

Direct perception of affective valence from vision

Received: 21 August 2023

Accepted: 16 October 2024

Published online: 30 December 2024

Saeedeh Sadeghi^{1,2}✉, Zijin Gu³, Eve De Rosa¹, Amy Kuceyeski^{4,5} & Adam K. Anderson¹✉

Subjective feelings are thought to arise from conceptual and bodily states. We examine whether the valence of feelings may also be decoded directly from objective ecological statistics of the visual environment. We train a visual valence (VV) machine learning model of low-level image statistics on nearly 8000 emotionally charged photographs. The VV model predicts human valence ratings of images and transfers even more robustly to abstract paintings. In human observers, limiting conceptual analysis of images enhances VV contributions to valence experience, increasing correspondence with machine perception of valence. In the brain, VV resides in lower to mid-level visual regions, where neural activity submitted to deep generative networks synthesizes new images containing positive versus negative VV. There are distinct modes of valence experience, one derived indirectly from meaning, and the other embedded in ecological statistics, affording direct perception of subjective valence as an apparent objective property of the external world.

Subjective feelings are thought to be supported by the representation of internal interoceptive bodily states^{1–4}, which construct emotions when integrated with the higher-order conceptual meaning of external cues and retrieved memories¹. However, the term affect, as coined by the nineteenth-century psychologist W. Wundt, more broadly reflects the basic feeling ingredient underlying all sensations⁵. Affect, as originally conceptualized by Wundt, is a fundamental component of all experience, extracted as if a feature from the external world⁶. While bears and sharks may “look” dangerous or babies are approachable⁵, this is thought to be mediated by an evoked aversion or fondness at a deeper level of abstraction^{7,8}, intermixing with afferent bodily states to affect us^{1,2,4}. It remains an open question whether such perceptions of valence, the positive and negative dimension of affect and the primary dimension of meaning⁹, is a core component of visual experience. Borrowing from the ecological tradition of perception, we ask whether valence can be derived from the regularities of the environment, affording its direct perception¹⁰.

Valenced feelings may be correlated with the visual statistics of our environment¹¹ and with activity patterns in the visual system^{12–16}. Here we examined the extent and variation of these correlations and

their causal status in the brain and for behavior. Deep neural networks may afford decoding of affective dimensions and emotional categories from visual inputs¹⁴ by deriving high-level internal representations related to specific objects and associated conceptual meanings. We took a different approach by distinguishing two potential components of valence. We define normative valence (NV) referring to the valence acquired in common affective rating paradigms used in developing standard affective image databases involving realistic images viewed for multiple seconds. Accordingly, NV, at least in part, reflects the normative conceptual categorization of objects as “positive” or “negative”. In contrast to NV which depends on high-level conceptual analysis, we define and estimate Visual Valence (VV) as a superficial valence originating from a perceptual gist¹⁷. We hypothesize VV originates from basic compositional visual features, affording ecological visual statistics reflecting valence. The visual system might exploit such ecological visual correlates of valence for efficient representation of a critical aspect of the external world. If so, one might expect VV to have a causal status with distinct behavioral and neural markers from the conceptual analysis of valence that is dominant in NV¹⁸. The competing hypothesis is that VV reflects aliased NV, having no independent

¹Department of Psychology, Cornell University, Ithaca, NY, USA. ²Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. ³School of Electrical and Computer Engineering, Cornell University and Cornell Tech, Ithaca, NY, USA. ⁴Department of Radiology, Weill Cornell Medicine, New York, NY, USA. ⁵Department of Computational Biology, Cornell University, Ithaca, NY, USA. ✉e-mail: sdsd@caltech.edu; aka47@cornell.edu

causal role, originating from conceptual associations, which reside outside the visual system. As such, we examined whether VV and NV have competing or shared origins in computation, behavior and the brain. These investigations would shed light on the notion of subjective affect as directly embedded in visual experience⁶, with core affect as an emergent property^{1,19}.

Results

Study 1: A visual valence model (VVM)

We compiled a dataset of nearly 8000 photos from nine publicly available affective image datasets to examine whether low-level visual features predict valence rating norms, i.e., normative valence. Feeding all image pixels into deep models primarily trained for object recognition and using the last layer(s) for emotion prediction has led to high predictive accuracy^{14,20}, heavily reliant on conceptually rich representations in the final layers. Instead, here we selected a set of basic global features that demonstrate abstract composition, or perceptual gist, independent of specific objects. We selected 142 total features, including variances of outputs of the first convolutional layer of AlexNet²¹, average color, proportion of different colors, spatial frequencies, and symmetry, as well as control for the presence of human faces to track potential contributions from higher-level features. While the choice of feature sets was heuristic-based on the existing literature, they were all chosen such that the same array of feature values could come from numerous images with different or no clear semantic content. Accordingly, extraction of this limited set of features acted as a filter, transforming the initial image into an abstract space, that is, at least partially, blind to the precise objects in an image and thus its content.

Prior to any model training, images were transformed from the large pixel space into this substantially smaller space of global visual features. The Visual Valence Model (VVM, Supplementary Fig. 1), was a trained random forest model that used these low-level features to predict NV, i.e., the average valence rating of each image obtained from the available affective datasets and scaled to 1–9 (extremely negative to extremely positive). Across the 7984 image set, the out-of-sample predictions were correlated with ground-truth NV ratings, indicating a small to moderate effect size ($t(7982) = 31.2$, $p < 0.0001$, $r = 0.33$, 95% confidence interval = (0.31, 0.35)). Presence of faces did not appreciably change this relationship ($\Delta r = 0.016$, $z = 1.19$, $p = 0.23$), nor was the relationship sensitive to the choice of regression algorithm (see Supplementary Material). Excluding features derived from the first layer of AlexNet (variances and symmetry) while keeping all other features led to a slightly lower, yet still significant predictive accuracy ($t(7982) = 26.1$, $p < 0.0001$, $r = 0.28$, 95% confidence interval = (0.26, 0.30)).

We further examined the contribution of different feature subsets in predicting valence. We performed the same procedure separately on different feature sets (Supplementary Table 1 and Supplementary Fig. 2). All feature sets were significantly ($p < 0.0001$) predictive of NV in the absence of other features with the following order from the most to the least predictiveness: proportion of colors ($r = 0.26$), variances ($r = 0.20$), average color ($r = 0.19$), frequencies ($r = 0.13$), symmetry ($r = 0.07$; full statistics in Supplementary Table 1). Predictions of color proportion features had the highest similarity with model's (VVM) predictions that included all the features, confirming the highest contribution of color in VVM's efficiency (Supplementary Fig. 2).

The other basic dimension of affect, i.e., arousal (here Normative Arousal, NA), was also predictable from visual features following a similar procedure as VVM ($t(7982) = 28.6$, $p < 0.0001$, $r = 0.30$, 95% confidence interval = (0.28, 0.32)). NV and NA were negatively correlated in the emotional image dataset ($t(7982) = -44.1$, $p < 0.0001$, $r = -0.42$, 95% confidence interval = (-0.46, -0.42)), such that the more negative images were more arousing. The VVM's predictions of valence (here referred to as Visual Valence or VV), however, were largely unchanged when controlling for NA ($t(7982) = 29.6$, $p < 0.0001$, $r = 0.31$, 95% confidence interval = (0.29, 0.33)), confirming predictions of the pleasantness and unpleasantness from visual features was distinct from the intensity of feeling, which is traditionally associated with attentional salience.

We expected better than chance, albeit not too large an association between visual features and NV^{11,22} since human NV ratings also reflect object recognition and conceptual valence processing of scene content. This was also reflected in how the VVM predictions were much more restricted in range than human observers' NV, with the distributions differing significantly in variance (Fig. 1a; $F(7983/7983) = 14.2$, $p < 0.001$). In sum, the VVM supports evidence for the visual valence (VV) hypothesis, that visual features correlate with valence experience.

Study 2: Transfer of VV to abstract paintings

VVM perceptions may be merely a correlate of conceptual content access. If so, then attenuating image conceptual content should impair model performance. We next examined whether the VVM transferred to valence ratings of art, using abstract paintings as the test data. Even though we trained the VVM on real photos of emotional scenes, its predictions not only remained correlated with human NV ratings of abstract paintings ($N = 500$, each rated by 20 individuals²³) but also increased greatly in predictive association ($t(498) = 11.9$, $p < 0.0001$, $r = 0.47$, 95% confidence interval = (0.40, 0.53); Fig. 1c).

Comparing the distributions of valence ratings to the VVM revealed a more restricted range similar to abstract art (Fig. 1b), which

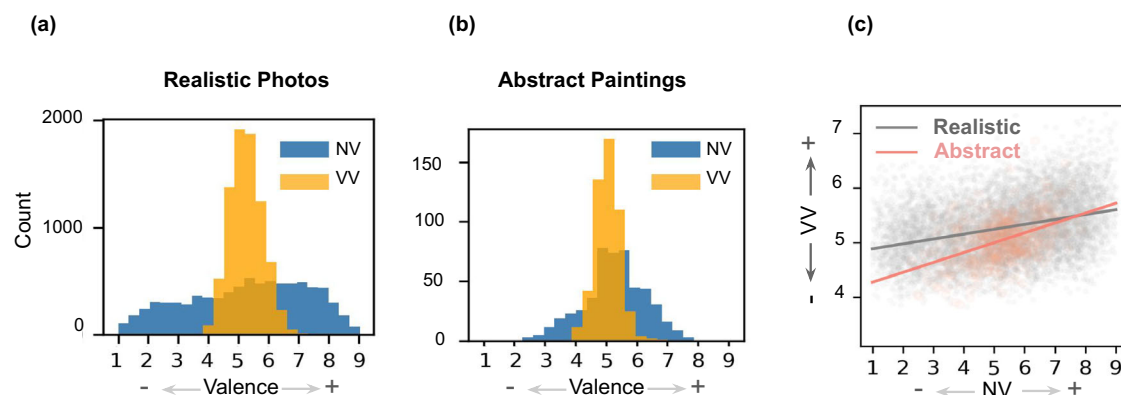


Fig. 1 | Normative valence (NV) and visual valence (VV) in study 1 and 2.

a Frequency distribution of NV and VV across realistic photos in study 1. **b** Distribution of NV and VV across abstract paintings in study 2. **c** Relationship

between VV (y-axis) and NV (x-axis) for realistic photos and abstract paintings. Each scattered point represents one image. The best-fitting lines are presented. Source data are provided as a Source Data file.

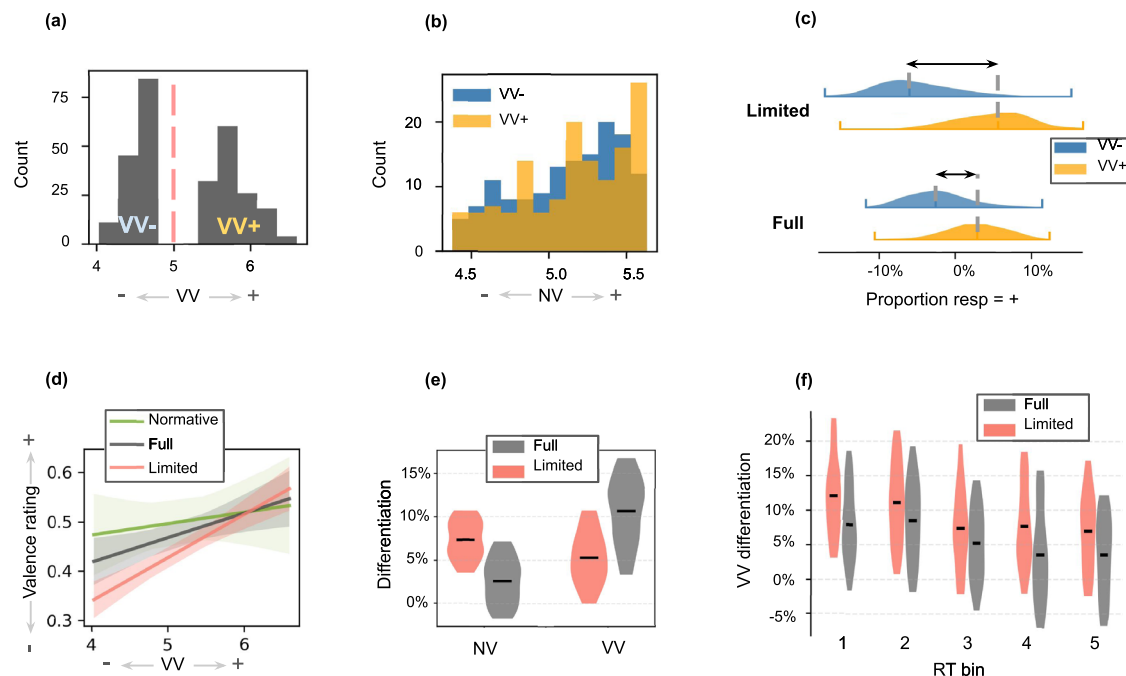


Fig. 2 | Study 3 results. **a** Visual Valence (VV) distribution for stimuli ($N = 280$). **b** Normative Valence (NV) distribution for VV+ and VV- stimuli. **c** Proportion of positive ratings for VV+ and VV- sets relative to subjects' baseline ($N = 87$ and 92 for full and limited viewing; center shows median). **d** Best-fitting lines showing the relationship between VV and different methods of subjective valence acquisition across 280 stimuli: normative ratings from the original dataset ($r = 0.04$, $p = 0.47$, green), full-viewing ($r = 0.16$, $p = 0.007$, gray), and limited viewing ($r = 0.34$, $p < 0.00001$, pink). Color shades show 95% CI. **e** Differentiation by NV or VV for full

vs. limited viewing across participants ($N = 87$ and 92 , respectively). Violin plots show the distribution of data between the 25th and 75th percentiles and the centers represent mean. **f** Differentiation by VV across Response Time (RT) quantile-based bins. Differentiation across subjects ($N = 87$, limited-viewing; $N = 92$, full-viewing) is shown for trials in each RT bin. Violin plots represent the distribution of data between the 25th and 75th percentiles, with centers indicating the mean. Source data are provided as a Source Data file.

differed from realistic photos with conceptual content (Fig. 1a; $F(7983/499) = 3.8$, $p < 0.0001$). VV now explained 20.1% of variance of NV in abstract paintings, nearly twice the amount of photos with conceptual content. These results are consistent with extremes in NV in photos reflecting human observer's access to conceptual content¹⁸. Without this access, machine and human observers demonstrated greater similarity in range and correspondence of valence experience. Since abstract art may contain attenuated conceptual content relative to real photos, VV may still be regulated by the processing of weak conceptual associations. Nonetheless, this provides evidence that VV has a causal role in valence experience. Use of VV may be limited by human observers' dependence on conceptual analysis for valence experience.

Study 3: Manipulating conceptual access

We next examined human observer's use of VV and its potential direct perception. Study 3 experimentally manipulated VV features and availability of conceptual processing during stimulus presentation. Direct perception was operationally defined as relative independence from access to higher-level conceptual representations for VV expression. We selected realistic photos ($N = 280$) with weak NV, i.e., varying marginally from neutral (5 ± 0.63 ; range width = 0.6 sd) on 1–9 scale but with strong VV, splitting the images into positive (VV+) and negative (VV-) sets based on VVM predictions (VV- range = 4–4.8 (width = 1.5 sd); VV+ range = 5.5–6.6 (width = 1.9 sd)) (Fig. 2a). By design, VV had little relation to NV in this subset of images ($t(278) = 0.67$, $r = 0.04$, $p = 0.50$) (Fig. 2b), such that access to VV and conceptual valence content would result in contrasting affective experiences.

To manipulate access to promote perceptual gist over conceptual processing, in one group (Limited-viewing condition, $N = 80$ ratings per image), images were inverted and viewing time restricted to 100 ms and requiring speeded judgments, while another group (Full-

viewing, $N = 80$ ratings per image) had unlimited exposure time and response window. Participants in both conditions were asked to report subjective valence with a 2-alternative forced choice response (positive or negative). Under these conditions, observers reported valence of images despite weak positive or negative NV (Fig. 2c, d). Under full-viewing conditions, the presence of weak conceptual valence content associated with NV explained valence responses better than VV (mixed-model fit comparison, $\Delta AIC = 92.7$). By contrast, in the limited-viewing condition, VV better explained responses than NV ($\Delta AIC = 286.5$; Fig. 2e). A significant interaction between VV and viewing condition confirmed that VV had a larger impact on valence responses during limited-viewing compared to full-viewing ($z = -5.5$, $p < 0.0001$, $\beta = -0.17$, 95% confidence interval = $(-0.23, -0.11)$); whereas an interaction in the opposite direction between NV and viewing condition revealed that NV having a greater impact during full-viewing ($z = 5.7$, $p < 0.0001$, $\beta = 0.33$, 95% confidence interval = $(0.22, 0.45)$). Overall, the image-by-image correspondence between machine and human observers' decoding of VV increased under limited and decreased under full-viewing (Fig. 2e).

While limited-viewing conditions would reduce access to high-level object or mid-level categories, we further accounted for the possible role of salient complex categories (i.e., presence of humans, animals, artifact vs natural, and indoor vs outdoor). The greater contributions of VV in limited viewing ($z = -5.4$, $p < 0.0001$, $\beta = -0.17$, 95% confidence interval = $(-0.24, -0.11)$) and effect of NV in full-viewing ($z = 5.8$, $p < 0.0001$, $\beta = 0.35$, 95% confidence interval = $(0.23, 0.47)$) held after statistically controlling for these salient content related effects. This is consistent with VV as generalized over many categorical features rather than specific high-level objects or scene inferences.

The salience of VV was not only limited to brief stimulus presentations, but it also predominated as observers made faster

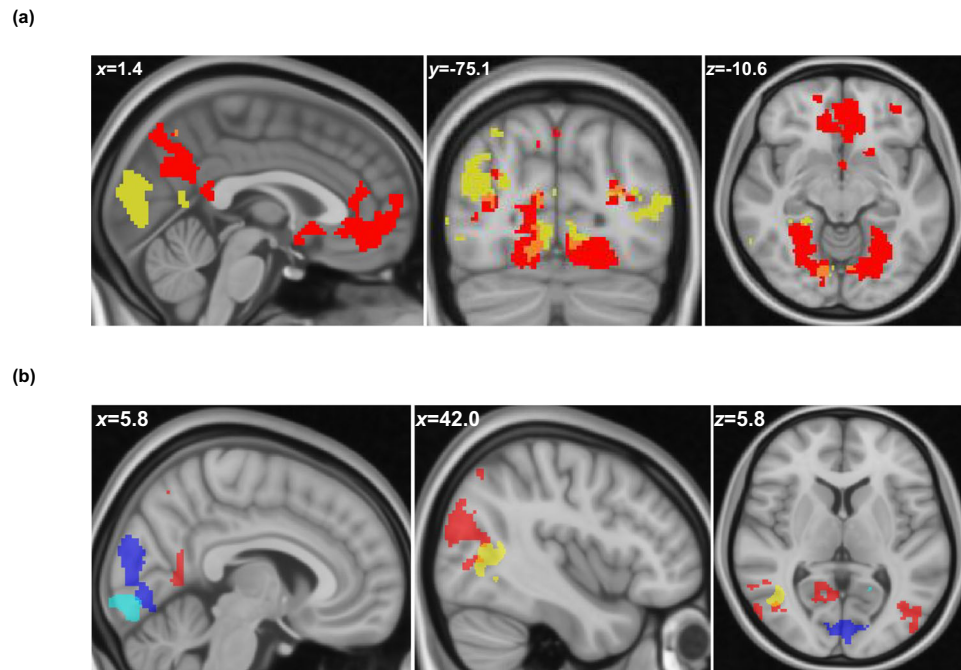


Fig. 3 | Neural correlates of normative and visual valence in study 4. a A computational metric of image visual valence (VV) corresponded with largely modality-specific posterior visual regions (yellow). Normative valence (NV) derived from observer ratings of valence experience revealed recruitment of higher-order multimodal regions traditionally associated with valence experience (red). Regions were defined by valence selectivity, increasing to either positive or negative

valence; whole-brain, corrected at false positive rate of 0.05. **b** Color-coded significant positive and negative visual valence clusters associated with Chikazoe data²⁴ (128 stimuli, 20 participants) and Natural Scene Dataset (NSD); whole-brain, corrected at false positive rate of 0.05; Chikazoe+: red; NSD+: yellow; Chikazoe-: dark blue; NSD-: light blue.

assessments. As expected, there were faster responses in the Limited-viewing compared to the Full-viewing condition ($z = 6.50$, $p < 0.0001$, $\beta = 0.32$, 95% confidence interval = (0.22, 0.42)). In general, faster valence judgments led to reports more driven by VV in both viewing conditions ($z = -3.7$, $p = 0.0002$; full-viewing: $z = -2.3$, $p = 0.02$, in limited-viewing: $z = -2.65$, $p = 0.008$; Fig. 2f). These effects were again persistent when controlling for higher-level content ($z = -3.9$, $p = 0.0001$; full-viewing: $z = -2.1$, $p = 0.04$; limited-viewing: $z = -2.7$, $p = 0.007$).

The double dissociation between valence type (visual versus conceptual) and processing mode (brief vs extended exposures and fast vs slow processing) demonstrates that although correlated in normal experience (Study 1 and 2), VV and NV are distinct competing valence signals, with NV related to deeper conceptual content and VV perceived directly from visual features, processed distinctly from mid-level scene types and conceptual categories.

Study 4: Neural bases of visual valence

If the VVM reflects a valid representation of valence embedded in visual feature space, then visual valence should reside in the visual system. By contrast, NV would correspond with regions traditionally associated with amodal valence^{24,25}. During fMRI BOLD scanning²⁴, participants ($N = 20$) viewed 128 images varying between extreme positive and negative emotional valence, rating valence experience under conditions of long exposure, favoring conceptual valence access. A whole-brain GLM univariate analysis of BOLD data identified brain regions whose activation, either positively or negatively, correlated with participant valence ratings for each image on a bipolar scale of most negative to most positive, where the average positivity minus negativity rating would reflect NV. Figure 3a shows the significant clusters (false positive rate; FPR = 0.05) associated with NV from the human observers and VV obtained from the VVM (see Supplementary Tables 2 and 3 for the full list of anatomical regions).

Although the stimulus selection was not based on the VVM, the model's predictions still transferred to this smaller image dataset, correlating with average subjective normative valence ratings (NV) across the images ($t(126) = 5.83$, $p < 0.0001$, $r = 0.46$, 95% confidence interval = (0.31, 0.58)). VV was not correlated with normative arousal ratings across images ($t(126) = 0.79$, $p = 0.4$, $r = 0.07$, 95% confidence interval = (-0.10, 0.24)). Despite the correlation between VV and NV, univariate substrates of VV overlapped with only 4% of voxels associated with NV (Supplementary Table 4). Specifically, NV (red in Fig. 3a) correlated with higher levels of the cortical hierarchy centered on the default mode network, including regions traditionally associated with the experience of valence across modalities. In addition, there were activations in regions associated with high-level visual scene content processing²⁶, including the lingual, fusiform and parahippocampal cortices. By contrast, VV (Fig. 3a yellow; Fig. 3b red and dark blue) correlated with lower-level modality-specific visual regions, including the occipital pole, intracalcarine cortex, and mid-level regions, including the lateral occipital cortices (LOC).

We performed further Region of Interest (ROI) analysis on Bilateral superior LOC (sLOC), which had the highest association with VV among visual regions (Supplementary Table 3), as well as medial prefrontal cortex (mPFC), which was highly associated with NV (Fig. 3a and Supplementary Table 2). Mean activity in the bilateral sLOC was significantly coupled with VV ($t(2525) = 3.1$, $p = 0.002$, $r = 0.12$, 95% confidence interval = (0.05, 0.21)) but not NV ($t(2525) = 0.8$, $p = 0.4$, $r = 0.005$, 95% confidence interval = (-0.007, 0.02)). This trend was generalized when examining the entire bilateral LOC, including both superior and inferior subregions VV: ($t(2525) = 2.6$, $p = 0.009$, $r = 0.1$, 95% confidence interval = (0.03, 0.2)); NV: ($t(2525) = 0.3$, $p = 0.8$, $r = 0.002$, 95% confidence interval = (-0.01, 0.01)). In contrast, mean activity in the mPFC was predictable from NV ($t(2525) = 3.9$, $p < 0.0001$, $r = 0.03$, 95% confidence interval = (0.01, 0.04)), but not VV ($t(2525) = 0.12$, $p = 0.9$, $r = 0.005$, 95%

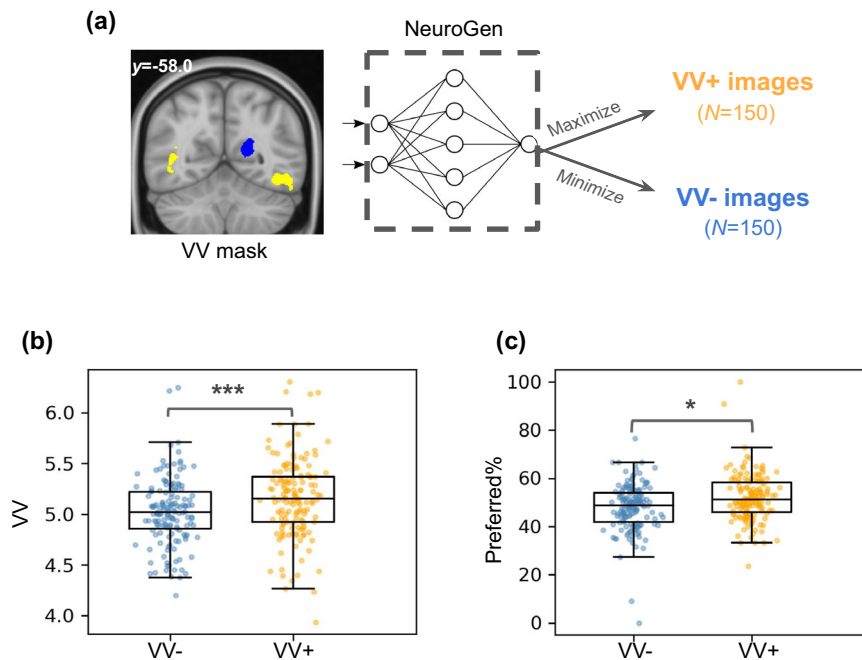


Fig. 4 | Synthesized image analysis in Study 5. a Image generation procedure: A mask of regions correlated with visual valence (VV) in the NSD dataset was fed into NeuroGen ($p < 0.001$ uncorrected, min cluster size = 10; yellow = positive, blue = negative correlation). NeuroGen synthesized VV+ and VV- image sets by maximizing or minimizing expected weighted sum of neural activation in the mask. **b** Within-class comparison of VVM's VV estimate for VV- and VV+ images ($N = 150$ each; two-sided paired t -test, $p < 0.0001$). **c** Behavioral preference between VV+ and

VV- images when shown side-by-side to participants. Each point shows a participant's average preference for VV- (blue) or VV+ (orange) synthesized images ($N = 128$ participants; two-sided paired t -test, $p = 0.01$). Preference is defined as the percentage of times images of one category (VV+ or VV-) were chosen as more positive. Boxplots show median ± 1.5 interquartile range overlaid by data points corresponding with images; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source data are provided as a Source Data file.

confidence interval = $(-0.08, 0.09)$). These results confirm the dissociation between the visual and frontal anatomical regions linked to VV and NV.

Psychophysiological interaction (PPI) analysis further revealed that the functional connectivity between mPFC and bilateral sLOC had a significant interaction with VV ($t(2523) = 2.7$, $p = 0.007$, interaction coefficient = 0.08 , 95% confidence interval = $(0.02, 0.14)$) but not NV ($t(2523) = -1.6$, $p = 0.11$, interaction coefficient = -0.008 , 95% confidence interval = $(-0.02, 0.001)$). This effect of VV generalized to the connectivity between mPFC and the entire bilateral LOC ($t(2523) = 2.15$, $p = 0.03$, interaction coefficient = 0.06 , 95% confidence interval = $(0.006, 0.12)$). The fact that this connectivity is specifically modulated by VV, although not a direct test of causal directionality, is in line with a potential bottom-up functional flow from VV embedded in mid-level visual processing to the mPFC that integrates a broader form of valence from different sources and modalities.

We next examined the reproducibility of the basic univariate findings in another dataset. We employed the open access 7T fMRI Natural Scenes Dataset of 8 participants viewing 8–10k images across 30–40 sessions distributed over one year (a total of 73k image stimuli)²⁷. Images were selected from various categories but were largely neutral in valence. Although the small number of participants in this dataset may challenge generalizability to new individuals, or the length of experiment may introduce habituation, the large stimulus set provides the opportunity to examine robustness of VV in the brain, generalized across tens of thousands of images. A small subset of images ($N = 100$) were rated on valence. When examining the transfer of VVM predictions to these images, we found a small but reliable correlation, ($t(98) = 2.23$, $p = 0.028$, $r = 0.22$, 95% confidence interval = $(0.02, 0.39)$). A group-level analysis in the entire dataset employing VV as predictors confirmed its representation primarily in the visual system (Fig. 3b; Supplementary Table 5). The larger number of stimuli allowed for the examination of regions that supported negative (VV-) and positive (VV+) valence with

greater reliability, not confounded by conceptual biases that potentially exist within a limited number of stimuli. Across datasets, we found that VV- was differentially associated with more posterior intracalcarine cortices and VV+ with more anterior fusiform and lateral occipito-temporal regions (Supplementary Table 5).

Beyond providing evidence for the distinct biological bases of VV and NV, it is evident that VV originates from the visual system. While the visual system is capable of representing high-level content²⁸, the current evidence suggests that the neural basis of VV relies more on decoding the external appearance of visual stimuli rather than internal neural representations of their conceptual attributes.

Study 5: Synthesizing images from visual system

While VV was largely correlated with activity in modality-specific visual regions, it is unclear if it is the generator of VV. In contrast to Study 4 that measured a brain response to stimuli, in Study 5 we reversed this approach, generating synthetic images to afford a visualization of images that maximize or minimize activation in visual regions most correlated with VV (Fig. 4a)²⁹. This also allowed for visualization of whether VV in the visual system did associate with any mid or high-level conceptual categories more than others. To do so, we employed NeuroGen²⁹, a brain-to-stimulus generative model, trained on the Natural Scenes Dataset.

Within the visual regions identified from the NSD, we considered voxels that maximally responded to positive (VV+) and minimally responded to negative (VV-) features and vice versa (Fig. 3b yellow and light blue). To determine whether these regions are generators of VV, the resulting synthesized images were then fed into the VVM. As a machine observer, the VVM discriminated image valence when visual features were optimized within classes of the same content (mean VV positive set = 5.16 , negative set = 5.02 ; paired t -test, $t(149) = 4.1$, $p < 0.0001$, 95% confidence interval = $(0.07, 0.20)$, within-sample; Fig. 4b). VV+ and VV- images were also slightly different in terms of

visual arousal, such that the VV- set had larger visual arousal (mean VV positive set = 4.78, negative set = 4.86; paired t -test, $t(149) = -2.3$, $p = 0.02$, 95% confidence interval = (0.14, -0.01)). The effect of image set (VV+ vs VV-) on VV was intact when controlling for visual arousal (effect of being in VV- group: $t(297) = -3.6$, $p = 0.0004$, $r = -0.11$, 95% confidence interval = (-0.17, -0.04)), confirming that the two sets are distinguishable in terms of VV independent of visual arousal. We repeated the image synthesis procedure, including only regions of the early visual system. Even though the generative model was optimized for VV, the synthesized VV+ and VV- images generated from early visual system were again significantly different in terms visual arousal ($t(149) = -2.9$, $p = 0.004$, 95% confidence interval = (-0.17, -0.03)), but not VV ($t(149) = 0.1$, $p = 0.9$, 95% confidence interval = (-0.06, 0.07)). These results suggest that VV, even if weakly embedded in early visual system, is entangled with visual salience (arousal), while it is more distinguishable from salience in mid-level visual processing.

Images synthesized from the entire visual system were further examined for their capacity to evoke valence experience. Synthesized image pairs within the same class were 180° rotated and displayed briefly (600 ms) side by side one at a time while participants decided ($N = 50$ per image) which image felt more positive or negative (binary decision). VV+ images were chosen as more pleasant (less unpleasant) significantly more often compared to VV- images (Fig. 4c, mixed-effect logistic regression, $z = 2.1$, $p = 0.03$, $r = 0.11$, 95% confidence interval = (0.006, 0.22)). There was no significant interaction with whether a participant was instructed to indicate the more positive or negative image ($z = 0.6$, $p = 0.6$, $r = 0.06$, 95% confidence interval = (-0.15, 0.28)), favoring that the effect of VV on choice originated from perceiving valence rather than the visual salience driving choices of both positivity and negativity.

To further examine the potential categorical correlates of VV, we performed an analysis of image categories based on WordNet tree hierarchy across the entire 1000-item list of classes from which images were generated (horizontal bars in Fig. 5). The most informative high-level category was “Artifact” (52.2%) and “Living thing” (40.7%) with VV- and VV+ classes following the same division ($\chi^2(1) = 88.5$, $p < 0.0001$) (see Supplementary Table 6 for the full list of positive and negative classes). As demonstrated in Studies 2–4, VV does not explicitly depend on conceptual analysis from the perspective of the potential neural generators of VV, VV+ and VV- might reflect distinct forms of visual analysis associated with different features.

Together, these results confirm that neural activity in visual regions is not only correlated with VV but are sufficient for synthesizing VV back as visual inputs, with distinct regions in the visual system differentially supporting positive and negative features.

Discussion

This study goes further than a correlation between visual features and affect^{14,20} by demonstrating the direct causal role of visual affect and its potential unique signatures in the brain and behavior. We provide support for a distinct form of valence that arises from the process of visual perception. Study 1 showed that VV accounts for a small but reliable amount of variance in NV. Studies 2 and 3 supported a causal role of VV in subjective valence that is distinct from the conceptual aspects of NV; by manipulating conceptual content (from real to abstract), or display condition (from full to limited viewing) subjective valence shifted from association with NV towards VV. These results support not only the causality of VV in subjective experience but also the context of behavioral relevance: use of VV increased under conditions that decreased reliance on conceptual analysis and being more reflexive than reflective. Study 4 supported neural substrates of VV residing mainly in visual processing regions, aligned with direct affective vision, as opposed to neutral vision followed by conceptual associations. While these results alone cannot prove visual neural origin of valence, the visual neural correlates were further tested for

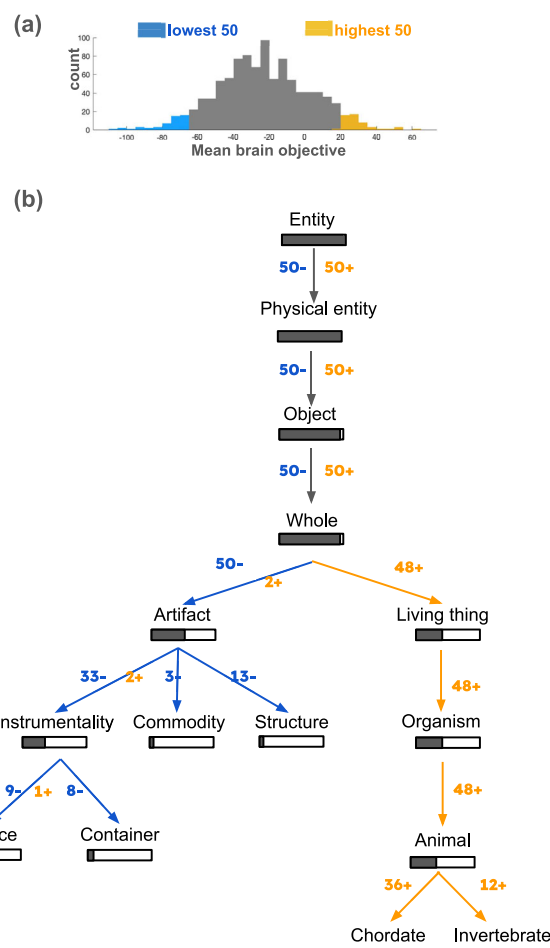


Fig. 5 | Analysis of class labels for synthesized images. **a** Distribution of the average neural objective function for the 1000 Alexnet classes. The top 50 largest and top 50 smallest classes represent entities evoking the most similar visual activity to the VV+ and VV- masks, respectively. **b** Hierarchical tree-structure semantic analysis of the labels of the top 50 positive (orange) and top 50 negative classes (blue). Horizontal gray bars represent the proportion of classes across the ImageNet's 1000 class list. Nodes are expanded as long as they correspond to at least 5% of the list (>20 classes). The number of classes that fall into each node of the tree is shown on arrows (blue with '-' showing the count out of the lowest 50, and orange with '+' showing the count out of the highest 50 classes). 'Artifact' versus 'Living thing' are the main nodes separating the two label sets. Source data are provided as a Source Data file.

causality in study 5, showing that synthesized VV+ and VV- images from the brain's visual system were distinguishable by model predictions as well as human observers.

Visual valence is related to but functionally distinct from high-level inferences about the content of what is perceived^{14,30–33}. This view is compatible with an emerging decentralized view of affect, with sensory modality-specific valence as a fundamental component of perceiving^{5,19}. The tradition of ecological vision¹⁰ argues visual perception develops as a result of navigating the external environment and discovering its inherent structure, more than forming and relying on internal higher-level inferences. Visual valence, while originating from real-world visual regularities associated with affective categories and concepts, we hypothesize is exploited to shape visual processing in the brain in a distinct manner. Just as texture gradients provide visual cues, or affordances, for perception of near or far distance, low or mid-level visual features can provide cues to positive or negative feelings. Events appear or look valenced, particularly when observers

have limited access to deep contents on which traditional valence depends.

In recent theorizations, salience is the main recognized scheme of the entanglement between the visual system and affective processing^{1,34}, supporting online reentrant feedback³⁵ and rapid scene analysis^{17,36}. Beyond salience, top-down input into visual processing regions may support representations of valence^{12–16}. Across two fMRI datasets, and thousands of images, we demonstrate a computational metric of valence correlated with activity within modality-specific visual regions, largely devoid of higher-order inputs strongly engaged by normative valence. Visual valence may then represent affective priors embedded in the visual system that depend upon intrinsic activity and reentrant feedback within visual regions. This may have similar mechanisms as in rapid object or scene category detection through coarse visual analysis¹⁷. The fact that VV was felt even in prolonged viewing of abstract paintings and reinforced in rapid viewing of realistic scenes suggests that valence, rather than a specific object category, is being directly computed from coarse analysis. Just as the visual system has acquired mechanisms to detect objects rapidly, it may directly estimate ‘positive’ and ‘negative’ based on the average visual features of all ‘good’ and ‘bad’ objects and scenes based on ecological statistics.

Interoceptive modalities, including those that derive their stimulation from the proximal external senses, such as pain or caress³⁷, sweet and bitter taste³⁸ and maybe even fruity or rancid smells³⁹, have sensory receptors that are evolutionarily tuned to information broadly consistent with valence^{19,40}. By contrast, in the exteroceptive senses, such as vision, there are no receptors for valence. They require discovering predictors of valence from the process of perception if such associations exist in the environment^{7,22,41}. This ecological perspective aligns with the Bayesian understanding of brain function⁴², where observers learn the transitional probabilities from seeing to feeling.

Visual valence only accounted for 5–20% of variance in valence. VV fails to predict valence of deep content at the individual item level much more often than not. Rather than serving as a valence predictor for specific events, VV would arise from an aggregate of experiences. As such VV operates as a distinct mode, competing with traditional valence, in which observers decode valence gist⁴³. The basis for such valence gist learning may originate from affective-motivational circuitry. Reward or punishment adaptively tunes the visual system, toward representation of predictive physical properties⁴⁴. This modified response is maintained even after the stimulus is no longer motivationally relevant^{45,46}, supporting long-term rewiring of the visual system. Until now, it has been thought that such learning alters the salience of associated physical features, rather than representing valence itself.

The function of visual valence aligns with an implicit versus explicit distinction in memory⁴⁷ and dual system theory⁴⁸, suggesting a fast, emotional heuristic system I, in contrast to a slow, deliberate system II, with the former supporting implicit affective judgements⁴⁹ and the latter associated with explicit emotional thoughts and memories¹. Implicit preferences need not require explicit conceptual inferences^{49,50}. Visual valence embedded in the global perceptual features can acquire an affective charge of its own, playing a role even when conceptual content may indicate to feel differently. VV would guide feelings and behavior in a more implicit way, similar to how facial features influence judgments of attraction or trustworthiness without access to knowledge about traits^{51,52}. Unlike faces, the present studies unbind visual features from specific localized objects to highly abstract universal features of the environment. Valence features may be represented in each perceptual system^{19,40} and may even have multiple representations within one perceptual system, such as for faces or body shapes.

Affect can arise from different levels of abstraction, ranging from conceptual processes rooted in identifying concrete objects and their relationships within scenes to experiences more closely tied to

abstract lower-level visual patterns. Combination of feedforward and recurrent feedback loops that integrate both concrete and abstract visual processing shape a complex decentralized network from which affect emerges. This distributed perspective aligns with theories on aesthetic appreciation of artworks⁵³, or visual appeal of nature^{54,55}. We show that VV resides in visual regions that distinguish the natural from artificial in synthesized images and, as such, relates to biophilia, i.e., the preference for nature⁵⁴. That valence is derived from visual features across highly distinct domains, from explicitly emotionally evocative events and abstract art to scenes from the natural and the built world, points to a fundamental convergence to which our brains are attuned. More than the contents of visual perception, visual valence may align with a way of seeing⁵⁶.

Methods

Online behavioral experiments in Studies 3 and 5 were determined to be exempt from Institutional Review Board (IRB) approval by the Cornell University IRB. Previously conducted fMRI experiment in study 4 had been approved by the University of Toronto Research Ethics Board and Sick Kids Hospital Research Ethics Board²⁴. The experimental protocol for collecting NSD fMRI dataset had been approved by the University of Minnesota institutional review board²⁷. All participants provided informed consent prior to their participation in these studies. We collected participants’ age and gender as reported in each study’s methodology. However, participant demographics were not included as covariates in our analysis, as they were not directly relevant to our main hypothesis.

1—Developing VVM

Realistic image database. A database of real-world naturalistic photographs was compiled by combining the following datasets: EmoMadrid⁵⁷, Emotional Picture System (EmoPics⁵⁸), Geneva Affective Picture Database (GAPED⁵⁹), International Affective Picture System (IAPS⁶⁰), Nencki Affective Picture System (NAPS⁶¹), Erotic subset for (NAPS ERO⁶²), Open Affective Standardized Image Set (OASIS⁶³), The Set of Fear Inducing Pictures (SFIP⁶⁴), Socio-Moral Image Database (SMID⁶⁵). Duplicate images were detected and excluded. Grayscale images were also excluded. The result was a set of 7984 images. Average valence and arousal ratings of images from each dataset were rescaled from their original range to the 1-to-9 scale, to be comparable across different datasets (1 indicating most negative and 9 indicating most positive). All analysis codes were implemented in Matlab 9.10.0

Visual feature extraction. The following visual features were extracted as the model input primarily based on the existing studies on aesthetics and emotion^{11,66,67}.

Average color. Colored images in the RGB space were transformed into HSV color map, and the average hue, average saturation, and average brightness across all pixels were estimated (3 features).

Color names. The proportion of pixels labeled as each basic color name was estimated based on a previous study⁶⁸, which provides a mapping from each point in the RGB space into one of 11 basic colors as perceived and labeled by humans (black, blue, brown, green, gray, orange, pink, purple, red, white, yellow). The result was 11 features indicating the amount of each color in an image.

Frequency. Spatial frequency amplitudes are employed by the visual system for rapid visual processing to obtain an initial sketch of scenes^{66,69,70}. These amplitudes reflect the texture of the image in different frequencies corresponding to different orientations. We followed the method used in ref. 66 for extracting frequency amplitudes based on the discrete Fourier transform. Briefly, images were first converted to grayscale. Then the 2-D space was divided into 4

orientations by 4 lines having 0, $\pi/4$, $\pi/2$, or $3\pi/4$ angles from the horizon. Each orientation line was then divided into 20 equally spaced points on a logarithmic scale. The amplitude in each point was interpolated, resulting in $20 \times 4 = 80$ total features.

Symmetry. Symmetry was estimated using the method proposed in ref. 21, based on filters from a Convolutional Neural Network (CNN) that take colors, edges, textures, shapes and objects into account, approximating the function of early visual processing in humans. We included the left-right, up-down, and left-right-up-down symmetry (3 features).

Statistical variances. We adopted a previous approach⁶⁷ to assess the variances in the response of the first convolutional layer of a CNN, trained on object recognition. This layer comprises 96 filters that detect edges at various frequencies and orientations or different color gradients, similar to early visual processing in the brain. To extract features, each image was segmented into $n \times n$ equally sized subregions, where n assumed 15 distinct values ranging from 2 to 30 (2, 4, 6,...,30). For each n , we computed three types of variances, culminating in a total of $15 \times 3 = 45$ features: (1) Richness: The overall variance computed across all filters and subregions, indicating the variability of features throughout the image subregions. (2) Median Regional Variability: For each filter, the variance is calculated across all subregions. The median of these values is then taken across all filters, reflecting how variance in filter responses differs from one subregion to another. (3) Variability: For each subregion, the variance is calculated across all filter responses. The median of these variances across all subregions then quantifies the diversity in visual statistics within individual subregions. These statistical variances provide a measure of both the uniformity and diversity in the visual characteristics of the image's subregions when the image is segmented into subregions of different sizes (i.e., with an increase in n , subregions of the grid get smaller while they would be higher in number). For instance, if an image contains a single object at the center with a solid background, size of the object in the image would influence variances for different n 's (grid sizes). If an image represents a uniform texture with no clear bounded objects, then the between region variances would be low independent of grid size n . Additionally, if the image or a subregion contains a rich and diverse set of visual features, then the filters are expected to make diverse responses within that region, leading to a high cross-filter variance within that region.

Faces. Three features were extracted related to faces in an image: number of faces, size of the largest face relative to the image size, total size of all faces relative to image size. We used the MTCNN toolbox in Matlab to automatically detect faces and compute the value of these three features for each image.

Random forest regression. We used the random-forest algorithm with 10-fold cross-validation to test the predictability of valence or arousal from each of the feature sets in the real-world image database. A random forest is an ensemble of decision trees each trained on a subset of features. The random forest's prediction for a set of input features is then the average prediction of all trees. While individual trees tend to overfit, this averaging prevents overfitting to the training data. Random forest was implemented using Matlab's regression tree ensembles with 500 ensemble trees. We defined the prediction accuracy of a random forest or linear regression model as the Pearson correlation coefficient between the out-of-sample predictions and ground-truth values.

2–Abstract paintings

Abstract artwork consisted of a database of 500 paintings from the Museum of Modern and Contemporary Art of Trento and Rovereto's

(MART) collection²³. With a total of 100 participants (74 females; ages 18–65; mean = 39.87), each painting in the MART database had been rated by 20 individuals in terms of emotional valence on a 1 (highly negative) to 7 (highly positive) scale. Raters did not see any information about the creator of the artwork or its title. We estimated the average valence rating of each painting as the ground truth. Valence ratings were rescaled to the 1–9 range to match the real photo database rating scale.

3–Full- and limited-viewing

Stimuli. We selected the most neutral images from the realistic image dataset, which was used to train VVM. The first criterion to select images was to have an arousal rating smaller than 3.5 (on a 1–9 scale). Among these low-arousing images, we selected $N = 600$ images with the most neutral valence rating (valence in a symmetrical range centered at 5). The valence range around the rating of 5 that included 600 images was between 4.37 and 5.63. Out of these 600 images, we then selected $N = 140$ with the highest VV, based on VVM's predicted valence (VV+ set) and $N = 140$ with the lowest VV (VV– set). The two image sets were further controlled for not having significantly different subjective valence and arousal based on the average human ratings in the original dataset (positive set valence: mean = 5.16, sd = 0.35; negative set valence: mean = 5.11, sd = 0.34; positive set arousal: mean = 2.42, sd = 0.68, negative set arousal: mean = 2.32, sd = 0.65). The two sets were also not statistically different in terms of visual arousal based on the random forest model that was trained to predict arousal with the large realistic image database (positive set mean = 4.56, negative set = 4.66, $t = -1.8$, $p > 0.05$). The positive valence set had an average visual valence of 5.78, significantly higher than the negative image set, which had an average visual valence of 4.56 ($p < 0.0001$). This procedure resulted in two sets of images that were indistinguishable in terms of human affective ratings, with one set having positive VV and the other having negative VV, merely based on VVM's predictions (Fig. 2a). The entire stimulus selection procedure was conducted with a Matlab script, with no active involvement of researchers, that could potentially bias the selection.

Participants. A total of 180 individuals participated in the study (42 male; age mean = 21.7, sd = 5.7). Participants included paid Amazon Mechanical Turk workers ($N = 25$), and Cornell University undergraduate students participating for course credit ($N = 155$). Amazon Mechanical Turk workers were paid 1\$ for rating each 56 images. Inclusion criteria for Mechanical Turk workers were having done at least 100 tasks on the Mechanical Turk website with at least a 95% approval rate.

Task and procedure. There were two valence rating conditions: Full-viewing and Limited-viewing. Each participant was randomly assigned to either condition (between-subject design). In each trial of both conditions, an image was displayed and the participant had to determine whether the image induced more positive versus more negative feeling (2-alternative valence response). In the Full-viewing condition, in each trial, the image was displayed until a response key was pressed with no time limit. In the Limited-viewing condition, the image was displayed upside-down and only for 100 ms, followed by a 3500 ms response time limit. Participants were instructed to enter their responses with 'A' and 'L' keys on the keyboard corresponding to positive and negative valence. All images were resized to have a width of 400 pixels while maintaining the original aspect ratio. The 280 image stimuli were randomly divided into 5 blocks of 56 images (28 from the negative and 28 from the positive set). Each participant performed between 1 to 5 blocks. Each stimulus was in total rated by 80 individuals in the Full-viewing and 80 individuals in the Limited-viewing condition. All conditions and instructions were implemented in Qualtrics online platform (with embedded HTML/JavaScript).

Participants performed the task in a remote online mode on a computer browser.

Analysis. Mixed-effect regression was used to analyze the results to account for within and between individual variability in responses or response times. In order to take mid and high-level conceptual categories into account, the set of 280 image stimuli was rated in terms of presence of humans (binary), animacy (binary), naturalness (1 = very artificial, 7 = very natural), and representing outdoors (1 = completely indoors, 7 = completely outdoors). Two raters independently performed the ratings. Internal consistency of the two ratings was confirmed (Cronbach's $\alpha > 0.9$ for all categories), and the average was used in the analysis. To control for the potential effect of these categories on VV in study 3, the interaction of each of these with the viewing condition (limited versus full-viewing) was included in the mixed-effect logistic regression predicting valence response. All analysis of study 3 was implemented in Python 3.11.3 and R 4.3.0.

Among the entire data, in the limited-viewing condition 1.6% of trials were timed out and dropped from further analysis.

4-fMRI experiment

Data collection and participants. Previously collected data from a published study from our lab consisting of 16 participants was included (the visual experiment of ref. 24). Additional data was collected for 4 individuals following the same procedure, resulting in a total of 20 participants (10 male; age mean = 26.1, sd = 2.1) included in the analysis.

Task and procedure. We analyzed data from the visual experiment of a previous study²⁴. Stimuli were a set of 128 colored images from the IAPS database, consisting of an equal number from each category of fear, disgust, neutral, or pleasant. In each trial of the task, one image was displayed for 3 s, followed by a blank screen for 4 s. Participants then rated the positivity and negativity of the image each for 3 s (on a scale of 1–7). There was an inter-trial interval of 4 s displaying a blank screen. Data was collected using a 3.0 Tesla fMRI machine with a repetition time (TR) of 2000 ms. Voxel size for the T1-weighted anatomical images was 1 mm and for the functional images was 3.5 mm in each dimension (see ref. 24 for more details about data collection).

Preprocessing. Regular preprocessing steps were conducted in SPM8, followed by normalization of each individual's functional images into the MNI space (interpolating data to 2 mm voxel dimensions)²⁴. An activation map was estimated on the un-smoothed data in response to the stimulus in each trial (using the canonical function in SPM). This resulted in one whole-brain t-map for each trial for each individual (20*128 total beta-maps). t-maps were normalized by subtracting the mean.

Whole-brain univariate analysis. Whole-brain univariate analysis was conducted to compare neural correlates of subjective valence and VV. Subjective valence for each image was defined as the participant's rating for positivity minus negativity. Note that VV for each stimulus was the same across all subjects (VVM's prediction based on low-level features), while subjective valence was distinct for each individual based on the participant's ratings. The Pearson correlation between each voxel's beta activation for each individual and visual valence across trials was estimated. The correlation coefficient was then converted to a z-score using Fisher transformation, resulting in one z-map for each subject. The z-maps were then tested across individuals using a two-tailed *t*-test to determine whether each voxel's value was significantly different from zero. Cluster thresholding was done using the 'ETAC' option (Equitable Thresholding And Clustering) in AFNI's 3dttest++ function, with a smoothing kernel size of FWHM = 6 mm. ETAC option for multiple comparison correction uses a voxelwise false

positive rate (FPR) of 5% at several different commonly used significance thresholds to find robustly significant clusters, not relying on arbitrary user parameters⁷¹. Both positive and negative clusters (voxels whose activity increased with either positivity or negativity) were considered as being sensitive to valence, while clusters were either positive or negative. A similar procedure was conducted to obtain whole-brain correlates of subjective normative valence. Structures associated with significant clusters were named based on probability maps from the Harvard-Oxford Cortical Atlas and the Harvard-Oxford subcortical atlas, using 'autoaq' script in FSL.

We also performed a second group-level univariate analysis across the 8 subjects NSD²⁷; see next section. We first estimated the z-score associated with the correlation between VV estimated for each stimulus and the beta activation in response to it. We then performed a one-sample *t*-test across the 8 subjects' z-maps. The voxel-level significance threshold was set at $p < 0.001$, and the minimum cluster size was estimated using AFNI's 3dclustsim routine to correct for multiple comparisons at $p < 0.05$.

ROI extraction: bilateral sLOC and bilateral LOC (superior+inferior) masks were obtained from the probabilistic Harvard-Oxford Cortical Structural Atlas in FSL, including voxels with at least 50% probability. mPFC ROI was created based on NeuroSynth meta-analysis maps (<https://neurosynth.org/>) using the keyword 'mPFC' (association test, false discovery rate < 0.01). To eliminate scattered voxels across the brain, the largest spatially adjacent set of voxels in the neurosynth-generated mask was used as the mPFC ROI.

ROI and PPI analysis were performed using the mixed-effect linear regression models ('lmer' package in R), with fixed coefficients for regressors and random intercept to account for individual differences in baseline activations.

5-Synthesizing images

We used a previously developed generative framework, NeuroGen²⁹, to synthesize images. NeuroGen works by concatenating an image generator with neural encoding models by feeding the generator output image to the encoding model and obtaining the predicted activation. During the optimization, the gradient flows from the region's response back to the synthetic image and then to the noise vector. Thus images that can achieve targeted brain responses (predicted) can be then iteratively synthesized via optimization of the noise vector, which is the input to the generator. NeuroGen was previously shown effective in modulating macro-scale brain responses⁷². Here we follow previous work and use the generator from BigGAN-deep⁷³, a conditional generative adversarial network (GAN) that synthesizes images conditioned on ImageNet classes. There is a truncation parameter during noise vector sampling that controls the balance between image fidelity and variety. Our encoding models were trained with eight individuals' brain responses in the NSD. Different from previous work that uses NeuroGen at a regional level⁷⁴, here the plugged encoding model into NeuroGen was a whole-brain voxelwise model. The model had VGG19⁷⁵ as a backbone and a spatial and feature-factorized linear readout⁷⁶.

Our objective function was:

$$L = \sum_{i=1}^8 (\text{predicted_activation} \times \text{visual_mask} \times \text{valence_mask}) \quad (\text{Eq.1})$$

The visual mask was a combination of 24 early and higher-order visual regions in NSD, identified using floc and retinotopic mapping, and the valence mask was obtained by estimating correlation coefficient between VV and trial-by-trial beta map, followed by an uncorrected voxelwise threshold of $p < 0.001$ for each individual's map and a minimum cluster size of 10 voxels. While all the following analyses were based on the objective that includes the visual mask of all early, mid and high-level visual regions²⁹; we also once synthesized images including only early visual regions mask, to compare the results (early

visual regions included: V1v, V1d, V2v, V2d, V3v, V3d, and humanV4; v = ventral; d = dorsal).

For image generation, we first selected 150 image classes and then generated an optimal VV+ and VV- image within each class. The class selection was done by first creating class-representative synthetic images generated from 100 random initializations, with the truncation parameter set to 1 for each of the 1000 ImageNet classes. Predicted brain activation of each random image was then used to estimate the corresponding objective value. The 150 classes that achieved maximal variance of the objective were selected. This approach would obtain classes that are capable of evoking both positive and negative VV responses since they span a wider range of objective values.

We then performed two optimizations for each selected class, one maximizing the objective (VV+) and the other minimizing the objective (VV-), to generate an image pair. For each optimization, the class information was embedded as a one-hot code into the class vector and fixed, and the noise vector was initialized from a truncated normal distribution and optimized. The truncation parameter here was set to 0.4 to have a good balance of image fidelity and variety. Overall, this procedure resulted in 150 (classes) × 2 (VV+ or VV-) = 300 images.

We also performed a between-class analysis, examining the top 50 image classes whose random initializations led to the maximum average objective (positive classes), or the minimum average objective (negative classes). We analyzed the resulting class labels, semantically, obtaining the hierarchical tree of categories associated with positive and negative VV in the brain. This was done using the *ntlk* package in Python interfacing with WordNet lexical database.

The generated 150 image pairs were fed into VVM to estimate VV for within-class comparison. They were also rated for their relative valence in a behavioral experiment. For the behavioral experiment, the 150 image pairs were randomly divided into 3 blocks of 50. In each trial of the task, a fixation cross was displayed for 400 ms, followed by the image pair displayed with 180° rotation and side by side (with random order) for 600 ms. In half of the blocks, participants were instructed to indicate whether the left or right image was perceived as more positive, and in the other half, they had to choose the more negative image. The aim of this manipulation was to counterbalance any potential confound of salience on choice. A trial was missed with a timeout message if the participant did not respond within 5 s. Participants were recruited from the Amazon Mechanical Turk with similar inclusion criteria as study 3, with a compensation of 0.9\$ per block. Each block was completed by an average of $N = 51.3$ participants, and the experiment included a total of 128 participants ($N = 98$ male, age mean = 34.7, $sd = 9.2$), each performing between 1 to 3 blocks. An average of 3.8% of trials were missed due to timeout and excluded from the analysis. Additionally, trials with response times shorter than 100 ms (16% of trials) were considered anticipatory and excluded from the analysis. Out of the 150 image pairs generated by the Neurogen, 16 images were identified as containing large faces with distorted facial features (resembling facial injury or fictional ghosts). These faces were distorted primarily because of the generative model's inadequacy in synthesizing realistic faces. Accordingly, these images were excluded from the analysis due to the potential to elicit fear responses unrelated to the neural substrates of VV.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available on GitHub repositories (<https://github.com/saeedeh/Visual-Valence-Model>, <https://github.com/saeedeh/sensoryValence-onlineStudy>), through Zenodo^{77,78}, or from the first author upon request. These repositories contain the VVM trained random forest model, the features extracted for the data

used in this project, the stimuli used in study 3, an online demo of the full and limited-viewing condition tasks, and the behavioral data. Source data are provided with this paper.

Code availability

The code that supports the findings of this study are available online on GitHub repositories (<https://github.com/saeedeh/Visual-Valence-Model>, <https://github.com/saeedeh/sensoryValence-onlineStudy>), through Zenodo^{77,78} or from the first author upon request. Code includes a sample script for extracting visual features and predicting VV for use in future research.

References

- Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **12**, 1–23 (2017).
- Critchley, H. D. & Garfinkel, S. N. Interoception and emotion. *Curr. Opin. Psychol.* **17**, 7–14 (2017).
- Damasio, A. & Carvalho, G. B. The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* **14**, 143–152 (2013).
- Seth, A. K. Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **17**, 565–573 (2013).
- Wundt, W. M. & Judd, C. H. *Outlines of Psychology* (W. Engelmann, 1902).
- Anderson, A. K. Toward an objective neural measurement of subjective feeling states. *Psychol. Conscious. Theory Res. Pract.* **2**, 30–33 (2015).
- Damiano, C., Walther, D. B. & Cunningham, W. A. Contour features predict valence and threat judgements in scenes. *Sci. Rep.* **11**, 19405 (2021).
- Goetschalckx, L., Andonian, A., Oliva, A. & Isola, P. Ganalyze: toward visual definitions of cognitive image properties. In *Proc. IEEE/CVF International Conference on Computer Vision*, 5744–5753 (2019).
- Osgood, C. E., May, W. H. & Miron, M. S. *Cross-Cultural Universals of Affective Meaning*, Vol. 1 (University of Illinois Press, 1975).
- Gibson, J. J. *The Ecological Approach to Visual Perception*. (Houghton, Mifflin and Company, Boston, MA, US, 1979).
- Redies, C., Grebenkina, M., Mohseni, M., Kaduham, A. & Dobel, C. Global image properties predict ratings of affective pictures. *Front. Psychol.* **11**, 953 (2020).
- Čeko, M., Kragel, P. A., Woo, C.-W., López-Solà, M. & Wager, T. D. Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760–770 (2022).
- Gao, C. & Shinkareva, S. V. Modality-general and modality-specific audiovisual valence processing. *Cortex* **138**, 127–137 (2021).
- Kragel, P. A., Reddan, M. C., LaBar, K. S. & Wager, T. D. Emotion schemas are embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358 (2019).
- Miskovic, V. & Anderson, A. Modality general and modality specific coding of hedonic valence. *Curr. Opin. Behav. Sci.* **19**, 91–97 (2018).
- Shinkareva, S. V. et al. Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Hum. Brain Mapp.* **35**, 3558–3568 (2014).
- Oliva, A. & Torralba, A. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* **155**, 23–36 (2006).
- Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A. & Kron, A. Dissociating affective and semantic valence. *J. Exp. Psychol. Gen.* **146**, 924 (2017).
- Kryklywy, J. H., Ehlers, M. R., Anderson, A. K. & Todd, R. M. From architecture to evolution: multisensory evidence of decentralized emotion. *Trends Cogn. Sci.* **24**, 916–929 (2020).
- Conwell, C., Graham, D., Konkle, T. & Vessel, E. Purely Perceptual Machines Robustly Predict Human Visual Arousal, Valence, and Aesthetics. *J. Vis.* **22**, 4266 (2022).

21. Brachmann, A. & Redies, C. Using convolutional neural network filters to measure left-right mirror symmetry in images. *Symmetry* **8**, 144 (2016).
22. Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J. & Midden, C. J. H. Brightness differences influence the evaluation of affective pictures. *Cogn. Emot.* **27**, 1225–1246 (2013).
23. Yanulevskaya, V. et al. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In *Proc. 20th ACM International Conference on Multimedia*, 349–358 (2012).
24. Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114 (2014).
25. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* **35**, 121–143 (2012).
26. Price, C. J., Devlin, J. T., Moore, C. J., Morton, C. & Laird, A. R. Meta-analyses of object naming: effect of baseline. *Hum. Brain Mapp.* **25**, 70–82 (2005).
27. Allen, E. J. et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).
28. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
29. Gu, Z. et al. NeuroGen: activation optimized image synthesis for discovery neuroscience. *Neuroimage* **247**, 118812 (2022).
30. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
31. Pourtois, G., Dan, E. S., Grandjean, D., Sander, D. & Vuilleumier, P. Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked potentials mapping. *Hum. Brain Mapp.* **26**, 65–79 (2005).
32. Schupp, H. T. et al. Selective visual attention to emotion. *J. Neurosci.* **27**, 1082–1089 (2007).
33. Vuilleumier, P. & Driver, J. Modulation of visual processing by attention and emotion: windows on causal interactions between human brain regions. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 837–855 (2007).
34. Miller, M. & Clark, A. Happily entangled: prediction, emotion, and the embodied mind. *Synthese* **195**, 2559–2575 (2018).
35. Edelman, G. M. & Gally, J. A. Reentry: a key mechanism for integration of brain function. *Front. Integr. Neurosci.* **7**, 63 (2013).
36. VanRullen, R. & Thorpe, S. J. The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* **13**, 454–461 (2001).
37. Löken, L. S., Wessberg, J., Morrison, I., McGlone, F. & Olsson, H. Coding of pleasant touch by unmyelinated afferents in humans. *Nat. Neurosci.* **12**, 547–548 (2009).
38. Wang, L. et al. The coding of valence and identity in the mammalian taste system. *Nature* **558**, 127–131 (2018).
39. Lapid, H. et al. Neural activity at the human olfactory epithelium reflects olfactory perception. *Nat. Neurosci.* **14**, 1455–1461 (2011).
40. Todd, R. M., Miskovic, V., Chikazoe, J. & Anderson, A. K. Emotional objectivity: neural representations of emotions and their interaction with cognition. *Annu. Rev. Psychol.* **71**, 25–48 (2020).
41. Palmer, S. E. & Schloss, K. B. An ecological valence theory of human color preference. *Proc. Natl Acad. Sci. USA* **107**, 8877–8882 (2010).
42. Yuille, A. & Kersten, D. Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* **10**, 301–308 (2006).
43. Bookbinder, S. H. & Brainerd, C. J. Emotionally negative pictures enhance gist memory. *Emotion* **17**, 102 (2017).
44. Hickey, C. & Peelen, M. V. Neural mechanisms of incentive salience in naturalistic human vision. *Neuron* **85**, 512–518 (2015).
45. Miskovic, V. & Keil, A. Escape from harm: linking affective vision and motor responses during active avoidance. *Soc. Cogn. Affect. Neurosci.* **9**, 1993–2000 (2014).
46. Rhodes, L. J., Ruiz, A., Ríos, M., Nguyen, T. & Miskovic, V. Differential aversive learning enhances orientation discrimination. *Cogn. Emot.* **32**, 885–891 (2018).
47. Graf, P. & Schacter, D. L. Implicit and explicit memory for new associations in normal and amnesic subjects. *J. Exp. Psychol. Learn. Mem. Cogn.* **11**, 501 (1985).
48. Kahneman, D. *Thinking, Fast and Slow* (Macmillan, 2011).
49. Zajonc, R. B. Feeling and thinking: preferences need no inferences. *Am. Psychol.* **35**, 151 (1980).
50. Zajonc, R. B. On the primacy of affect. *Am. Psychol.* **39**, 117–123 (1984).
51. Stirrat, M. & Perrett, D. I. Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychol. Sci.* **21**, 349–354 (2010).
52. Thornhill, R. & Gangestad, S. W. Facial attractiveness. *Trends Cogn. Sci.* **3**, 452–460 (1999).
53. Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, K. & O'Doherty, J. P. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nat. Hum. Behav.* **5**, 743–755 (2021).
54. Kardan, O. et al. Is the preference of natural versus man-made scenes driven by bottom-up processing of the visual features of nature? *Front. Psychol.* **6**, 471 (2015).
55. Kellert, S. R. & Wilson, E. O. *The Biophilia Hypothesis* (Island Press, 1995).
56. Schmitz, T. W., De Rosa, E. & Anderson, A. K. Opposing influences of affective state valence on visual cortical encoding. *J. Neurosci.* **29**, 7199–7207 (2009).
57. Carretié, L., Tapia, M., López-Martín, S. & Albert, J. EmoMadrid: an emotional pictures database for affect research. *Motiv. Emot.* **43**, 929–939 (2019).
58. Wessa, M. et al. EmoPics: subjektive und psychophysiologische evaluation neuen bildmaterials für die klinisch-bio-psychologische forschung. *Z. Klin. Psychol. Psychother.* **39**, 77 (2010).
59. Dan-Glauser, E. S. & Scherer, K. R. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behav. Res. Methods* **43**, 468–477 (2011).
60. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. International affective picture system (IAPS): technical manual and affective ratings. *NIMH Cent. Study Emot. Atten.* **1**, 39–58 (1997).
61. Marchewka, A., Żurawski, Ł., Jednoróg, K. & Grabowska, A. The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behav. Res. Methods* **46**, 596–610 (2014).
62. Wierzbna, M. et al. Erotic subset for the Nencki Affective Picture System (NAPS ERO): cross-sexual comparison study. *Front. Psychol.* **6**, 1336 (2015).
63. Kurdi, B., Lozano, S. & Banaji, M. R. Introducing the open affective standardized image set (OASIS). *Behav. Res. Methods* **49**, 457–470 (2017).
64. Michałowski, J. M. et al. The Set of Fear Inducing Pictures (SFIP): development and validation in fearful and nonfearful individuals. *Behav. Res. Methods* **49**, 1407–1419 (2017).
65. Crone, D. L., Bode, S., Murawski, C. & Laham, S. M. The Socio-Moral Image Database (SMID): a novel stimulus set for the study of social, moral and affective processes. *PLoS ONE* **13**, e0190954 (2018).
66. Rhodes, L. J. et al. The role of low-level image features in the affective categorization of rapidly presented scenes. *PLoS ONE* **14**, e0215975 (2019).
67. Brachmann, A., Barth, E. & Redies, C. Using CNN features to better understand what makes visual artworks special. *Front. Psychol.* **8**, 830 (2017).
68. Van De Weijer, J., Schmid, C. & Verbeek, J. Learning color names from real-world images. In *Proc. 2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8 (IEEE, 2007).
69. Crouzet, S. M. & Thorpe, S. J. Low-level cues and ultra-fast face detection. *Front. Psychol.* **2**, 342 (2011).

70. Gaspar, C. M. & Rousselet, G. A. How do amplitude spectra influence rapid animal detection? *Vis. Res.* **49**, 3001–3012 (2009).
71. Cox, R. W. Equitable thresholding and clustering: a novel method for functional magnetic resonance imaging clustering in AFNI. *Brain Connect.* **9**, 529–538 (2019).
72. Gu, Z., Jamison, K., Sabuncu, M. R. & Kuceyeski, A. Human brain responses are modulated when exposed to optimized natural images or synthetically generated images. *Commun. Biol.* **6**, 1–12 (2023).
73. Brock, A., Donahue, J. & Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=B1xsqj09Fm> (2019).
74. Gu, Z., Jamison, K., Sabuncu, M. & Kuceyeski, A. Personalized visual encoding model construction with small data. *Commun. Biol.* **5**, 1382 (2022).
75. Simonyan, K. & Zisserman, V. Very deep convolutional networks for large-scale image recognition. In *Proc. International Conference on Learning Representations* (2014).
76. Klindt, D., Ecker, A. S., Euler, T. & Bethge, M. Neural system identification for large populations separating “what” and “where”. *Adv. Neural Inf. Process. Syst.* **30**, (2017).
77. saeedeh. saeedeh/Visual-Valence-Model: v1. Zenodo <https://doi.org/10.5281/zenodo.13345696> (2024).
78. Sadeghi, S. saeedeh/sensoryValence-onlineStudy: v1. Zenodo <https://doi.org/10.5281/zenodo.13345691> (2024).

Acknowledgements

We gratefully acknowledge David J Field, Xinyi Li, Elif Celikors, and Junichi Chikazoe for their valuable insights and helpful feedback throughout the course of this research.

Author contributions

S.S. and A.K.A. conceptualized the project and performed data analysis and modeling. S.S., A.K.A., and E.D. designed the study 3 experiment. Z.G. and A.K. performed the image synthesis in study 5. S.S. and A.K.A. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53668-6>.

Correspondence and requests for materials should be addressed to Saeedeh Sadeghi or Adam K. Anderson.

Peer review information *Nature Communications* thanks Judith Domínguez-Borràs, Steven Scholte and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints> **Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024