

The Spontaneous Emergence of “A Sense of Beauty” in Untrained Deep Neural Networks

Tianxin Shu^{1, 2, 3, 4}, Huawei Xu^{1, 2, 3, 4}, Xingxing Chen^{1, 2, 3, 4}, Yuxuan Cai^{1, 2, 3, 4}, Ming Liu^{1, 2, 3, 4}, and Delong Zhang^{1, 2, 3, 4}

¹ Key Laboratory of Brain, Cognition and Education Sciences

² Ministry of Education Department, South China Normal University

³ School of Psychology, Center for Studies of Psychological Application

⁴ Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University

The sense of facial beauty has long been observed in both infants and nonhuman primates, yet the neural mechanisms of this phenomenon are still not fully understood. The current study employed generative neural models to produce facial images of varying degrees of beauty and systematically investigated the neural response of untrained deep neural networks (DNNs) to these faces. Representational neural units for different levels of facial beauty are observed to spontaneously emerge even in the absence of training. Furthermore, these neural units can effectively distinguish between varying degrees of beauty. Additionally, the perception of facial beauty by DNNs relies on both configuration and feature information of faces. The processing of facial beauty by neural networks follows a progression from low-level features to integration. The tuning response of the final convolutional layer to facial beauty is constructed by the weighted sum of the monotonic responses in the early layers. These findings offer new insights into the neural origin of the sense of beauty, arising the innate computational abilities of DNNs.

Keywords: aesthetic neurocomputation, generative neural models, untrained deep neural networks, facial beauty, linear weighted summation


Supplemental materials: <https://doi.org/10.1037/aca0000690.supp>

In both everyday life and scientific research, there has been ongoing debate about beauty. Interestingly, the fundamental questions of the debate have changed little: What is beauty? From philosophy to empirical aesthetics, from neuroaesthetics to computational aesthetics, whether there exists a universally applicable aesthetic theory remains a contentious issue (Seeley, 2013). While beauty lacks a unified definition, a recurring debate revolves around the objectivity or subjectivity of aesthetics (Fedrizzi, 2012). Both perspectives

found support among early Greek philosophers, with the Pythagoreans attempting to prove the objectivity of beauty, while sophists argued for its subjectivity. The objectivity of aesthetics is based on the concept that beauty exists within objects and is therefore measurable (Birkhoff, 1933). When considering faces, various standards for evaluating facial beauty exist, such as the golden ratio (Bashour, 2006), triple symmetry, and the five-eye width ratio (H. K. Zhang et al., 2017). On the other hand, subjective beauty in aesthetics focuses more on human experience and subjective evaluations. The current study aims to encompass the subjectivity of beauty using human-assessed data and enrich the understanding of image attributes influencing human aesthetics (i.e., the objectivity of beauty) through deep neural network (DNN) models. The term “aesthetics” originates from the ancient Greek word “aisthesis,” which can be translated as “sensual perception” (Reicher, 2015). It is evident that aesthetics is closely tied to sensation (Santayana, 1896). A visual-image-based aesthetic model posits that aesthetic perception is a rapid, bottom-up, and universal process (Redies, 2015). Consequently, we characterize “beauty sense” as the capacity or faculty to discern, appreciate, and assess elements of beauty. With this in mind, the question then arises: How is this ability to identify and judge beauty—referred to as “beauty sense”—developed?

For decades, the debate has oscillated between whether our sense of beauty is innate or an acquired predisposition. The argument centered on its association with the acquisition of knowledge (Damon et al., 2017; Sarasso et al., 2020). Recent research unraveling the ties between beauty perception and neural circuitry activation posits that such perception arises from motor-predominant, simultaneous

Dirk B. Walther served as action editor.

Tianxin Shu  <https://orcid.org/0009-0000-2282-9774>

This work was supported by National Natural Science Foundation of China (Grant 31600907) to Delong Zhang. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article. All data needed to evaluate the conclusions in the article are present in the article and/or in the online supplemental materials. The code and data supporting the findings of this study are available from the corresponding author upon reasonable request.

Tianxin Shu served as lead for methodology, formal analysis, and writing—original draft, reviewing, and editing and contributed equally to resources. Huawei Xu served in a supporting role for formal analysis and resources. Xingxing Chen served in a supporting role for formal analysis and resources. Yuxuan Cai served as lead for writing—reviewing and editing. Ming Liu served as lead for supervision. Delong Zhang served as lead for project administration and funding acquisition and contributed equally to methodology.

Correspondence concerning this article should be addressed to Yuxuan Cai, or Delong Zhang, South China Normal University, 55 Zhongshan Avenue West, Tianhe District 510631, Guangzhou City, Guangdong Province, China. Email: y.cai@m.scnu.edu.cn or delong.zhang@m.scnu.edu.cn

perceptual processing within sensory cortices, which are linked to reward-related circuits (Sarasso et al., 2020). The reward-related circuits exhibit complex interaction with learning progresses to seek further knowledge acquisition. These findings have indicated that the neural underpinnings of our perception of beauty span beyond basic sensory processing, encompassing more complex, high-level cognitive domains. However, the processes by which low-level sensory properties innately shape and refine an aesthetic experience are not fully understood.

Presently, leveraging deep generative neural networks to synthesize facial images presents a novel opportunity to quantitatively investigate the neural underpinnings of aesthetic perception. Faces have served as an ideal subject in aesthetics due to their distinctive physiological and evolutionary properties (Geldart, Maurer, & Carney, 1999; Grammer & Thornhill, 1994; Little, 2014). Individuals often use facial beauty as a criterion to assess the potential mate value and health status of an individual (Fink & Penton-Voak, 2002; Thornhill & Gangestad, 1993). Conducting research on facial aesthetics provides an opportunity to gain insights into the complex interactions between biological and psychological factors, as well as aesthetic perception. Human brains may have an innate sense of beauty prior to the acquisition of visual experience given that newborn infants tend to look at attractive faces for a longer period (Slater et al., 1998). In addition, the effect of facial beauty on 6-month-old infants' gaze could span race, gender, and age (Langlois et al., 1991; Slater et al., 1998). However, it has not been determined whether the preference for beautiful faces is due to innate aesthetic schema or rapid acquisition of facial attributes after birth (Damon et al., 2017; Goren et al., 1975; Johnson et al., 2015). For example, Quinn et al. (2008) discovered that infants aged three to four months exhibited preferential attention toward the faces of cats and tigers, suggesting the potential for an inherent aesthetic schema. Nonetheless, none of these studies completely ruled out the influence of visual experience.

Investigating the innate and intuitive aspects of aesthetic perception presents significant challenges. On the one hand, it is difficult to obtain valuable neural activity data from newborn infants. On the other hand, adults' sense of facial beauty is susceptible to various environmental factors, such as culture, observer gender, and mental state (Thiruchselvam et al., 2016; Vartanian et al., 2013; Y. Zhang et al., 2016). In this context, DNN models that mimic brain information processing offer a viable approach to studying aesthetic perception by eliminating numerous confounding factors. Crucially, they enable the exclusion of visual aesthetic experience effects. Recently, biologically inspired artificial neural network models have been increasingly utilized to probe multiple dimensions of visual information processing (Alzubaidi et al., 2021). Research has demonstrated that a feedforward network, even when randomly initialized and untrained, is capable of initiating a range of cognitive functions (Ullman et al., 2012). Furthermore, a subnetwork within a similarly randomized neural network setup can effectively execute image classification tasks (Ramanujan et al., 2020). Evidence suggests that even without the benefit of learning from visual experience, a randomly initialized network retains the capacity for visual feature extraction. Among various deep neural models, AlexNet, developed by Krizhevsky et al. (2012), emulates the processing pattern of the ventral visual stream and has gained prominence in the field of visual information processing. This model is rooted in a hierarchical framework, initially capturing elementary visual features

and subsequently discerning more complex features at higher levels of the hierarchy (Jo et al., 2019; Kravitz et al., 2013; Rust & DiCarlo, 2010; Tong, 2003). Remarkably, an untrained AlexNet demonstrates selective responsiveness to numerosity and faces (Baek et al., 2021; Kim et al., 2021). However, within the realm of facial beauty research, considerable emphasis has been placed on utilizing DNNs for the classification or prediction of facial attractiveness (Bougourzi et al., 2022; Gan et al., 2014; Zhai et al., 2020), as well as the extraction and synthesis of aesthetic features in faces (Chiang et al., 2014; Schmid et al., 2008; Zhan et al., 2020). Despite this focus, there is a paucity of studies investigating the intrinsic developmental mechanism of aesthetics.

The DNN may employ a hierarchical processing mode for facial beauty. The findings imply that brain regions conventionally associated with facial feature processing are also implicated in responses to facial beauty. A "core system" for facial processing comprises the inferior occipital gyrus (IOG), the fusiform gyrus (FG), and the superior temporal sulcus (STS) (Haxby et al., 2000). The IOG is engaged in the perception of facial features, while the STS and FG recognize changeable and invariant features, respectively. The perceptual processing of facial beauty extends beyond this core system to include regions such as the amygdala, sublentiform extended amygdala, ventral tegmental area, orbitofrontal cortex, and nucleus accumbens (Senior, 2003). Specifically, the amygdala and ventral tegmental area are principally implicated in the assessment of beauty, whereas aesthetic processing culminates within the orbitofrontal cortex and nucleus accumbens. Magnetic resonance imaging studies have revealed a positive correlation between facial attractiveness and neural activity within the IOG and FG (Kranz & Ishai, 2006; Mende-Siedlecki et al., 2013; Tsukiura & Cabeza, 2011a, 2011b; Vartanian et al., 2013). Furthermore, activation of the STS was regulated by facial beauty (Kranz & Ishai, 2006). According to the dual-code theory of facial processing, this comprises two primary types of information: configuration information, which pertains to the spatial arrangement of facial features, and feature information, which entails the specific details of facial traits (Cabeza & Kato, 2000). Both elements are integral to the overall processing of faces (Yovel & Kanwisher, 2004). For instance, linear responses of the posterior central gyrus, caudate nucleus, and bilateral inferior frontal gyrus were observed for facial proportion-induced facial attractiveness (Shen et al., 2016). Specifically, features such as ear length, nose size, eye spacing, and distance between lips and chin considerably contribute to these linear responses. These results indicate that configuration and feature information play an important role in the perception of facial beauty, and the nervous system may have a linear response to these two kinds of information. Tanaka and Simonyi (2016) posited that the recognition of complex objects, including faces, requires a process from early processing of simple low-level features to high-level of complex information. On this basis, we hypothesize that the neural perception of facial beauty may begin with a linear processing of configuration and feature information, eventually leading to an integrated holistic representation of facial beauty.

The responses of the earlier layers in DNNs may be integrated in a weighted sum to form the overall processing for facial beauty in the final layer. Neuroeconomic research has demonstrated that values can be constructed by considering different underlying features or properties of stimuli (Farashahi et al., 2019; Koechlin, 2020). The valuation process for a stimulus, including works of art, entails an

initial decomposition into distinct features followed by a weighted recombination to derive a subjective evaluation of the whole entity (Iigaya et al., 2021; O’Doherty et al., 2021). Combined with computer vision and machine learning techniques, some studies have proposed the Linear Feature Summation (LFS) Model as a framework for understanding the construction of aesthetic value (Iigaya et al., 2023). Visual stimuli are first decomposed into multiple visual features representing the color or shape of the painting, which are then transformed into abstract high-level features. Simple linear regression analysis applied to these features can generate reliable value judgments about novel visual stimuli. Given that attributes of facial beauty possess gradable characteristics and considering that neural responses to facial beauty are linearly triggered by these attributes, we use a summation coding model to examine the integrative process underlying the perception of beauty (Q. Chen & Verguts, 2013). The proposed model aligns with the LFS approach in terms of utilizing a linear weighted summation technique; however, it diverges by segmenting features into tuning curves corresponding to neural units in the early layers that exhibit monotonically increasing or decreasing responses. Building on these foundations, we put forth the concept that the neural network’s perception of facial beauty involves a form of summation encoding—specifically, certain neural units in the early layers respond monotonically to different levels of facial beauty, ultimately constructing the overall response to facial beauty.

Here, using a completely randomly initialized neural network, we discovered that neural units selectively responsive to beauty spontaneously emerge within the network. These neural units can accomplish the task of beauty comparison. Additionally, the quantification of facial beauty appears to be limited, reflected by the tuning curves of neural units with varying preferences for beauty aligning along a linear scale. Closer examination suggests that the response of the untrained AlexNet model to facial beauty likely adopts a hierarchical processing structure. These findings support the notion that different layers within the neural network fulfill distinct roles in beauty perception, and the culminating response observed at the last convolutional layer comes from the contribution of inputs from earlier layers. Therefore, drawing on insights from the LFS model and the summation coding model (Q. Chen & Verguts, 2013; Iigaya et al., 2021; Stoianov & Zorzi, 2012), we investigated whether neural units tuned to various levels of beauty emerge due to the assembly of units in the early layers that exhibit monotonically increasing or decreasing responses to beauty.

Method

DNN Model

AlexNet (Krizhevsky et al., 2012), which emulates the ventral visual stream of the brain, was employed for the extraction of aesthetic features from visual stimuli. This model is composed of five convolutional layers followed by three fully connected layers. We initialized each convolutional layer by randomly assigning weights drawn from either a normal or a uniform distribution, both with a zero mean. The standard deviation (*SD*) for the weights was determined using the inverse square root of the count of neural units in the previous layer to balance the influence of input signals throughout the network’s convolutional layers (bias is set to 0; He et al., 2015).

Stimulus Data Set

Stimulus sets were created using StyleGAN2 (Karras et al., 2019, 2020), an advanced generative algorithm developed by Nvidia capable of producing high-resolution facial images. We initially utilized a pre-trained model, which had been trained on the Flickr-Faces-High-Quality (<https://github.com/NVLabs/ffhq-dataset>) data set, to generate and obtain latent codes for Asian female faces (a312863063, 2022). After purging distorted images resulting from the generation process, we retained a total of 150 images (original dimensions: $1,024 \times 1,024 \times 3$ pixels; resized to $227 \times 227 \times 3$ pixels using MATLAB). The stimuli were categorized into three sets based on facial orientation: frontal, left profile, and right profile (refer to Figure 1B). Each set comprised 50 distinct images. Subsequently, these images’ latent codes were manipulated to adjust facial attributes. By employing an embedding algorithm paired with a direction vector derived from pretrained aesthetic labeling data, it became possible to edit the aesthetic qualities of faces.

Following the creation of facial images across different beauty levels, three adults (average age 25.6) assessed their physical authenticity. They determined that images falling within a range of -21 to 25 most accurately reflected the structure of human faces as perceived with the naked eye. Utilizing adjustment parameters set between -21 and 25 , we established five degrees of beauty alterations (2, 3, 4, 5, and 6). This resulted in the generation of sets containing 23, 16, 12, 10, and eight beauty levels, respectively. Consequently, a total of $(23 + 16 + 12 + 10 + 8) \times 3 \times 50 = 10,350$ images was used to analyze the response of networks.

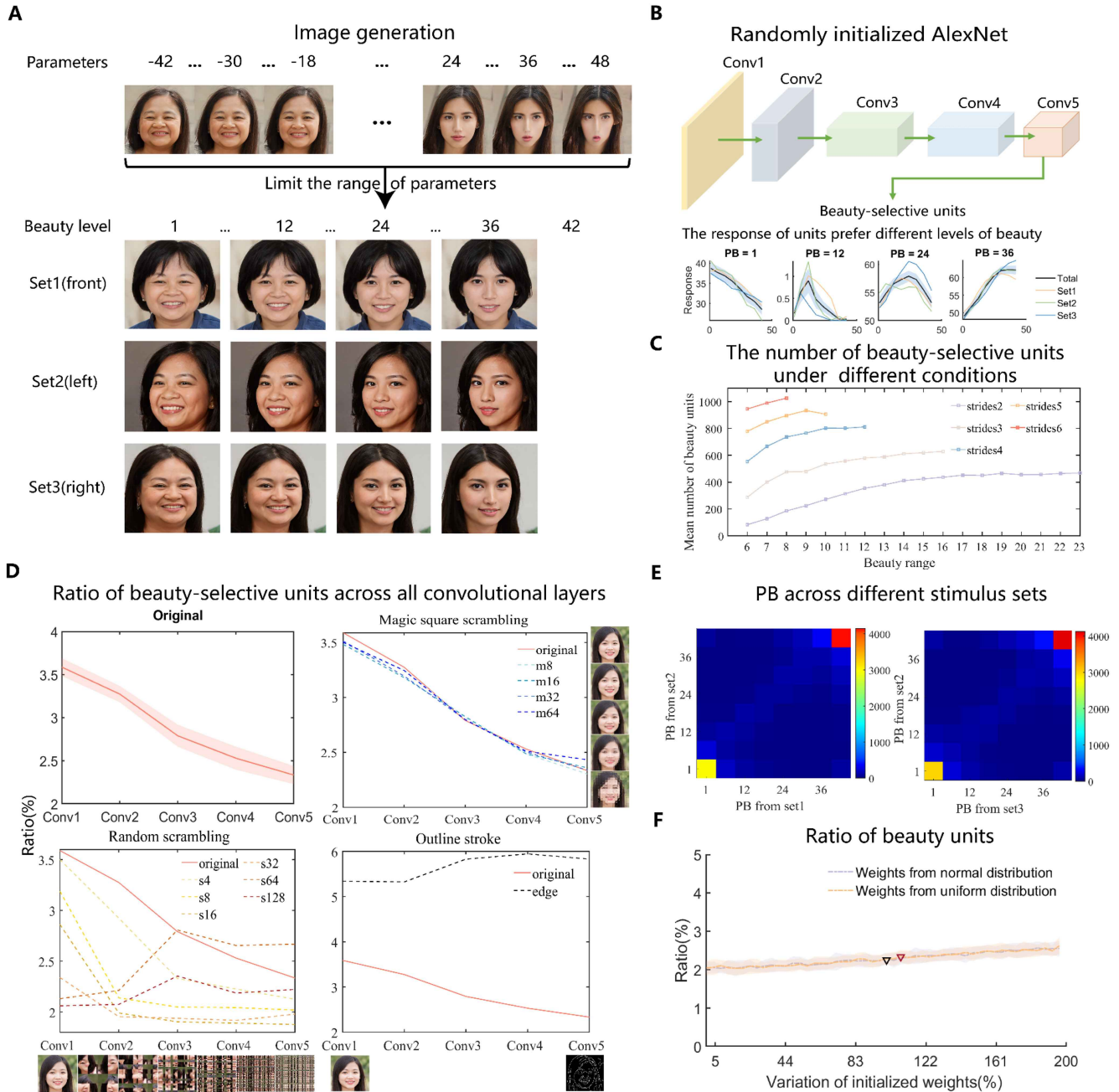
To elucidate the specific contributions of each layer within the DNN to the perception of facial beauty, we subjected the facial images to a series of manipulations. First, to eliminate the influence of feature proportions and overall face contours while maintaining localized information, we applied six levels of random scrambling to the original facial images. Second, aiming to preserve the outline and proportions of facial features and simultaneously exclude detailed facial data, we executed four graded magic square scrambling procedures on the original images. Furthermore, contour extraction was performed on the facial images to remove background elements, color, and skin texture, thus conserving only the contour and spatial arrangement of features.

Investigation of DNN Unit Responses

We conducted an analysis on the responses within AlexNet, a DNN composed of eight layers, including five convolutional and three fully connected layers. In light of the study’s aim to concentrate on the selective response of neural units, the three fully connected layers were omitted from the analysis, thereby focusing exclusively on the response of units in the fifth convolutional layer (see Figure 1C, top).

A network unit is identified as a beauty-selective unit if its response exhibited a significant variation in relation to the level of beauty (two-way analysis of variance, $p < .01$) without substantial fluctuations caused by orientation or interaction between these two factors. Conversely, a nonselective unit refers to a unit that is exclusively sensitive to alterations in orientation (two-way analysis of variance, $p < .01$) and remains unaffected by changes in the level of beauty or their interplay. Namely, beauty-selective unit refers to a specific unit within a neural network that demonstrates selectivity or responsiveness to features associated with beauty. The preferred beauty (PB) for a unit is defined as the beauty level at which the unit, on average, elicits its maximal response across all examined levels of beauty.

Figure 1
Spontaneous Emergence of Beauty-Selective Units in Untrained Neural Networks



Note. (A) Top: Image generation range diagram. Bottom: Examples of visual stimuli for facial beauty. (B) Top: The architecture of randomly initialized AlexNet. Bottom: Examples of tuning curves for individual beauty-selective units observed in the untrained AlexNet. (C) The number of beauty-selective units under different conditions. The horizontal coordinate represents the number of levels under different stride lengths. (D) Top left: The proportion of emerging beauty-selective units in five convolutional layers. Top right and bottom: The ratio of the five convolution layers' selective units after scrambling and contouring the stimulus images. The legends represent different degrees of scrambling. (E) The PB estimates measured with different stimulus conditions are significantly correlated with each other. (F) When the weight variation was substantially changed, the proportion of selective units in the fifth convolution layer under different initialization conditions was consistently observed. Conv = convolutional layer; PB = preferred beauty; m8 = magic square scrambling degree 8; m16 = magic square scrambling degree 16; m32 = magic square scrambling degree 32; m64 = magic square scrambling degree 64; s4 = random scrambling degree 4; s8 = random scrambling degree 8; s16 = random scrambling degree 16; s32 = random scrambling degree 32; s64 = random scrambling degree 64; s128 = random scrambling degree 128. See the online article for the color version of this figure.

The definition of the tuning width of each unit, as well as the methodology for computing the average tuning curves for all beauty-selective units, adheres to protocols established by Kim et al. (2021). Specifically, the tuning width of each unit was considered as the SD (sigma) of the Gaussian fit of the average tuning curve on a linear scale of beauty. To determine the average tuning curves of all beauty-selective units, the tuning curve of each unit was normalized and then averaged across units using the PB level as a reference point. Further, to obtain the average tuning curves across different beauty levels, the tuning curve of each unit was averaged and normalized across units preferring the same beauty level.

Beauty Comparison Task for the DNN

To determine the efficacy of beauty-selective units in evaluating facial beauty, we employed a beauty comparison task based on the methods described by previous research (Kim et al., 2021). Initially, we trained a support vector machine (SVM) by randomly sampling the responses of 256 units (with 10 trials of sampling for each untrained network). The SVM was provided with a sample stimulus and a test stimulus to predict whether the sample stimulus was more beautiful than the test stimulus. Specifically, we generated 100 sample stimuli for each level of beauty, followed by creating one test stimulus for each sample stimulus, while ensuring that the test stimuli do not align with the same level of beauty as the sample stimuli. Finally, unit tuning curves were normalized to calculate the average tuning curves of both correct and incorrect trials during the task.

To verify the homogeneity of the generated face images under the same beauty level, we collected facial beauty ratings from 40 participants (14 male and 26 female, $M_{\text{age}} = 21.18$) using the Psychtoolbox. During this assessment, we presented the participants with two images of faces on a screen at the same time, asked them to choose the face they thought was more beautiful and recorded their press response. The images were selected from the conditions of beauty range from 1 to 42 with a stride of 6 and a beauty level of 8. A total of 800 images were assessed, 400 of which were used as training sets and the remaining 400 as testing sets, with 50 images for each of the 8 levels of beauty. The two images presented were randomly selected 50 times from 50 images at the same level.

Model Simulation for Weighted Summation of Monotonically Responsive Units

Based on the statistics obtained from the untrained AlexNet, a model was designed to simulate the activity of decreasing and increasing units (Kim et al., 2021). The tuning curve of a model output unit (R) was defined by:

$$R = \text{ReLU} \left[\sum w_{\text{Dec}, i} r_{\text{Dec}, i} + \sum w_{\text{Inc}, i} r_{\text{Inc}, i} \right], \quad (1)$$

where $w_{\text{Dec}, i}$ and $w_{\text{Inc}, i}$ are the weight of the i th decreasing or increasing units, respectively, and $r_{\text{Dec}, i}$ and $r_{\text{Inc}, i}$ indicate their tuning curve. To simulate the tuning curves in Conv4 of the untrained AlexNet, the decreasing and increasing unit activities (r) were modeled as normal distributions peaking at 1 and 42, respectively. The Gaussian distribution is modeled from the statistics measured in the untrained AlexNet, $M \pm SD = 16.5 \pm 1.79$ (decreasing), 17.1 ± 1.63 (increasing), from which the tuning width of 60 decreasing or increasing units was then

randomly sampled. The feedforward weight (w) was also randomly sampled from the Gaussian distribution estimated from the untrained AlexNet. Ten thousand output units were generated in a trial, and a total of 1,000 trials were performed.

The definition of increasing (or decreasing) units in Conv4 of the untrained AlexNet was adapted from previous work (Stoianov & Zorzi, 2012). The beauty is expressed linearly, and those monotonic units were defined by regressing the response of unit i (R_i) with the beauty level (B) and the coefficient that manipulates beauty (C).

$$R_i = \beta_B B + \beta_C C + \varepsilon. \quad (2)$$

For a unit, if the regression explains at least 10% of the variance in its response and the regression coefficient of adjustment amplitude is less than 0.1 when editing the beauty degree, β_C , it is defined as an increasing (or decreasing) unit.

Validation Analysis

To ascertain the stability of the research results across different sampling trials, all experiments concerning DNNs in this study went through 20 trials, each with varying weight sampling of the convolutional layers in random initialization.

In order to further verify the stability of the untrained DNN’s representation of facial beauty, we used randomly initialized VGG-16 (Visual Geometry Group 16-layer network) to respond to beautiful images to exclude the effects of network depth, convolution kernel size, etc. Specifically, we examined the selective response to the beauty of the 13 convolution layers of untrained VGG-16 and the monotonic response to the beauty of the first 12 convolution layers.

In an effort to corroborate the generalizability of beauty-selective units, we selected 200 female faces (average age 28.39) from the Chicago Face Database (<https://www.chicagofaces.org/>) as a validation set (Figure 7A). This data set comprises 57 Asian American faces, 52 Indian faces, 90 White faces, and one multiracial American face (Lakshmi et al., 2021; Ma et al., 2015, 2021), all displaying neutral expressions. We recruited 100 participants (average age 23.08, 31 male and 69 female) to evaluate the attractiveness level of these faces. The newly collected data was used as labels for the validation set, testing the generalization capability of beauty-selective units in performing comparative tasks.

Statistical Analysis

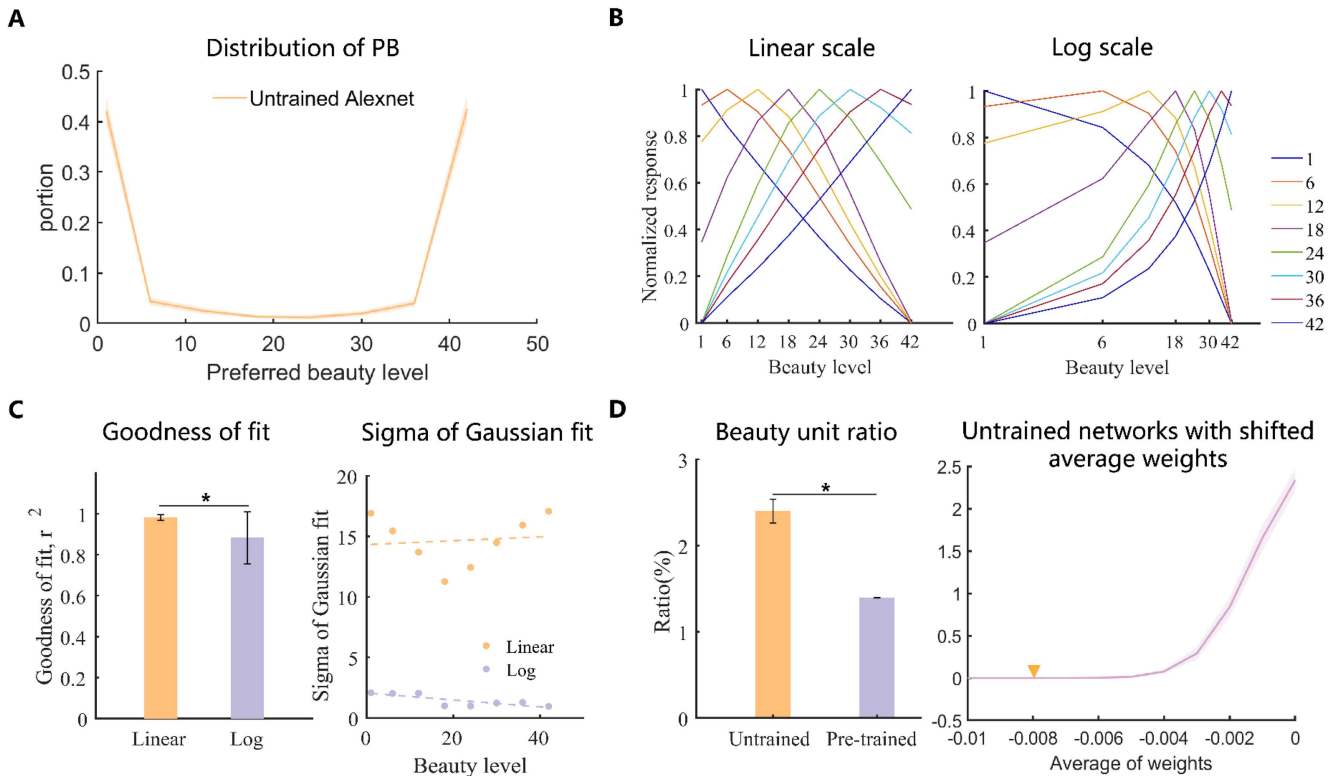
All sample sizes, exact p values, and statistical methods are indicated in the corresponding texts or figure legends. The one-sided Wilcoxon rank sum test was used for most comparison analyses. Shaded areas or error bars indicate the SD in Figures 1D and 1F, 2A–2D, 3B–3E, 4C, 4F and 4G, 5, 6B–6I, and 7B, and indicate the standard error (SE) in Figures 1B and 6A.

Results

Quantification of Facial Beauty

We used StyleGAN2 to generate facial images with different degrees of beauty and found that the level of facial beauty could be quantified with a unique scale beyond the boundaries of this scale, facial stimuli failed to maintain their physiological plausibility (Figure 1A, top). Consequently, we confined the adjustment

Figure 2
Tuning Properties of Beauty-Selective Units in the Untrained Network



Note. (A) Distribution of PB level in the untrained network. (B) Average tuning curves of different beauty levels on a linear scale and a logarithmic scale. (C) Left: The goodness of the Gaussian fit (r^2) is greater on a linear scale. $*p < 1.84 \times 10^{-5}$, Wilcoxon rank sum test. Orange (dark gray) bars represent the linear scale, and lilac (light gray) bars represent the logarithmic scale. Right: The tuning width (sigma of the Gaussian fitting) is U-shaped on a linear scale and remains constant on a logarithmic scale. (D) Left: Beauty-selective units emerging in the untrained AlexNet are more than trained. $*p < 6.8 \times 10^{-163}$, Wilcoxon rank sum test. Orange (dark gray) bars represent the untrained neural network, and lilac (light gray) bars represent the pre-trained neural network. Right: When the average of weights shifted to -0.008 , the beauty-selective units disappeared. Orange (dark gray) triangles indicate the average of weights -0.008 . PB = preferred beauty. See the online article for the color version of this figure.

parameters between -21 and 25 to represent the different levels of facial beauty (Figure 1A, bottom).

Beauty-Selective Units in Untrained DNNs

We found that there were neural units selectively responding to different degrees of facial beauty (Figure 1B, bottom) in the last layer (Conv5) of the untrained AlexNet (Figure 1B, top). By investigating the neural responses in the Conv5 of beauty stimulus images with strides of 2, 3, 4, 5, and 6, we found that for the same number of classes, more selective units were found with the increase of stride length. Roughly, at the same stride length, more selective units emerged in the network when the number of classes was higher (Figure 1C).

We focused on the condition exhibiting the highest proportion of beauty-selective units, namely, the condition of six strides-and-eight levels, to explore other properties of the network's response to facial beauty. We examined the proportion of beauty-selective units in all five convolution layers and found that the proportion of beauty-selective units showed a decreasing trend from the first layer to the fifth layer (Figure 1D, top left).

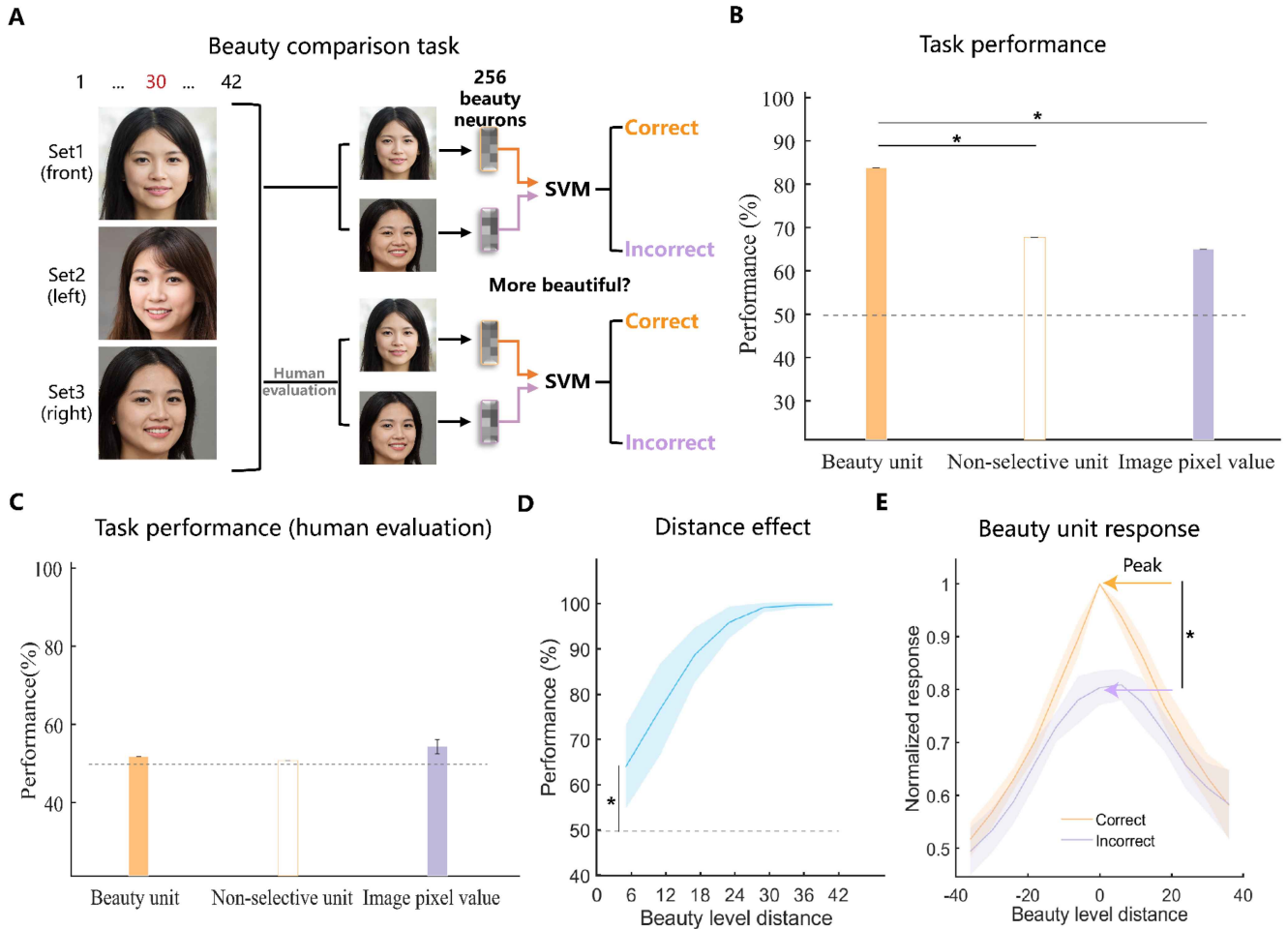
To investigate the response characteristics of different network layers, we made scrambling and contour processing of face images. The

degree of random scrambling indicates the randomization and segmentation of the face stimuli. We found that the proportion of beauty-selective units began to decline sharply from the first layer, and the decline trend disappeared with the scrambling degree increased. The same phenomenon was observed when we applied magic square scrambling to the stimuli, which largely remained the configuration information of the face. In contrast, when only the facial contour was used in the network, the proportion of beauty-selective units varied little across the five convolutional layers with the majority emerging in the later three layers (Figure 1D, top right and bottom).

Beauty-selective units consistently emerged within the untrained AlexNet. The face orientation did not affect the PB outcomes (Figure 1E). Additionally, neither the method of random initialization nor variations in weight distributions impacted the emergence of beauty-selective units (Figure 1F). We changed the width of the random weight distribution (normal and uniform) of each layer under different initialization methods and confirmed that beauty-selective units were continuously observed even when the weight variation was greatly reduced under different initialization conditions.

Notably, responses to beauty were more concentrated in the minimum and maximum levels of beauty (Figure 2A). Furthermore, neural unit responses to the preference for beauty exhibited a Gaussian

Figure 3
Beauty Comparison Performance of Beauty Units



Note. (A) Beauty comparison task. (B) Task performance of the response of beauty units, nonselective units, and the pixel values of raw stimulus images at the different levels (orange (dark gray) open bar: beauty units versus nonselective units, $*p = 9.03 \times 10^{-5}$, Wilcoxon rank sum test; lilac (light gray) solid bar: beauty units versus images, $*p = 8.47 \times 10^{-5}$, Wilcoxon rank sum test). The dashed line indicates the chance level. (C) Task performance of the response of beauty units, nonselective units, and the pixel values of raw stimulus images at the same level. (D) The performance increases as the beauty difference increases and is significantly higher than the chance level for all cases ($*p = 4.78 \times 10^{-5}$, Wilcoxon rank sum test). (E) Average activity of beauty-selective units as a function of the beauty level distance. The response during correct trials is significantly higher than that during incorrect trials ($*p = 4 \times 10^{-9}$, Wilcoxon rank sum test). SVM = support vector machine. See the online article for the color version of this figure.

distribution (Figure 2B), with linear models demonstrating superior fit compared to logarithmic ones (Figure 2C, left). The Gaussian fit's SD (σ) of the mean tuning curve displayed a U-shaped pattern across linear scales corresponding with increases in PB; however, it remained stable on logarithmic scales (Figure 2C, right).

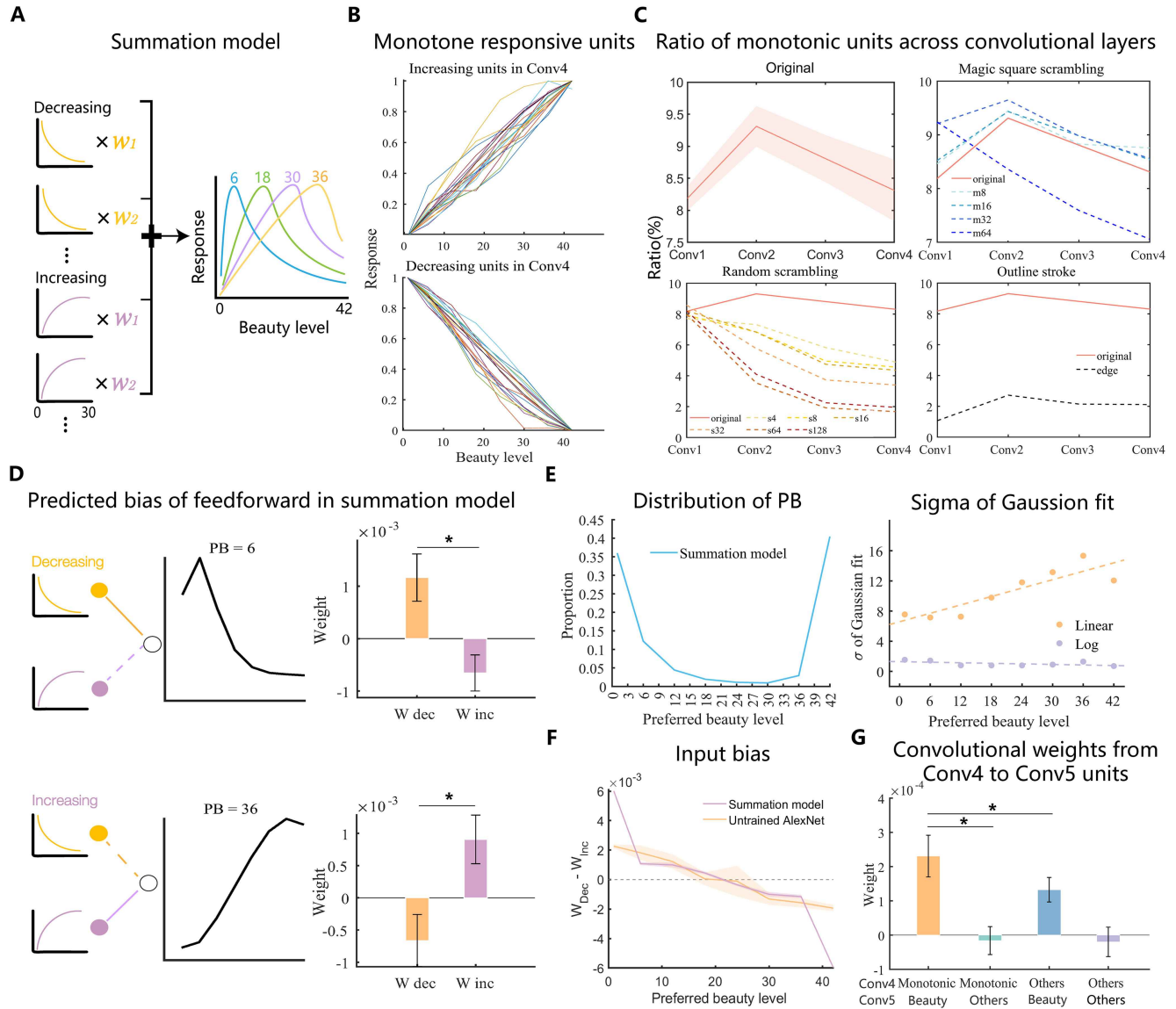
We also observed that the untrained AlexNet exhibited a greater number of beauty-selective units in the fifth layer compared to an AlexNet trained on natural images for classification (2.4% of Conv5 units in the untrained network vs. 1.39% in the pretrained network; Figure 2D, left; $*p < 6.8 \times 10^{-163}$, Wilcoxon rank sum test). Further analysis indicated a negative shift in the average convolutional weight at the fifth layer of the pretrained network relative to the untrained counterpart (-0.13 , in units of the SD of convolutional weights). However, while introducing a similar negative offset to the weights of the untrained network caused a decline in the proportion of beauty-selective units, this reduction did not align with

the pretrained network's level at -0.13 , but rather, the proportion decreased to zero when the weight shift reached -0.008 (Figure 2D, right).

Beauty Comparison Task

We investigated the capability of beauty-selective units to execute comparison tasks across levels and beauty within levels (Figure 3A). The findings indicated that beauty-selective units were indeed able to differentiate between different levels of beauty more effectively than nonselective units and pixel-based image information (Figure 3B; orange open bar: beauty units vs. nonselective units, $*p = 9.03 \times 10^{-5}$, Wilcoxon rank sum test; lilac solid bar: beauty units vs. images, $*p = 8.47 \times 10^{-5}$, Wilcoxon rank sum test). However, when comparing the beauty of faces at the same level, the SVM did not demonstrate the ability to reliably distinguish

Figure 4
Emergence of Beauty Tuning From the Weighted Sum of Increasing and Decreasing Unit Activities

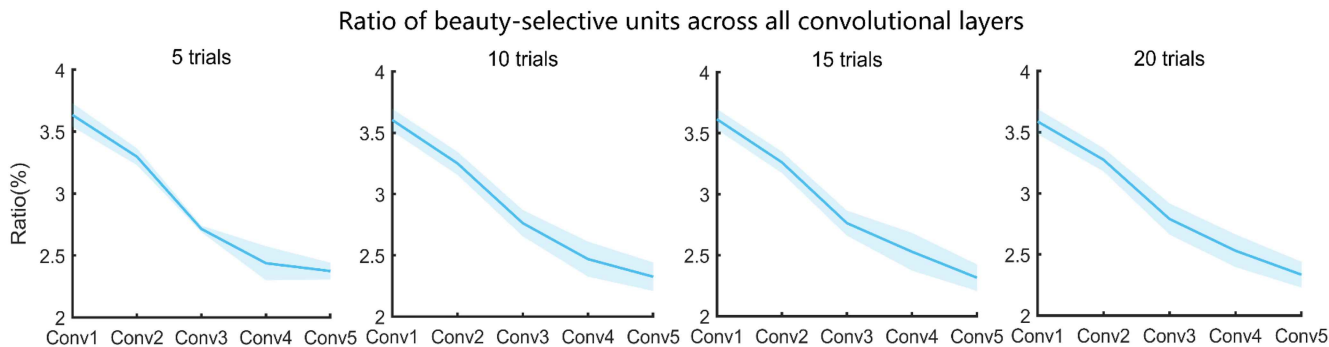


Note. (A) Summation coding model. (B) Monotonically decreasing/increasing neuronal activities as the beauty level increases were observed in Conv4 of untrained AlexNet. (C) Top left: The ratio of monotone units in the first four convolutional layers of the network. Top right and bottom: The monotone unit ratio of the first four convolutional layers of the network after scrambled and contoured face images. (D) According to the model's assumptions, units tune to lower levels of beauty to receive stronger inputs from decreasing units (orange (dark gray) vs. lilac (light gray); $*p = 9.77 \times 10^{-4}$), and vice versa (orange (dark gray) vs. lilac (light gray); $*p = 9.77 \times 10^{-4}$). Right: The average weights of units preferring beauty level 6 and 36. (E) Left: Distributions of the PB level from the model simulation. Right: The tuning width on linear and logarithmic scales predicted by the model. (F) Weight bias of all beauty units observed in Conv5 of the untrained AlexNet. The results of untrained AlexNet and model simulation are basically consistent. (G) Monotonic units in Conv4 provide stronger inputs to beauty units than to the other units in Conv5 (orange vs. green; $*p = 9.13 \times 10^{-5}$, Wilcoxon rank-sum test). Beauty units in Conv5 also connect to monotonic units more strongly than the other Conv4 units (orange vs. blue; $*p = 1.1 \times 10^{-3}$). W_1 – W_2 = examples of the weight of units; Conv = convolutional layer; PB = preferred beauty; W dec = the weight of decreasing units; W inc = the weight of increasing units; m8 = magic square scrambling degree 8; m16 = magic square scrambling degree 16; m32 = magic square scrambling degree 32; m64 = magic square scrambling degree 64; s4 = random scrambling degree 4; s8 = random scrambling degree 8; s16 = random scrambling degree 16; s32 = random scrambling degree 32; s64 = random scrambling degree 64; s128 = random scrambling degree 128. See the online article for the color version of this figure.

beauty (Figure 3C; orange open bar: beauty units vs. nonselective units, $p = .22$, Wilcoxon rank sum test; lilac solid bar: beauty units vs. images, $p = .32$, Wilcoxon rank sum test).

Moreover, the network performed better as the distance from beauty increased (Figure 3D). To further investigate the contributions of the beauty-selective units for correct choices, we compared

Figure 5
Validation Analysis of the Neural Network’s Iterations



Note. The proportion of beauty-selective units among the five convolutional layers remained consistent across 20 trials. Conv = convolutional layer. See the online article for the color version of this figure.

the average tuning curves obtained in correct and incorrect trials. The results showed that the average response to the PB in correct trials was significantly greater than that in incorrect trials (Figure 3E).

Beauty Tuning by Weighted Sum of Monotonically Responsive Units

We analyzed the emergence of beauty-selective units in untrained neural networks by employing a summation coding model and assessed how monotonically responsive units represented beauty across hierarchical layers (Figure 4A; Q. Chen & Verguts, 2013; Stoianov & Zorzi, 2012). We identified neural units exhibiting monotonically increasing or decreasing responses to beauty levels within the early network layers (Figure 4B), with the highest proportion of such monotonic units occurring at the second layer (Figure 4C, top left). Further experimentation involving random scrambling of faces revealed a gradual reduction in the ratio of monotonic units from the first to the fourth layer, where the overall proportion typically decreased as the degree of scrambling intensified (Figure 4C, top right). Even after applying magic square scrambling—designed to preserve facial contour information—the proportion of monotonic units still peaked at the second layer. This peak diminished when more profound degrees of scrambling were introduced, pointing toward a consistent decrement from the first layer onward (Figure 4C, bottom left). After contour treatment, the proportion of monotonic units reached the peak at the second layer, but the whole proportion was much smaller than in other treatment conditions (Figure 4C, bottom right).

Using computational simulations, we discovered that the beauty tuning curves are the result of integrating weighted contributions from both increasing and decreasing unit activities (Figure 4B and 4C). As PB tends to 1 or 42, the number of tuned units increases (Figure 4E, left). The simulated sigma of Gaussian fitting of the average tuning curve was stable on a logarithmic scale but did not show a U-shape on a linear scale (Figure 4E, right). Moreover, the simulations were consistent with the hypothesis that units tuned to low levels of beauty receive stronger inputs from decreasing units (Figure 4D, top right; $*p = 9.77 \times 10^{-4}$), while units tuned to high levels of beauty receive stronger inputs from increasing units (Figure 4D, bottom right; $*p = 9.77 \times 10^{-4}$), showing a feedforward bias. We also observed this bias in Conv5 of the untrained AlexNet (Figure 4F; orange solid

curve). Further, we observed that monotone units in Conv4 contributed more to beauty-selective units than other units in Conv5 (Figure 4G; orange vs. green; $*p = 9.61 \times 10^{-21}$, Wilcoxon rank-sum test).

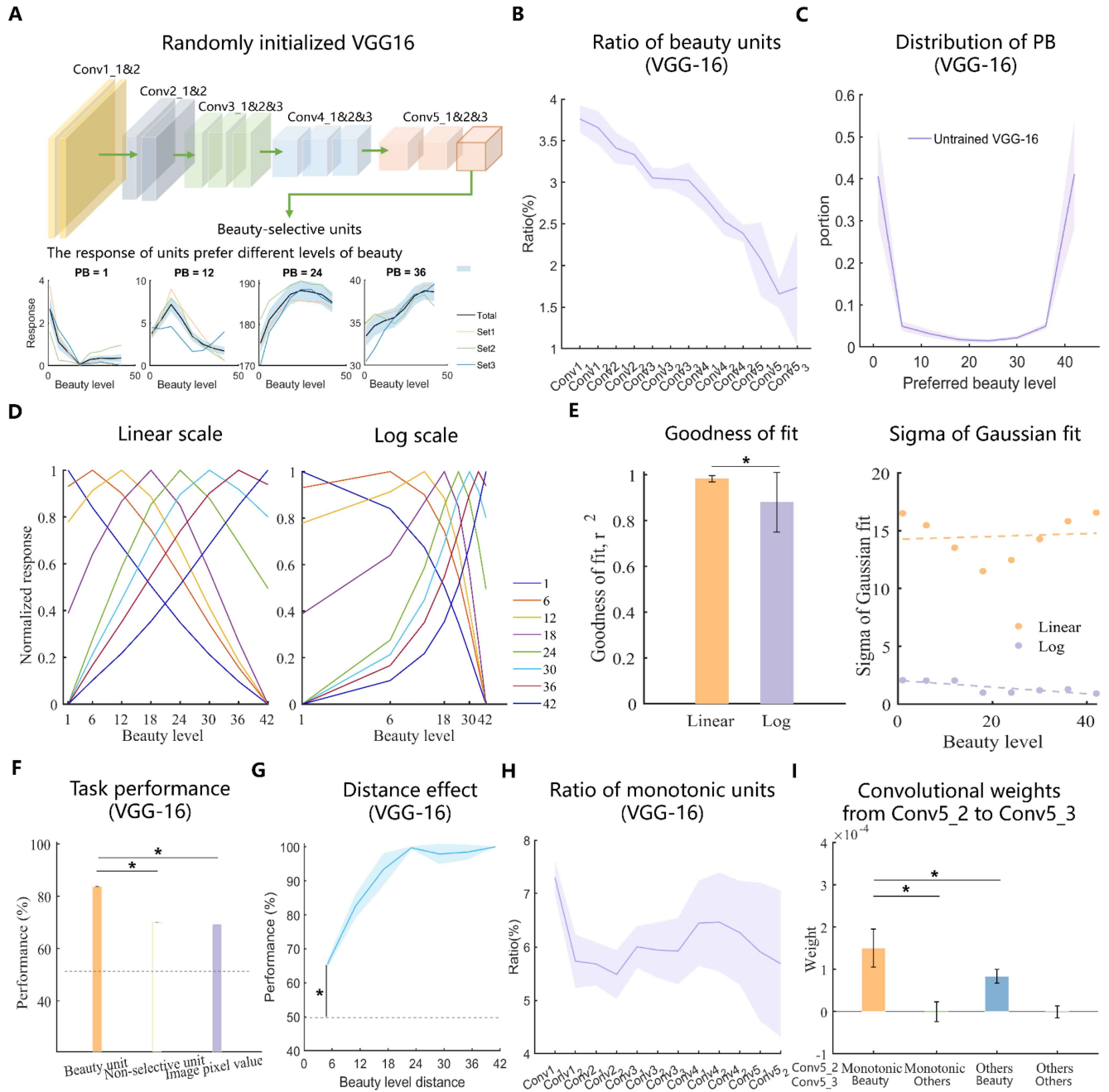
Validation Analysis

The results of experiments with DNNs vary very little from trial to trial (Figure 5).

As with untrained AlexNet, we found that beauty-selective units also spontaneously emerged in untrained VGG-16 (Figure 6A), and the selective response to beauty appeared at the first convolutional layer, the proportion of beauty-selective units decreased from the first layer to the last layer (Figure 6B). Moreover, these beauty-selective units were more distributed at low and high levels of beauty (Figure 6C). Again, just like in the untrained AlexNet, the response of beauty-selective units in untrained VGG-16 was more consistent with linear scale expression (Figure 6D). As the level of beauty increases, the σ of Gaussian fitting is U-shaped on the linear scale and remains stable on the logarithmic scale (Figure 6E, right). These beauty-selective units could also perform the beauty comparison task (Figure 6F), the units performed better as the distance from beauty increased (Figure 6G). In untrained VGG-16, we found units that monotonically increased or decreased in response as the level of beauty increased. However, the distribution trend of monotonic units in the first 12 layers of untrained VGG-16 showed a horizontal S-shape (Figure 6H). We examined the weight contribution of untrained VGG-16 convolution layer monotonic units at the 12th to selective units at the last convolution layer and found that monotonic units in Conv5_2 provided stronger inputs to beauty units than to the other units in Conv5_3 (Figure 6I, right; orange vs. green; $*p = 9.13 \times 10^{-5}$, Wilcoxon rank-sum test).

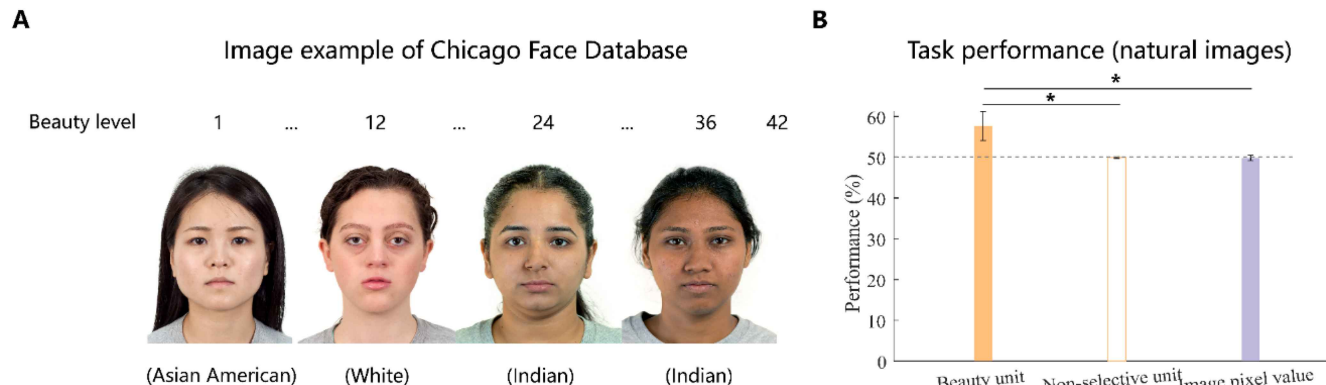
Using images generated by StyleGAN2 as the training set for SVM and the real human images from the Chicago Face Database as its test set (Figure 7A), the results revealed that compared to nonselective units and pixel information, the beauty-selective units performed better in distinguishing levels of beauty in natural images (Figure 7B; orange open bar; beauty units vs. nonselective units, $*p = 6.81 \times 10^{-5}$, Wilcoxon rank sum test; lilac solid bar; beauty units vs. images, $*p = 4.05 \times 10^{-7}$, Wilcoxon rank sum test; dashed line; beauty units vs. chance level, $*p = 1.17 \times 10^{-6}$, Wilcoxon rank sum test).

Figure 6
Validation Analysis of Untrained VGG-16



Note. (A) Top: Architecture of randomly initialized VGG-16. Bottom: Examples of tuning curves for individual beauty-selective units observed in the untrained VGG-16. (B) The proportion of emerging beauty-selective units in thirteen convolutional layers. (C) Distribution of PB in the untrained VGG-16. (D) Average tuning curves of different beauty levels on a linear scale and a logarithmic. (E) Left: The goodness of the Gaussian fit (r^2) is greater on a linear scale. $*p < 7.77 \times 10^{-18}$, Wilcoxon rank sum test. Right: The tuning width (sigma of the Gaussian fitting) is U-shaped on a linear scale and remains constant on a logarithmic scale. (F) Task performance of the response of beauty units in untrained VGG-16, nonselective units, and the pixel values of raw stimulus images at the different levels (orange [dark gray] open bar; beauty units vs. nonselective units, $*p = 3.3 \times 10^{-4}$, Wilcoxon rank sum test; lilac [light gray] solid bar; beauty units vs. images, $*p = 1.37 \times 10^{-4}$, Wilcoxon rank sum test). (G) Beauty distance effect ($*p = 3.17 \times 10^{-6}$, Wilcoxon rank sum test). (H) The ratio of monotone units in the first 12 convolutional layers of the VGG-16. (I) Monotonic units in Conv5_2 provide stronger inputs to beauty units than to the other units in Conv5_3 (orange vs. green; $*p = 9.13 \times 10^{-5}$, Wilcoxon rank-sum test), implying that beauty tuning in Conv5_3 arises from the monotonic units in Conv5_2. Beauty units in Conv5_3 also connect to monotonic units more strongly than the other Conv5_2 units (orange vs. blue; $*p = 2.9 \times 10^{-3}$). VGG = Visual Geometry Group; Conv = convolutional layer; PB = preferred beauty. See the online article for the color version of this figure.

Figure 7
Validation Analysis of the Chicago Face Database



Note. (A) Examples of visual stimuli of different facial beauty in the Chicago Face Database. (B) Task performance of the beauty units, nonselective units and the pixel values of raw stimulus images at the different level used the Chicago Face Database (orange [dark gray] open bar; beauty units vs. nonselective units, $*p = 6.81 \times 10^{-5}$, Wilcoxon rank sum test; lilac [light gray] solid bar; beauty units vs. images, $*p = 4.05 \times 10^{-7}$, Wilcoxon rank sum test; dashed line; beauty units vs. chance level, $*p = 1.17 \times 10^{-6}$, Wilcoxon rank sum test). The dashed line indicates the chance level. Visual stimuli are from “The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data,” by D. S. Ma, J. Correll, and B. Wittenbrink, 2015, *Behavior Research Methods*, 47, pp. 1122–1135 (<https://www.chicagofaces.org/>). CC BY 4.0. See the online article for the color version of this figure.

Discussion

In the process of generating face stimuli with StyleGAN2, we identified a scale for measuring beauty. Using untrained DNNs, after excluding the effects of low-level visual features (e.g., orientation) and random initialization methods, we found that beauty-selective units spontaneously emerged in DNNs and that these units can perform beauty comparison tasks. Within the scale of beauty that we found the tuning curve of the response followed the linear distribution. The units tuned to different levels of beauty came from the construction of units that monotonically increase and decrease in response to the level of beauty in the early layers. The random-initialized DNN processed facial beauty in a hierarchical manner. Verification was performed using an untrained VGG-16, which showed similar representations of facial beauty, suggesting that the architecture of the DNN does not affect the innate representation of beauty.

Echoing Pythagoras’s assertion that beauty resides in the proportionality and harmony of numbers, the quantification of beauty is often anchored in mathematical criteria. Research leveraging geometric features to quantify facial attractiveness has discovered that facial configurations occupy a narrow and specific region within a space (D. Zhang et al., 2011). Similarly, our use of StyleGAN2 to generate facial images has revealed the existence of a measurable range of facial beauty. Faces that deviate significantly from this established continuum tend to deteriorate in aesthetic value and may even lose physiological believability. This suggests a central bias in aesthetic preferences, with extreme alterations to facial features—either excess or insufficiency—leading to a departure from what is conventionally recognized as beautiful.

While previous studies have revealed that infants might possess an inborn facial representation (Slater & Kirby, 1998; Slater et al., 1998), they have not ruled out visual experience or further explored innate representation patterns of facial beauty. In light of our study’s findings, we proposed that representations of beauty emerged independent of experience-based learning, suggesting that beauty may exhibit an intrinsic development pattern within untrained DNNs. Additionally, our

results indicated that representations of beauty manifest in the network’s early layers, implying that the primary processing of facial beauty might commence at the structural encoding phase—a stage as early as when facial features begin to be systematically organized. In fact, facial beauty processing occurs so early that brain regions involved in facial feature processing are already involved in facial beauty processing (Haxby et al., 2000; Kranz & Ishai, 2006). This supports the notion that the neural substrates for feature-based facial processing overlap with, and likely contribute to, the circuits evaluating facial beauty. Evidence from electroencephalography studies indicated that powerful neural representations of facial beauty appear within the visual range of 150–200 ms (Kaiser & Nyga, 2020; Lu et al., 2014). Besides, event-related potential studies on face recognition have shown that components sensitive to facial configuration information and feature information appear at 90–120 and 200–260 ms, respectively (H. L. Wang et al., 2015, 2016).

In the untrained AlexNet, we discovered neural units capable of detecting subtle differences in facial beauty. This suggests that single units do have the ability to differentiate beauty. While no single-neuron studies of human facial beauty have yet been conducted in humans, single-neuron studies in nonhuman primates have found face-selective neurons in the inferotemporal cortex (Gross, 1994; Tsao et al., 2003). Studies of facial identity perception have found that monkey neurons show a tendency to tune around the average human face using a caricaturalized axis (Leopold et al., 2006) and that such neurons can reliably differentiate between extremely subtle differences in successive faces. Furthermore, the units in the network that were sensitive to facial beauty functioned in a linear response mode, but the sigma of the tuning curve presented a “U” shape on the beauty scale. This indicates that the units in the network that respond preferentially to extreme beauty are very active, and they also respond strongly to nearby beauty. One possible explanation is that the units responding to both extremes of beauty levels have cognitive functions associated with rewards or punishments, and therefore, they exhibit broader responses compared to units favoring intermediate beauty levels (Liang et al., 2010).

By scrambling and contouring the images, we provide a glimpse into the hierarchical processing of facial beauty by randomly initialized

neural network models and confirm the contribution of configuration information and feature information to the representation of facial beauty. The configuration and feature information of faces are very important for the brain, which lacks knowledge of beauty, to process aesthetic values (Leo & Simion, 2009; Slater et al., 2000). It has been widely discussed whether configuration information or feature information is more important in facial aesthetics (Abbas & Duchaine, 2008; Liu et al., 2022; Orghian & Hidalgo, 2020). The research findings indicate that individuals make accurate attractiveness judgments of faces only when the upper and lower parts of the face are aligned (Abbas & Duchaine, 2008). Blurred faces, which retain the main configuration information but lose most of the feature information, receive higher attractiveness ratings compared to normal faces (Orghian & Hidalgo, 2020; Sadr & Krowicki, 2019). Other studies have shown that presenting only a part of the face, such as the left or right half, can lead to higher attractiveness ratings than presenting the entire face (Liu et al., 2022; Orghian & Hidalgo, 2020; Sadr & Krowicki, 2019). Certain facial features play important roles in facilitating attractiveness judgments, such as women's faces with features like large eyes, small noses, and narrow chins often receiving higher attractiveness ratings (Baudouin & Tiberghien, 2004; Rhodes, 2006). It can be observed that both configuration information and feature information are crucial for attractiveness judgments of faces, and it is difficult to discern which is more important. Based on the current research findings, we believe that configuration information and feature information may have different roles in different processing stages, and both contribute to the completion of aesthetic processing.

In terms of monotonic response, when the structure of the face was randomly disturbed or the degree of magic square scrambling was high, the peak value of the second layer disappeared, which implies that the second convolution layer may have a unique monotonic response to configuration information such as face contouring and the arrangement of facial features. In the case of random scrambling, the distribution of monotone responses in the convolutional layers shows a decreasing trend, indicating that the early layers are more involved in the monotone processing of feature information. Meanwhile, contour processing did not affect the monotonic response distribution, but the overall ratio was sharply reduced, implying that the feature information is very important for the linear processing of facial beauty.

In the case of beauty-selective response, it was a different story. Magic square scrambling had little effect on the selective response trend and the high ratio of selective response units after contouring indicated that configuration information contributes a lot to the beauty-selective response. As the degree of random scrambling deepened, it was difficult to extract local information and the advantage of early layer processing disappeared, which showed the contribution of feature information to low-level feature processing in early layers. The performance after contour processing could support not only the importance of configuration information in overall facial processing but also the contribution of feature information in feature extraction at earlier layers. In summary, our results filled a gap in the neural processing of innate aesthetics and validated the hierarchical processing of facial beauty by an untrained neural network.

By applying the summation coding model, we discovered that neuronal representations of facial beauty emerge as a result of integrating linearly responsive neural units with varying weights within the network's early layers. A recent study that utilized neuroimaging and DNNs demonstrated that the aesthetic valuation of artwork is computed hierarchically through the linear weighted summation of

low-level and high-level stimulus features, tracing a pathway from the early visual cortex to the parietal and prefrontal cortices (Iigaya et al., 2023). This finding underscores that both cognitive neural processes in the brain and computational models in artificial neural networks can manifest similar patterns in appraising aesthetic value, transcending differences in the vehicles of knowledge and conception of beauty. Neural units tuned to lower levels of beauty tend to receive more pronounced inputs from units displaying decreasing responses, while those tuned to higher levels of attractiveness are predominantly influenced by units with increasing activity profiles. This suggests that neural units with monotonically increasing responses may be more attuned to specific attractive features. Corresponding with previous empirical findings, certain aspects of facial attractiveness, such as larger eyes, smaller nose, and chin, follow a linear trend along a scale of perceived beauty (Baudouin & Tiberghien, 2004; Cunningham et al., 1995; Shen et al., 2016), thereby indicating that incrementally responsive neural units could be selectively attuned to these attributes.

Overall, the results of this study fill a gap in the research on the neural computation of intrinsic aesthetic perception, revealing the hierarchical processing pattern of facial beauty by untrained neural networks. Aesthetics is a crucial part of knowledge acquisition (Perlovsky, 2014; Perlovsky & Schoeller, 2019; Sarasso et al., 2020), where individuals, while appreciating a piece of art, gather meanings and engage in learning (Tracy, 1946). Aesthetic appreciation triggers the instinct to seek knowledge, enabling individuals to gather information and learn from their environment (Schoeller & Perlovsky, 2016; Schoeller et al., 2018). Our research findings indicated that a sense of beauty can emerge without prior learning, potentially supporting an innate sense of aesthetic perception, suggesting that humans might have an inherent ability to learn from aesthetics. Furthermore, perception and evaluation of facial beauty contribute to our ability to identify and select social partners, establish and maintain social connections, and aid in our survival and reproduction in social environments (J. Chen et al., 2012; Eagly et al., 1991; J. Wang et al., 2017). Facial aesthetics are closely linked to information such as mate selection and genetic inheritance (Fink & Penton-Voak, 2002; Little et al., 2011; Thornhill & Gangestad, 1999). Therefore, aesthetic abilities may have played a significant role in human evolution, and revealing the inherent neural computational patterns of facial aesthetics can help us gain deeper insights into the origins and evolution of social cognition and adaptive behavior in human evolution.

While our study was not a substitute for neuroimaging studies, it bore a striking resemblance to functional neuroimaging studies of adult aesthetics of artworks (Iigaya et al., 2023). This suggests that the emergence of initial beauty sense has a similar neural mechanism to the formation of adult aesthetic value, which makes sense why infants would prefer to focus on faces rated as attractive by adults (Quinn et al., 2008; Samuels & Ewy, 1985). Due to the limitations of existing technology in directly studying the neural processing patterns of human intrinsic aesthetic perception, this study, conducted using biologically inspired DNN models, offers a neurocomputational perspective on the inherent developmental patterns of aesthetics. It provides an interpretable approach to understanding the mechanisms underlying aesthetic value generation.

Although our current study identified units that respond specifically to facial beauty in untrained neural networks and made a preliminary exploration of the response patterns of these units, there are still gaps that need to be filled by more systematic and rich studies.

Firstly, based on the current algorithms used in this research, it is still difficult to rigorously control factors like lighting, background, facial shape, and hairstyle. While it is challenging to avoid certain systematic alterations, it also retains the authenticity of the natural facial features. Moreover, we have validated our beauty-selective units by utilizing additional data sets, confirming their response to facial beauty rather than systemic facial changes. In the future, achieving fine control over specific dimensions of the face might require overcoming certain technical challenges in the field of artificial intelligence. Second, the standards of facial beauty are not uniform. The current study only verified the activity of units under one standard, more research is needed in the future to focus on different standards of beauty (such as averageness, symmetry, nose size, eye spacing, and so on). Morton and Johnson (1991) propose the “CONSPEC” mechanism, arguing that infants are born with some information about facial structure, indicating the presence of an innate aesthetic pattern. Exploring beauty representation using different criteria helps identify the innate aesthetic pattern. Finally, although facial beauty is representative in aesthetic terms, the scope of beauty is not limited to the face. Humans can find beauty in bodies, faces, landscapes, fine works of art, and even mathematical proofs (Chatterjee, 2013). In terms of the initialized brain, infants can not only preferentially gaze at the attractive faces of the field (Quinn et al., 2008) but also distinguish and respond preferentially to paintings (Cacchione et al., 2011; Krentz & Earl, 2013), suggesting that infants may have an innate preference for general structural features (Turati et al., 2002; Wilkinson et al., 2014). More specifically, the initial setup of our perceptual system prompts infants to prefer looking at certain entities because we have a range of preferred perceptual features, including but not limited to certain specific features such as large eyes (Geldart, Maurer, & Carney, 1999) and complex geometric properties (Cassia et al., 2004; Geldart, Maurer, & Henderson, 1999). The initial neural responses of aesthetic objects such as natural landscapes, nonhuman faces, and paintings can be further studied to systematically explain the internal mechanism of beauty sense in the randomly initialized DNNs.

Conclusion

In conclusion, our research explored the neural origin of beauty sense from the innate computational abilities of DNNs. We found that there are units selectively responsive to facial beauty in the completely randomly initialized DNNs, and the responses of these units are linearly distributed. Additionally, representations of beauty have emerged in the initial layers of DNNs. Untrained DNNs perceive facial beauty in a hierarchical manner, where both configuration information and feature information of faces contribute to the completion of aesthetic processing. The selective responses in the final layer of the DNN are constructed through linear weighting of monotonic responses in the early layers.

References

- a312863063. (2022). Generators-with-StyleGAN2. *GitHub repository*. <https://github.com/a312863063/generators-with-stylegan2.git>
- Abbas, Z. A., & Duchaine, B. (2008). The role of holistic processing in judgments of facial attractiveness. *Perception, 37*(8), 1187–1196. <https://doi.org/10.1068/p5984>
- Alzubaidi, L., Zhang, J. L., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*(1), Article 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Baek, S., Song, M., Jang, J., Kim, G., & Paik, S. B. (2021). Face detection in untrained deep neural networks. *Nature Communications, 12*(1), Article 7328. <https://doi.org/10.1038/s41467-021-27606-9>
- Bashour, M. (2006). An objective system for measuring facial attractiveness. *Plastic and Reconstructive Surgery, 118*(3), 757–774. <https://doi.org/10.1097/01.prs.0000207382.60636.1c>
- Baudouin, J. Y., & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychologica, 117*(3), 313–332. <https://doi.org/10.1016/j.actpsy.2004.07.002>
- Birkhoff, G. D. (1933). *Aesthetic measure*. Harvard University Press.
- Bougourzi, F., Dornaika, F., & Taleb-Ahmed, A. (2022). Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. *Knowledge-Based Systems, 242*, Article 108246. <https://doi.org/10.1016/j.knosys.2022.108246>
- Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science, 11*(5), 429–433. <https://doi.org/10.1111/1467-9280.00283>
- Cacchione, T., Möhring, W., & Bertin, E. (2011). What is it about Picasso? Infants’ categorical and discriminatory abilities in the visual arts. *Psychology of Aesthetics, Creativity, and the Arts, 5*(4), 370–378. <https://doi.org/10.1037/a0024129>
- Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns’ face preference? *Psychological Science, 15*(6), 379–383. <https://doi.org/10.1111/j.0956-7976.2004.00688.x>
- Chatterjee, A. (2013). *The aesthetic brain: How we evolved to desire beauty and enjoy art* (Vol. 48, No. 1). Oxford University Press. https://doi.org/10.1162/LEON_r_00948
- Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience Letters, 517*(2), 65–70. <https://doi.org/10.1016/j.neulet.2012.02.082>
- Chen, Q., & Verguts, T. (2013). Spontaneous summation or numerosity-selective coding? *Frontiers in Human Neuroscience, 7*, Article 886. <https://doi.org/10.3389/fnhum.2013.00886>
- Chiang, W. C., Lin, H. H., Huang, C. S., Lo, L. J., & Wan, S. Y. (2014). The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D Moire features. *Pattern Recognition, 47*(3), 1249–1260. <https://doi.org/10.1016/j.patcog.2013.09.007>
- Cunningham, M. R., Roberts, A. R., Wu, C. H., Barbee, A. P., & Druen, P. B. (1995). Their ideas of beauty are, on the whole, the same as ours—Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68*(2), 261–279. <https://doi.org/10.1037/0022-3514.68.2.261>
- Damon, F., Mottier, H., Méary, D., & Pascalis, O. (2017). A review of attractiveness preferences in infancy: From faces to objects. *Adaptive Human Behavior and Physiology, 3*(4), 321–336. <https://doi.org/10.1007/s40750-017-0071-2>
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*(1), 109–128. <https://doi.org/10.1037/0033-2909.110.1.109>
- Farshahi, S., Donahue, C. H., Hayden, B. Y., Lee, D., & Soltani, A. (2019). Flexible combination of reward information across primates. *Nature Human Behaviour, 3*(11), 1215–1224. <https://doi.org/10.1038/s41562-019-0714-3>
- Fedrizzi, L. (2012). Beauty and its perception: Historical development of concepts, neuroaesthetics, and gender-differences. *Rendiconti Lincei, 23*(3), 259–269. <https://doi.org/10.1007/s12210-012-0177-1>
- Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*(5), 154–158. <https://doi.org/10.1111/1467-8721.00190>

- Gan, J. Y., Li, L. C., Zhai, Y. K., & Liu, Y. H. (2014). Deep self-taught learning for facial beauty prediction. *Neurocomputing*, *144*, 295–303. <https://doi.org/10.1016/j.neucom.2014.05.028>
- Geldart, S., Maurer, D., & Carney, K. (1999). Effects of eye size on adults' aesthetic ratings of faces and 5-month-olds' looking times. *Perception*, *28*(3), 361–74. <https://doi.org/10.1068/p2885>
- Geldart, S., Maurer, D., & Henderson, H. (1999). Effects of the height of the internal features of faces on adults' aesthetic ratings and 5-month-olds' looking times. *Perception*, *28*(7), 839–850. <https://doi.org/10.1068/p2943>
- Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549. <https://doi.org/10.1542/peds.56.4.544>
- Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection—The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*(3), 233–242. <https://doi.org/10.1037/0735-7036.108.3.233>
- Gross, C. G. (1994). How inferior temporal cortex became a visual area. *Cerebral Cortex*, *4*(5), 455–469. <https://doi.org/10.1093/cercor/4.5.455>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification*. IEEE Computer Society.
- Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., & O'Doherty, J. P. (2021). Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nature Human Behaviour*, *5*(6), 743–755. <https://doi.org/10.1038/s41562-021-01124-6>
- Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, S., Cross, L., & O'Doherty, J. P. (2023). Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nature Communications*, *14*(1), Article 127. <https://doi.org/10.1038/s41467-022-35654-y>
- Jo, H. G., Ito, J., Holthausen, B. S., Baumann, C., Grun, S., Habe, U., & Kellermann, T. (2019). Task-dependent functional organizations of the visual ventral stream. *Scientific Reports*, *9*(1), Article 9316. <https://doi.org/10.1038/s41598-019-45707-w>
- Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169–179. <https://doi.org/10.1016/j.neubiorev.2014.10.009>
- Kaiser, D., & Nyga, K. (2020). Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Scientific Reports*, *10*(1), Article 16852. <https://doi.org/10.1038/s41598-020-74009-9>
- Karras, T., Laine, S., Aila, T., & Soc, I. C. (2019, June 16–20). *A style-based generator architecture for generative adversarial networks*. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). *Analyzing and improving the image quality of StyleGAN* [Conference Paper]. 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Proceedings (pp. 8107–8116). <https://doi.org/10.1109/cvpr42600.2020.00813>
- Kim, G., Jang, J., Baek, S., Song, M., & Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. *Science Advances*, *7*(1), Article eabd6127. <https://doi.org/10.1126/sciadv.abd6127>
- Koechlin, E. (2020). Human decision-making beyond the rational decision theory. *Trends in Cognitive Sciences*, *24*(1), 4–6. <https://doi.org/10.1016/j.tics.2019.11.001>
- Kranz, F., & Ishai, A. (2006). Face perception is modulated by sexual preference. *Current Biology*, *16*(1), 63–68. <https://doi.org/10.1016/j.cub.2005.10.070>
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. <https://doi.org/10.1016/j.tics.2012.10.011>
- Krentz, U. C., & Earl, R. K. (2013). The baby as beholder: Adults and infants have common preferences for original art. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(2), 181–190. <https://doi.org/10.1037/a0030691>
- Krizhevsky, A., Sutskever, I. & Hinton, G. (2012). *Imagenet classification with deep convolutional neural networks* (Vol. 25). Advances in neural information processing systems.
- Lakshmi, A., Wittenbrink, B., Correll, J., & Ma, D. S. (2021). The India face set: International and cultural boundaries impact face impressions and perceptions of category membership. *Frontiers in Psychology*, *12*, Article 627678. <https://doi.org/10.3389/fpsyg.2021.627678>
- Langlois, J. H., Ritter, J. M., Roggman, L. A., & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, *27*(1), 79–84. <https://doi.org/10.1037/0012-1649.27.1.79>
- Leo, I., & Simion, F. (2009). Face processing at birth: A Thatcher illusion study. *Developmental Science*, *12*(3), 492–498. <https://doi.org/10.1111/j.1467-7687.2008.00791.x>
- Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575. <https://doi.org/10.1038/nature04951>
- Liang, X. Y., Zebrowitz, L. A., & Zhang, Y. (2010). Neural activation in the “reward circuit” shows a nonlinear response to facial attractiveness. *Social Neuroscience*, *5*(3), 320–334. <https://doi.org/10.1080/17470911003619916>
- Little, A. C. (2014). Domain specificity in human symmetry preferences: Symmetry is most pleasant when looking at human faces. *Symmetry*, *6*(2), 222–233. <https://doi.org/10.3390/sym6020222>
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1638–1659. <https://doi.org/10.1098/rstb.2010.0404>
- Liu, C. H., Young, A. W., Li, J. X., Tian, X. R., & Chen, W. F. (2022). Predicting attractiveness from face parts reveals multiple covarying cues. *British Journal of Psychology*, *113*(1), 264–286. <https://doi.org/10.1111/bjop.12532>
- Lu, Y. J., Wang, J. M., Wang, L., Wang, J. L., & Qin, J. L. (2014). Neural responses to cartoon facial attractiveness: An event-related potential study. *Neuroscience Bulletin*, *30*(3), 441–450. <https://doi.org/10.1007/s12264-013-1401-4>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- Ma, D. S., Kantner, J., & Wittenbrink, B. (2021). Chicago Face database: Multiracial expansion. *Behavior Research Methods*, *53*(3), 1289–1300. <https://doi.org/10.3758/s13428-020-01482-5>
- Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*(3), 285–299. <https://doi.org/10.1093/scan/nsr090>
- Morton, J., & Johnson, M. H. (1991). CONSPEC And CONLERN: A two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181. <https://doi.org/10.1037/0033-295X.98.2.164>
- O'Doherty, J., Rutishauser, U., & Iigaya, K. (2021). The hierarchical construction of value. *Current Opinion in Behavioral Sciences*, *41*, 71–77. <https://doi.org/10.1016/j.cobeha.2021.03.027>
- Orghian, D., & Hidalgo, C. A. (2020). Humans judge faces in incomplete photographs as physically more attractive. *Scientific Reports*, *10*(1), Article 110. <https://doi.org/10.1038/s41598-019-56437-4>
- Perlovsky, L. (2014). Aesthetic emotions, what are their cognitive functions? [Opinion]. *Frontiers in Psychology*, *5*, Article 98. <https://doi.org/10.3389/fpsyg.2014.00098>
- Perlovsky, L., & Schoeller, F. (2019). Unconscious emotions of human learning. *Physics of Life Reviews*, *31*, 257–262. <https://doi.org/10.1016/j.plrev.2019.10.007>

- Quinn, P. C., Kelly, D. J., Lee, K., Pascalis, O., & Slater, A. M. (2008). Preference for attractive faces in human infants extends beyond conspecifics. *Developmental Science*, *11*(1), 76–83. <https://doi.org/10.1111/j.1467-7687.2007.00647.x>
- Ramanujan, V., Wortsman, M., Kembhavi, A., Farhadi, A., & Rastegari, M. (2020). *Whats hidden in a randomly weighted neural network?* [Conference Paper]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Proceedings (pp. 11890–11899). <https://doi.org/10.1109/cvpr42600.2020.01191>
- Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in Human Neuroscience*, *9*, Article 218. <https://doi.org/10.3389/fnhum.2015.00218>
- Reicher, M. E. (2015). *Einführung in die philosophische Ästhetik* [Introduction to philosophical aesthetics]. Wissenschaftliche Buchgesellschaft.
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*(1), 199–226. <https://doi.org/10.1146/annurev.psych.57.102904.190208>
- Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *The Journal of Neuroscience*, *30*(39), 12978–12995. <https://doi.org/10.1523/JNEUROSCI.0179-10.2010>
- Sadr, J., & Krowicki, L. (2019). Face perception loves a challenge: Less information sparks more attraction. *Vision Research*, *157*, 61–83. <https://doi.org/10.1016/j.visres.2019.01.009>
- Samuels, C. A., & Ewy, R. (1985). Aesthetic perception of faces during infancy*. *British Journal of Developmental Psychology*, *3*(3), 221–228. <https://doi.org/10.1111/j.2044-835X.1985.tb00975.x>
- Santayana, G. (1896). *The sense of beauty: Being the outline of aesthetic theory* (Vol. 238). Dover Publications.
- Sarasso, P., Neppi-Modona, M., Sacco, K., & Ronga, I. (2020). “Stopping for knowledge”: The sense of beauty in the perception-action cycle. *Neuroscience & Biobehavioral Reviews*, *118*, 723–738. <https://doi.org/10.1016/j.neubiorev.2020.09.004>
- Schmid, K., Marx, D., & Samal, A. (2008). Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios. *Pattern Recognition*, *41*(8), 2710–2717. <https://doi.org/10.1016/j.patcog.2007.11.022>
- Schoeller, F., & Perlovsky, L. (2016). Aesthetic chills: Knowledge-acquisition, meaning-making, and aesthetic emotions [Original Research]. *Frontiers in Psychology*, *7*, Article 1093. <https://doi.org/10.3389/fpsyg.2016.01093>
- Schoeller, F., Perlovsky, L., & Arseniev, D. (2018). Physics of mind: Experimental confirmations of theoretical predictions. *Physics of Life Reviews*, *25*, 45–68. <https://doi.org/10.1016/j.plrev.2017.11.021>
- Seeley, W. P. (2013). Art, meaning, and perception: A question of methods for a cognitive neuroscience of art. *The British Journal of Aesthetics*, *53*(4), 443–460. <https://doi.org/10.1093/aesthj/ayt022>
- Senior, C. (2003). Beauty in the brain of the beholder. *Neuron*, *38*(4), 525–528. [https://doi.org/10.1016/S0896-6273\(03\)00293-9](https://doi.org/10.1016/S0896-6273(03)00293-9)
- Shen, H., Chau, D. K. P., Su, J. P., Zeng, L. L., Jiang, W. X., He, J. F., Fan, J., & Hu, D. W. (2016). Brain responses to facial attractiveness induced by facial proportions: Evidence from an fMRI study. *Scientific Reports*, *6*(1), Article 35905. <https://doi.org/10.1038/srep35905>
- Slater, A., Bremner, G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. (2000). Newborn infants’ preference for attractive faces: The role of internal and external facial features. *Infancy*, *1*(2), 265–274. https://doi.org/10.1207/S15327078IN0102_8
- Slater, A., & Kirby, R. (1998). Innate and learned perceptual abilities in the newborn infant. *Experimental Brain Research*, *123*(1-2), 90–94. <https://doi.org/10.1007/s002210050548>
- Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, *21*(2), 345–354. [https://doi.org/10.1016/S0163-6383\(98\)90011-X](https://doi.org/10.1016/S0163-6383(98)90011-X)
- Stoianov, I., & Zorzi, M. (2012). Emergence of a ‘visual number sense’ in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196. <https://doi.org/10.1038/nn.2996>
- Tanaka, J. W., & Simonyi, D. (2016). SI Classics The “parts and wholes” of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*(10), 1876–1889. <https://doi.org/10.1080/17470218.2016.1146780>
- Thiruchselvam, R., Harper, J., & Homer, A. L. (2016). Beauty is in the belief of the beholder: Cognitive influences on the neural response to facial attractiveness. *Social Cognitive and Affective Neuroscience*, *11*(12), 1999–2008. <https://doi.org/10.1093/scan/nsw115>
- Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, *4*(3), 237–269. <https://doi.org/10.1007/BF02692201>
- Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460. [https://doi.org/10.1016/S1364-6613\(99\)01403-5](https://doi.org/10.1016/S1364-6613(99)01403-5)
- Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience*, *4*(3), 219–229. <https://doi.org/10.1038/nrn1055>
- Tracy, H. L. (1946). Aristotle on aesthetic pleasure. *Classical Philology*, *41*(1), 43–46. <https://doi.org/10.1086/362919>
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*(9), 989–995. <https://doi.org/10.1038/nn1111>
- Tsukiura, T., & Cabeza, R. (2011a). Shared brain activity for aesthetic and moral judgments: implications for the Beauty-is-Good stereotype. *Social Cognitive and Affective Neuroscience*, *6*(1), 138–148. <https://doi.org/10.1093/scan/nsq025>
- Tsukiura, T., & Cabeza, R. (2011b). Remembering beauty: Roles of orbitofrontal and hippocampal regions in successful memory encoding of attractive faces. *Neuroimage*, *54*(1), 653–660. <https://doi.org/10.1016/j.neuroimage.2010.07.046>
- Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns’ preference for faces: What is crucial? *Developmental Psychology*, *38*(6), 875–882. <https://doi.org/10.1037/0012-1649.38.6.875>
- Ullman, S., Harari, D., & Dorfman, N. (2012). From simple innate biases to complex visual concepts. *Proceedings of the National Academy of Sciences*, *109*(44), 18215–18220. <https://doi.org/10.1073/pnas.1207690109>
- Vartanian, O., Goel, V., Lam, E., Fisher, M., & Granic, J. (2013). Middle temporal gyrus encodes individual differences in perceived facial attractiveness. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 38–47. <https://doi.org/10.1037/a0031591>
- Wang, H. L., Guo, S. C., & Fu, S. M. (2016). Double dissociation of configural and featural face processing on P1 and P2 components as a function of spatial attention. *Psychophysiology*, *53*(8), 1165–1173. <https://doi.org/10.1111/psyp.12669>
- Wang, H. L., Sun, P., Ip, C. T., Zhao, X., & Fu, S. M. (2015). Configural and featural face processing are differently modulated by attentional resources at early stages: An event-related potential study with rapid serial visual presentation. *Brain Research*, *1602*, 75–84. <https://doi.org/10.1016/j.brainres.2015.01.017>
- Wang, J., Xia, T., Xu, L., Ru, T., Mo, C., Wang, T. T., & Mo, L. (2017). What is beautiful brings out what is good in you: The effect of facial attractiveness on individuals’ honesty. *International Journal of Psychology*, *52*(3), 197–204. <https://doi.org/10.1002/ijop.12218>
- Wilkinson, N., Paikan, A., Gredebäck, G., Rea, F., & Metta, G. (2014). Staring us in the face? An embodied theory of innate face preference. *Developmental Science*, *17*(6), 809–825. <https://doi.org/10.1111/desc.12159>
- Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*(5), 889–898. [https://doi.org/10.1016/s0896-6273\(04\)00728-7](https://doi.org/10.1016/s0896-6273(04)00728-7)

- Zhai, Y. K., Huang, Y., Xu, Y., Gan, J. Y., Cao, H., Deng, W. B., Labati, R. D., Piuri, V., & Scotti, F. (2020). Asian Female facial beauty prediction using deep neural networks via transfer learning and multi-channel feature fusion. *IEEE Access*, 8, 56892–56907. <https://doi.org/10.1109/ACCESS.2020.2980248>
- Zhan, J. Y., Liu, M., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2020, October 20–22). *A generative model of cultural face attractiveness*. 20th ACM International Conference on Intelligent Virtual Agents (ACM IVA) 2020, Electr Network.
- Zhang, D., Zhao, Q. J., & Chen, F. M. (2011). Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, 44(4), 940–950. <https://doi.org/10.1016/j.patcog.2010.10.013>
- Zhang, H. K., Wang, M., Luo, K., Zheng, Z. J., & Jiang, W. (2017). A study of the best fixed-point of facial proportions. *Medicine & Philosophy(B)*, 38(05), 64–69. <https://doi.org/10.12014/j.issn.1002-0772.2017.05b.18>
- Zhang, Y., Wei, B., Zhao, P. Q., Zheng, M. X., & Zhang, L. L. (2016). Gender differences in memory processing of female facial attractiveness: Evidence from event-related potentials. *Neurocase*, 22(3), 317–323. <https://doi.org/10.1080/13554794.2016.1151532>

Received August 2, 2023

Revision received February 13, 2024

Accepted February 23, 2024 ■

Supplementary Materials for Spontaneous Emergence of “A Sense of Beauty” in Untrained Deep Neural Networks

Tianxin Shu *et al.*

*Corresponding author. Email: delong.zhang@m.scnu.edu.cn

This PDF file includes:

Figure. S1 to S3

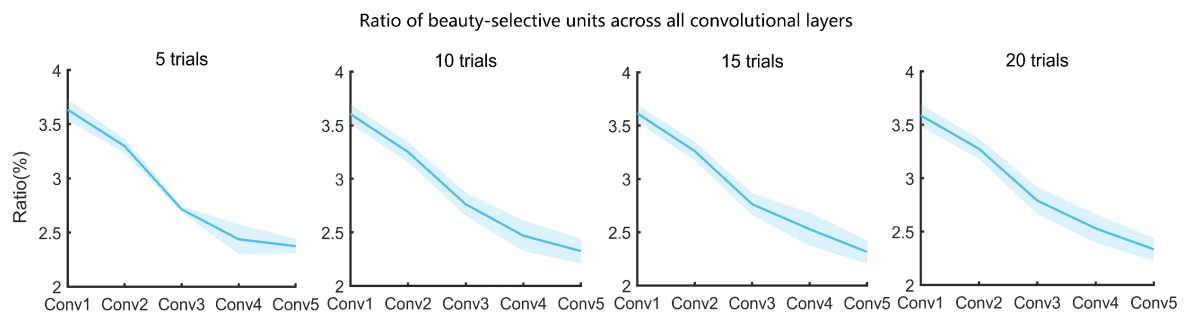


Figure S1. Validation analysis of the neural network's iterations. The proportion of beauty-selective units among the five convolutional layers remained consistent across 20 trials.

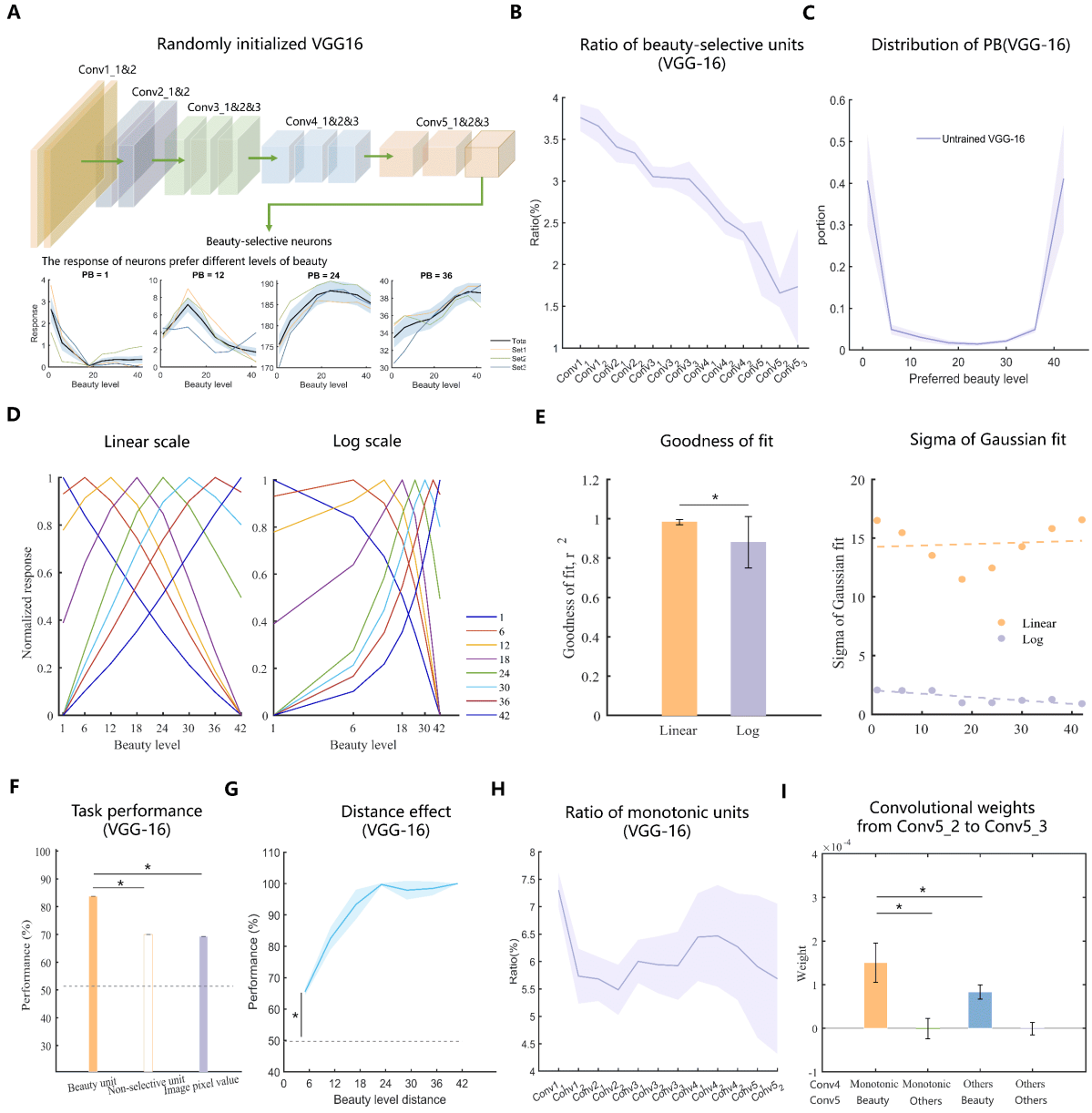


Figure S2. Validation analysis of untrained VGG-16. (A) Top: Architecture of randomly initialized VGG-16. Bottom: Examples of tuning curves for individual beauty-selective units observed in the untrained VGG-16. (B) The proportion of emerging beauty selective units in thirteen convolutional layers. (C) Distribution of preferred beauty in the untrained VGG-16. (D) Average tuning curves of different beauty levels on a linear scale and on a logarithmic. (E) Left: The goodness of the Gaussian fit (r^2) is greater on a linear scale. $*P < 7.77 \times 10^{-18}$, Wilcoxon rank sum test. Right: The tuning width (sigma of the Gaussian fitting) is U-shaped on a linear scale and remains constant on a logarithmic scale. (F) Task performance of the response of beauty units in untrained VGG-16, nonselective units, and the pixel values of raw stimulus images at the different level (orange open bar; beauty units versus nonselective units, $*P = 3.3 \times 10^{-4}$, Wilcoxon rank sum test; lilac solid bar; beauty units versus images, $*P = 1.37 \times 10^{-4}$, Wilcoxon rank sum test). (G) Beauty distance effect ($*P = 3.17 \times 10^{-6}$, Wilcoxon rank sum test). (H) The ratio of monotone units in the first twelve convolutional layers of the VGG-16. (I) Monotonic units in Conv5_2 provide stronger inputs to beauty units than to the other units in Conv5_3 (orange vs. green; $*P = 9.13 \times 10^{-5}$, Wilcoxon rank-sum test), implying that beauty tuning in Conv5_3 arises from the monotonic units in Conv5_2. Beauty units in Conv5_3 also connect to monotonic units more strongly than the other Conv5_2 units (orange vs. blue; $*P = 2.9 \times 10^{-3}$).

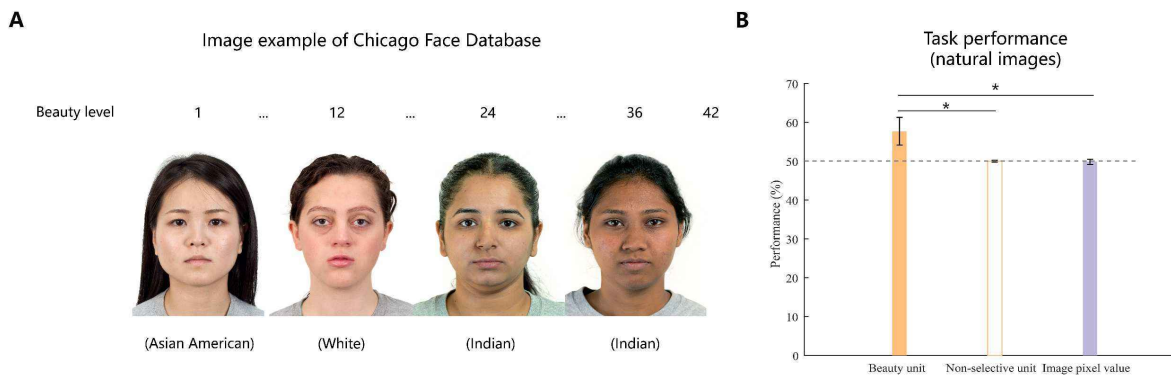


Figure S3. Validation analysis of the Chicago Face Database. (A) Examples of visual stimuli of different facial beauty in Chicago Face Database. (B) Task performance of the beauty units, nonselective units and the pixel values of raw stimulus images at the different level used the Chicago Face Database (orange open bar; beauty units versus non-selective units, $*P = 6.81 \times 10^{-5}$, Wilcoxon rank sum test; lilac solid bar; beauty units versus images, $*P = 4.05 \times 10^{-7}$, Wilcoxon rank sum test; Dashed line; beauty units versus chance level, $*P = 1.17 \times 10^{-6}$, Wilcoxon rank sum test). The dashed line indicates the chance level.