featured 32 fixation trials, during which the black fixation cross turned gray. Participants were instructed to press a button on these trials. Stimulus presentation was controlled using the Psychtoolbox (*103*).

fMRI acquisition and preprocessing

MRI data was acquired using a 3T Siemens Magnetom PRISMA Scanner equipped with a 64-channel head coil. T2*-weighted gradient-echo echo-planar images were collected as functional volumes (TR = 1850 ms, TE = 30 ms, 75° flip angle, 2.2-mm^3 voxel size, 58 slices, 20% gap / distance factor, 220-mm field of view, 100 × 100 matrix size, interleaved acquisition). In addition, a T1-weighted image (Magnetization Prepared Rapid Acquisition with Gradient Echoes; 1-mm^3 voxel size) was obtained as a high-resolution anatomical reference. During preprocessing, the functional volumes were realigned and coregistered to the T1 image using SPM12 (www.fil.ion.ucl.ac.uk/spm/). The functional data were then modeled using a general linear model (GLM) with separate predictors for the 32 images and the three presentation conditions (the whole image and the two parts), separately for each run. The GLM also contained six movement regressors obtained during realignment and thus 102 regressors for each run.

fMRI analyses

Multivariate analyses were performed using the CoSMoMVPA toolbox (*104*). Multivoxel response patterns were obtained for seven regions of interest. Four early visual cortex regions were defined using a template atlas (*105*): V1, V2, V3, and V4. In addition, three scene-selective regions were defined using functional group maps (*106*): the OPA (also termed transverse occipital sulcus), the MPA (also termed retrosplenial cortex), and the PPA. For each region, multivoxel response patterns were extracted by unfolding the GLM beta weights for each image and each run into a one-dimensional vector.

We then performed an analysis similar to the DNN analysis. For each region, we averaged response patterns evoked by the two halves and correlated the resulting response pattern to the response pattern evoked by the full image, separately for each image. Here, the response pattern to the whole scene always stemmed from half of the fMRI runs and the average response pattern to the halves stemmed from the other half of the runs (*34*). Correlations were computed across all possible 50/50 splits among the 10 runs and averaged across splits. By flipping the sign of the resulting correlations, we obtained a quantification of integration, where lower correlations index a higher degree of integration. These values were averaged across participants before comparing them to the beauty ratings.

From the fMRI data, we additionally obtained a quantification of response reliability for the whole images. This was done by correlating the response pattern to each whole image in half of the runs with the response pattern to the same image in the other half of the runs. Correlations were again averaged across all possible 50/50 splits among the 10 runs. This yielded a correlation for each whole image that indexed the stability of the response patterns across repetitions.

We additionally constructed two alternative predictors: First, we constructed a part-based similarity measure by correlating the fMRI activation pattern for one half of the image with the activation pattern for the other half of the image, separately for each region. As for the integration measure, the activation pattern for each half was computed by averaging the corresponding patterns from half of the runs. Correlations were thus always performed across runs (see above). Second, we constructed an overall activation measure (akin to the L2 measure for the DNN) by extracting the univariate activation for each whole image in each of the regions. Raw values of the integration measure, the part-based similarity measure, and the univariate activation measure across regions are reported in fig. S10.

Predicting beauty ratings from fMRI integration

To assess how beauty ratings were predicted by integration in the human visual cortex, we correlated (Spearman correlations) the image-specific mean beauty ratings for the 32 images used in the fMRI study with the image-specific fMRI quantifications of integration, separately for each brain region. All *P* values corresponding to these correlations were FDR corrected across regions.

We additionally repeated this analysis while partialing out the quantification of reliability obtained for each image. In this way, we could ensure that differences in integration (i.e., differences in how well the combined response to the image halves predicted the response to the whole image) were not simply resulting from differences in the reliability of responses to each of the individual images (where less reliable response would be harder to approximate in the first place). Unlike the integration measure, the part-based similarity measure and the univariate activation measure did not significantly predict beauty ratings (fig. S15).

Supplementary Materials

This PDF file includes:

Figs. S1 to S15

REFERENCES AND NOTES

1. H. Leder, B. Belke, A. Oeberst, D. Augustin, A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* **95**, 489–508 (2004).
2. H. Leder, M. Nadal, Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode–Developments and challenges in empirical aesthetics. *Br. J. Psychol.* **105**, 443–464 (2014).
3. C. Redies, Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* **9**, 218 (2015).
4. M. Enquist, A. Arak, Symmetry, beauty and evolution. *Nature* **372**, 169–172 (1994).
5. S. E. Palmer, K. B. Schloss, J. Sammartino, Visual aesthetics and human preference. *Annu. Rev. Psychol.* **64**, 77–107 (2013).
6. K. B. Schloss, S. E. Palmer, Aesthetic response to color combinations: Preference, harmony, and similarity. *Atten. Percept. Psychophys.* **73**, 551–571 (2011).
7. P. J. Silvia, C. M. Barona, Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empir. Stud. Arts* **27**, 25–42 (2009).
8. E. Van Geert, J. Wagemans, Order, complexity, and aesthetic appreciation. *Psychol. Aesthet. Creat. Arts* **14**, 135–154 (2020).
9. E. A. Vessel, A. I. Isik, A. M. Belfi, J. L. Stahl, G. G. Starr, The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proc. Natl. Acad. Sci.* **116**, 19155–19164 (2019).
10. X. Yue, E. A. Vessel, I. Biederman, The neural basis of scene preferences. *Neuroreport* **18**, 525–529 (2007).
11. X. Zhao, J. Wang, J. Li, G. Luo, T. Li, A. Chatterjee, W. Zhang, X. He, The neural mechanism of aesthetic judgments of dynamic landscapes: An fMRI study. *Sci. Rep.* **10**, 20774 (2020).
12. D. Kaiser, Characterizing dynamic neural representations of scene attractiveness. *J. Cogn. Neurosci.* **34**, 1988–1997 (2022).
13. M. Forster, in *The Oxford Handbook of Empirical Aesthetics,* M. Nadal, O. Vartanian, Ed. (Oxford Univ. Press, 2020), pp. 430–446.
14. R. Reber, N. Schwarz, P. Winkielman, Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* **8**, 364–382 (2004).
15. L. K. Graf, J. R. Landwehr, A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Pers. Soc. Psychol. Rev.* **19**, 395–410 (2015).
16. D. E. Broadbent, *Perception and Communication* (Pergamon Press, 1958).
17. R. Desimone, J. Duncan, Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
18. R. Marois, J. Ivanoff, Capacity limits of information processing in the brain. *Trends Cogn. Sci.* **9**, 296–305 (2005).
19. S. Kastner, P. De Weerd, R. Desimone, L. G. Ungerleider, Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* **282**, 108–111 (1998).
20. E. K. Miller, P. M. Gochin, C. G. Gross, Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.* **616**, 25–29 (1993).
21. E. T. Rolls, M. J. Tovee, The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Exp. Brain Res.* **103**, 409–420 (1995).
22. D. Kaiser, T. Stein, M. V. Peelen, Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc. Natl. Acad. Sci.* **111**, 11217–11222 (2014).
23. S. A. McMains, S. Kastner, Defining the units of competition: Influences of perceptual organization on competitive interactions in human visual cortex. *J. Cogn. Neurosci.* **22**, 2417–2426 (2010).
24. D. Kaiser, G. L. Quek, R. M. Cichy, M. V. Peelen, Object vision in a structured world. *Trends Cogn. Sci.* **23**, 672–685 (2019).
25. J. M. Wolfe, G. A. Alvarez, R. Rosenholtz, Y. I. Kuzmova, A. M. Sherman, Visual search for arbitrary objects in real scenes. *Atten. Percept. Psychophys.* **73**, 1650–1671 (2011).
26. R. M. Cichy, D. Kaiser, Deep neural networks as scientific models. *Trends Cogn. Sci.* **23**, 305–317 (2019).
27. N. Kriegeskorte, Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015).
28. J. S. Bowers, G. Malhotra, M. Dujmovic, M. L. Montero, C. Tsvetkov, V. Biscione, G. Puebla, F. Adolfi, J. E. Hummel, R. F. Heaton, B. D. Evans, J. Mitchell, R. Blything, Deep problems with neural network models of human vision. *Behav. Brain Sci.* **46**, e385 (2022).
29. K. Hermann, T. Chen, S. Kornblith, The origins and prevalence of texture bias in convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **33**, 19000–19015 (2020).
30. Y. Xu, M. Vaziri-Pashkam, Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nat. Commun.* **12**, 2065 (2021).
31. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs.CV] (4 September 2014).
32. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 1452–1464 (2017).
33. S. K. Jeong, Y. Xu, Task-context-dependent linear representation of multiple visual objects in human parietal cortex. *J. Cogn. Neurosci.* **29**, 1778–1789 (2017).
34. D. Kaiser, M. V. Peelen, Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage* **169**, 334–341 (2018).
35. D. Kaiser, L. Strnad, K. N. Seidl, S. Kastner, M. V. Peelen, Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J. Neurophysiol.* **111**, 82–90 (2014).
36. L. Kliger, G. Yovel, The functional organization of high-level visual cortex determines the representation of complex visual stimuli. *J. Neurosci.* **40**, 7545–7558 (2020).
37. S. P. MacEvoy, R. A. Epstein, Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* **19**, 943–947 (2009).
38. J. Kubilius, A. Baeck, J. Wagemans, H. P. O. de Beeck, Brain-decoding fMRI reveals how wholes relate to the sum of parts. *Cortex* **72**, 5–14 (2015).
39. G. Jacob, R. Pramod, H. Katti, S. Arun, Qualitative similarities and differences in visual object representations between brains and deep networks. *Nat. Commun.* **12**, 1872 (2021).
40. V. Mocz, S. K. Jeong, M. Chun, Y. Xu, Multiple visual objects are represented differently in the human brain and convolutional neural networks. *Sci. Rep.* **13**, 9088 (2023).
41. A. Jha, S. Agarwal, in *Applied Cognitive Science and Technology: Implications of Interactions Between Human Cognition and Technology*. (Springer, 2023), p. 123–138.
42. S. Verhavert, J. Wagemans, M. D. Augustin, Beauty in the blink of an eye: The time course of aesthetic experiences. *Br. J. Psychol.* **109**, 63–84 (2018).
43. I. I. Groen, M. R. Greene, C. Baldassano, L. Fei-Fei, D. M. Beck, C. I. Baker, Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *eLife* **7**, e32962 (2018).
44. K. Ramakrishnan, H. S. Scholte, I. I. Groen, A. W. Smeulders, S. Ghebreab, Visual dictionaries as intermediate features in the human brain. *Front. Comput. Neurosci.* **8**, 168 (2015).
45. A. A. Brielmann, D. G. Pelli, Intense beauty requires intense pleasure. *Front. Psychol.* **10**, 2420 (2019).
46. B. Kurdi, S. Lozano, M. R. Banaji, Introducing the open affective standardized image set (OASIS). *Behav. Res. Methods* **49**, 457–470 (2017).
47. D. M. Beck, S. Kastner, Stimulus similarity modulates competitive interactions in human visual cortex. *J. Vis.* **7**, 19.1–19.12 (2007).
48. V. Willenbockel, J. Sadr, D. Fiset, G. O. Horne, F. Gosselin, J. W. Tanaka, Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).
49. L. Kauffmann, S. Ramanoël, C. Peyrin, The neural bases of spatial frequency processing during scene perception. *Front. Integr. Neurosci.* **8**, 37 (2014).
50. A. Oliva, A. Torralba, Chapter 2 Building the gist of a scene: The role of global image features in recognition. *Prog. Brain Res.* **155**, 23–36 (2006).
51. R. A. Epstein, J. S. Higgins, W. Parker, G. K. Aguirre, S. Cooperman, Cortical correlates of face and scene inversion: A comparison. *Neuropsychologia* **44**, 1145–1158 (2006).
52. D. Kaiser, G. Haberle, R. M. Cichy, Cortical sensitivity to natural scene structure. *Hum. Brain Mapp.* **41**, 1286–1295 (2020).
53. D. B. Walther, E. Caddigan, L. Fei-Fei, D. M. Beck, Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* **29**, 10573–10581 (2009).
54. D. Kaiser, G. Häberle, R. M. Cichy, Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *J. Neurophysiol.* **124**, 145–151 (2020).

55. C. Damiano, J. Wilder, E. Y. Zhou, D. B. Walther, J. Wagemans, The role of local and global symmetry in pleasure, interest, and complexity judgments of natural scenes. *Psychol. Aesthet. Creat. Arts* **17**, 322–337 (2023).
56. D. M. Oppenheimer, The secret life of fluency. *Trends Cogn. Sci.* **12**, 237–241 (2008).
57. D. E. Berlyne, *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation* (Hemisphere, 1974).
58. J. Friedenberg, B. Liby, Perceived beauty of random texture patterns: A preference for complexity. *Acta Psychol. (Amst)* **168**, 41–49 (2016).
59. R. Hübner, M. G. Fillinger, Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Front. Psychol.* **7**, 335 (2016).
60. P. Cavanagh, The artist as neuroscientist. *Nature* **434**, 301–307 (2005).
61. C. D. Green, All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception* **24**, 937–968 (1995).
62. J. Hönekopp, 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).
63. H. Leder, J. Goller, T. Rigotti, M. Forster, Private and shared taste in art and face appreciation. *Front. Hum. Neurosci.* **10**, 155 (2016).
64. E. A. Vessel, N. Maurer, A. H. Denker, G. G. Starr, Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition* **179**, 121–131 (2018).
65. D. Kaiser, K. Nyga, Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Sci. Rep.* **10**, 16852 (2020).
66. M. F. Bonner, R. A. Epstein, Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nat. Commun.* **12**, 4081 (2021).
67. T. R. Hayes, J. M. Henderson, Looking for semantic similarity: What a vector-space model of semantics can tell us about attention in real-world scenes. *Psychol. Sci.* **32**, 1262–1270 (2021).
68. T. C. Kietzmann, P. McClure, N. Kriegeskorte, Deep neural networks in computational neuroscience. *Neuroscience* (2019); https://doi.org/10.1093/acrefore/9780190264086.013.46.
69. G. W. Lindsay, Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *J. Cogn. Neurosci.* **33**, 2017–2031 (2021).
70. F. A. Wichmann, R. Geirhos, Are deep neural networks adequate behavioral models of human visual perception? *Annu. Rev. Vis. Sci.* **9**, 501–524 (2023).
71. C. Conwell, D. Graham, E. A. Vessel, The Perceptual Primacy of Feeling: Affectless machine vision models robustly predict human visual arousal, valence, and aesthetics. PsyArXiv (2021). https://osf.io/preprints/psyarxiv/5wg4s.
72. X. Jin, D. Zou, L. Wu, G. Zhao, X. Li, X. Zhang, B. Zhou, S. Ge, X. Zhou, in *27th ACM International Conference on Multimedia* (2019), pp. 311–319.
73. X. Lu, Z. Lin, H. Jin, J. Yang, J. Z. Wang, in *ACM Conference on Multimedia* (2014), pp. 457–466.
74. C. I. Seresinhe, T. Preis, H. S. Moat, Using deep learning to quantify the beauty of outdoor places. *R. Soc. Open Sci.* **4**, 170170 (2017).
75. L. Chen, R. M. Cichy, D. Kaiser, Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. *Sci. Adv.* **9**, eadi2321 (2023).
76. D. Kaiser, R. M. Cichy, Parts and Wholes in Scene Processing. *J. Cogn. Neurosci.* **34**, 4–15 (2021).
77. R. A. Epstein, C. I. Baker, Scene Perception in the Human Brain. *Annu. Rev. Vis. Sci.* **5**, 373–397 (2019).
78. A. Harel, D. J. Kravitz, C. I. Baker, Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cereb. Cortex* **23**, 947–957 (2013).
79. S. Nasr, C. E. Echavarria, R. B. Tootell, Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *J. Neurosci.* **34**, 6721–6735 (2014).
80. D. M. Watson, T. Hartley, T. J. Andrews, Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex. *Cortex* **92**, 162–174 (2017).
81. D. M. Watson, M. Hymers, T. Hartley, T. J. Andrews, Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage* **124**, 107–117 (2016).
82. D. D. Coggan, D. H. Baker, T. J. Andrews, Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *Eur. J. Neurosci.* **49**, 1587–1596 (2019).
83. K. Dwivedi, R. M. Cichy, G. Roig, Unraveling representations in scene-selective brain regions using scene-parsing deep neural networks. *J. Cogn. Neurosci.* **33**, 2032–2043 (2021).
84. D. Berman, J. D. Golomb, D. B. Walther, Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *PLOS ONE* **12**, e0189828 (2017).
85. B. Long, C.-P. Yu, T. Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9015–E9024 (2018).
86. F. S. Kamps, J. B. Julian, J. Kubilius, N. Kanwisher, D. D. Dilks, The occipital place area represents the local elements of scenes. *Neuroimage* **132**, 417–424 (2016).
87. C. M. Jones, J. Byland, D. D. Dilks, The occipital place area represents visual information about walking, not crawling. *Cereb. Cortex* **33**, 7500–7505 (2023).
88. D. Kaiser, J. Turini, R. M. Cichy, A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife* **8**, e48182 (2019).
89. F. S. Kamps, V. Lall, D. D. Dilks, The occipital place area represents first-person perspective motion information through scenes. *Cortex* **83**, 17–26 (2016).
90. K. Iigaya, S. Yi, I. A. Wahle, S. Tanwisuth, L. Cross, J. P. O'Doherty, Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nat. Commun.* **14**, 127 (2023).
91. A. I. Isik, E. A. Vessel, From visual perception to aesthetic appeal: Brain responses to aesthetically appealing natural landscape movies. *Front. Hum. Neurosci.* **15**, 676032 (2021).
92. D. Kaiser, Spectral brain signatures of aesthetic natural perception in the α and β frequency bands. *J. Neurophysiol.* **128**, 1501–1505 (2022).
93. M. van Elk, M. A. Arciniegas Gomez, W. van der Zwaag, H. T. van Schie, D. Sauter, The neural correlates of the awe experience: Reduced default mode network activity during feelings of awe. *Hum. Brain Mapp.* **40**, 3561–3574 (2019).
94. K. Koffka, *Principles of Gestalt Psychology* (Kegan Paul, 1935).
95. A. L. Anwyl-Irvine, J. Massonnie, A. Flitton, N. Kirkham, J. K. Evershed, Gorilla in our midst: An online behavioral experiment builder. *Behav. Res. Methods* **52**, 388–407 (2020).
96. M. Schrimpf, J. Kubilius, H. Hong, N. J. Majaj, R. Rajalingham, E. B. Issa, K. Kar, P. Bashivan, J. Prescott-Roy, K. Schmidt, D. L. K.Yamins, J. J. DiCarlo, Brain-score: Which artificial neural network for object recognition is most brain-like? bioRxiv 407007 [Preprint]. 5 September 2018. https://doi.org/10.1101/407007.
97. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional Architecture for Fast Feature Embedding. *Proceedings of the 22nd ACM International Conference on Multimedia*, New York, NY, 3 November 2014 (2014), pp. 675–678.
98. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database" in *2009 IEEE Conference on computer Vision and Pattern Recognition* (2009), pp. 248–255.
99. C. Baldassano, D. M. Beck, L. Fei-Fei, Human-object interactions are more than the sum of their parts. *Cereb. Cortex* **27**, 2276–2288 (2017).
100. D. Bersch, K. Dwivedi, M. Vilas, R. M. Cichy, G. Roig, Net2Brain: A Toolbox to compare artificial vision models with human brain responses. arXiv:2208.09677 [cs.CV] (20 August 2022).
101. L. Myers, M. J. Sirois, Spearman correlation coefficients, differences between. *Encyclopedia Stat. Sci.*, (2004).
102. X.-L. Meng, R. Rosenthal, D. B. Rubin, Comparing correlated correlation coefficients. *Psychol. Bull.* **111**, 172–175 (1992).
103. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
104. N. N. Oosterhof, A. C. Connolly, J. V. Haxby, CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front. Neuroinform.* **10**, 27 (2016).
105. L. Wang, R. E. Mruczek, M. J. Arcaro, S. Kastner, Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* **25**, 3911–3931 (2015).
106. J. B. Julian, E. Fedorenko, J. Webster, N. Kanwisher, An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* **60**, 2357–2364 (2012).

Acknowledgments

Funding: D.K. is supported by the German Research Foundation (DFG; SFB/TRR135, project number 222641018; KA4683/5-1, project number 518483074); "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art; and an ERC Starting Grant (PEP, ERC-2022-STG 101076057). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. **Author contributions:** Conceptualization: D.K. Data curation: S.N. and D.K. Formal analysis: S.N. and D.K. Funding acquisition: D.K. Investigation: D.K. Methodology: S.N. and D.K. Project administration: D.K. Resources: S.N. and D.K. Software: S.N. and D.K. Supervision: D.K. Validation: S.N. and D.K. Visualization: D.K. Writing—original draft: D.K. Writing—review and editing: S.N. and D.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data, materials, and code can be found at https://zenodo.org/doi/10.5281/zenodo.10277386.

Submitted 26 May 2023
Accepted 29 January 2024
Published 1 March 2024
10.1126/sciadv.adi9294