



---

# Large Language Models as Machines of Beauty: Cognitive Averaging, Latent Space Geometry, and the Entropic Foundations of Aesthetic Preference

Daniel Plate<sup>1\*</sup>, James Hutson<sup>2</sup>

Lindenwood University, USA.

---

## \*Corresponding Author

**Daniel Plate**

Lindenwood University,  
USA.

## Article History

Received: 10.09.2025

Accepted: 28.10.2025

Published: 13.12.2025

**Abstract:** This study advances the position that large language models (LLMs) and human perceptual systems are governed by a shared computational drive toward prototypicality, entropy reduction, and aesthetic coherence. Drawing on developmental evidence that infants exhibit early preferences for facial symmetry and averageness, the analysis situates aesthetic preference within broader research on processing fluency and predictive coding, emphasizing that biological perception rewards stimuli that reduce uncertainty and support efficient information compression. This foundation is used to examine how LLMs, through cross-entropy optimization, perplexity minimization, and latent space clustering, converge on high-density representational regions that operate as statistical prototypes of linguistic and conceptual categories. The examination shows that centroids within latent space function as computational counterparts to psychological prototypes, while attention mechanisms act as filters that amplify structured regularity and suppress idiosyncratic variation. Through the integration of perspectives from cognitive psychology, computational neuroscience, and machine learning, the study reframes aesthetic qualities as emergent properties of systems optimized to stabilize input and maximize predictive coherence. This perspective also clarifies phenomena such as mode collapse and embedding drift as consequences of excessive convergence toward prototypical structure, paralleling aesthetic degradation observed when biological systems over-attenuate variability. The significance of this argument lies in demonstrating that beauty can be modeled as a measurable outcome of intelligent information processing, linking infant cognition, neural prediction dynamics, and the generative capacities of artificial systems through the common logic of prototype formation and entropy minimization.

**Keywords:** *Large language models, Prototypicality, Processing fluency, Latent space geometry, Neuroaesthetics.*

---

## Cite this article:

Plate, D., Hutson, J., (2025). Large Language Models as Machines of Beauty: Cognitive Averaging, Latent Space Geometry, and the Entropic Foundations of Aesthetic Preference. *ISAR Journal of Arts, Humanities and Social Sciences*, 3(12), 37-47.

---



## 1. Introduction: Beauty, Prototypicality, and Entropy in Brains and Machines

Across much of aesthetic theory, beauty has been framed as a matter of subjective taste or cultural convention, even as empirical work has increasingly revealed robust regularities in what observers find aesthetically compelling across contexts. Contemporary neuroaesthetic syntheses emphasize that symmetry, balance, and structural coherence reliably predict aesthetic preference, even when modulated by expertise and cultural background (Frascaroli et al., 2024; Popescu & Holman, 2024; Soydaner & Wagemans, 2024). Developmental studies show that such regularities emerge early: infants display preferences for adult-rated attractive faces well before extensive socialization, and these preferences are systematically linked to mathematically averaged, prototypical configurations and to facial symmetry (Rhodes et al., 2002; Rubenstein et al., 1999). In other words, even nascent visual systems appear to favor central tendencies in the “space” of possible faces, treating prototypicality itself as attractive. Recent work in empirical and computational aesthetics extends these findings beyond faces, demonstrating that prototypical scenes and forms, quantified in high-dimensional feature spaces, tend to be judged more beautiful than atypical exemplars (Redies et al., 2025; Soydaner & Wagemans, 2024). Parallel developments in computational image aesthetics use machine learning (ML) models trained on attributes such as color harmony, compositional balance, and lighting to predict aesthetic scores, illustrating that algorithmic systems also learn to privilege structurally regular, high-density regions of visual feature space (Qi et al., 2025; Soydaner & Wagemans, 2024). Together, these examples challenge a purely relativist view of beauty and instead suggest that both human and artificial perceivers are attracted to statistically average, coherently organized patterns. The present study extends this claim by arguing that infant visual systems and large language (LLMs) models are tuned to prototypicality as a means of reducing representational uncertainty and stabilizing their respective perceptual or predictive states. Beauty, on this account, is not a mysterious surplus added after perception but an index of successful alignment between incoming stimuli and high-probability regions within a shared representational geometry shaped by symmetry, averageness, and internal coherence.

Processing fluency provides the central psychological construct that links this developmental and computational picture to the dynamics of prediction and entropy reduction in both biological and artificial systems. Classic work defined fluency as the ease with which information is processed and showed that higher fluency produces more positive evaluations, an idea that recent research has integrated with predictive processing and motivational accounts of aesthetic judgment (Reber et al., 2004). Contemporary reviews and empirical studies argue that aesthetic pleasure emerges when stimuli are processed more fluently than expected, given an observer’s prior experiences and epistemic goals, thereby linking fluency to prediction-error dynamics rather than mere familiarity (Gao & Liu, 2024; Popescu & Holman, 2024; Yoo et al., 2024). Experiments with complex digital scenes demonstrate that images affording “cognitive naturalness” and rapid integration, even when physically implausible, elicit heightened aesthetic pleasure compared with more realistic but less fluently processed alternatives (Hwang & Cho, 2025; Sarasso et al., 2020). Predictive processing theorists have generalized this logic by modeling the brain as a hierarchically organized generative system that

minimizes prediction error or variational free energy, with aesthetic experiences marking episodes of especially efficient entropy reduction under tight precision control (Frascaroli et al., 2024; Van de Cruys et al., 2024). Within this framework, beauty is increasingly understood as a state in which a system maintains an optimal balance between order and surprise, avoiding both trivial predictability and unmanageable chaos. Computational aesthetics operationalizes these ideas by relating measures of visual entropy, feature redundancy, and typicality to aesthetic ratings, revealing that preferred images cluster in restricted bands of complexity and regularity (Redies et al., 2025; Soydaner & Wagemans, 2024). In generative AI, analogous optimization criteria appear in loss functions that minimize cross-entropy and perplexity, effectively rewarding outputs that conform to entrenched patterns and thereby instantiate a machine analogue of processing fluency (Chávez-Autor, 2025; Qi et al., 2025; Xu et al., 2025). The present study builds on these converging strands to treat beauty, fluency, and entropy minimization as three aspects of a single computational imperative that shapes both infant gaze behavior and large-scale generative modeling.

Recent research on LLMs and multimodal architectures shows that systems trained purely on predictive objectives spontaneously organize their internal representations into clustered latent spaces with strikingly human-like conceptual structure. Multimodal LLMs develop low-dimensional embeddings in which object categories form cohesive clusters with identifiable central tendencies, and these embeddings correlate with activity in high-level visual cortical regions, indicating partial alignment between model-based and neural representational geometries (Du et al., 2025). Comparative work on lexical concepts demonstrates that text-only models recover much of the non-sensorimotor structure of human conceptual knowledge, while diverging in embodied dimensions, thereby clarifying the domains in which latent-space prototypes approximate human semantic averages (Xu et al., 2025). Studies in cognitive science and natural language processing further show that these models support category induction and typicality judgments that track graded membership patterns in human data, suggesting that distances to cluster centroids in latent space function as computational analogues of psychological prototypicality (Jones et al., 2024; Michels, 2024; Pedrotti et al., 2025). At the same time, work in artificial creativity and aesthetic assessment documents systematic tendencies toward homogenization, mode collapse, and over-reliance on high-density regions of latent space whenever generative systems are optimized solely for fluency and accuracy, with diversity and originality emerging as secondary or even sacrificed objectives (Chávez-Autor, 2025; Qi et al., 2025; Soydaner & Wagemans, 2024). Prompt-engineering frameworks that explicitly score and refine prompts along aesthetic dimensions such as color harmony, compositional balance, and lighting show that these models can be steered toward particular aesthetic prototypes, and that such prototypes are quantifiable and manipulable through structured interaction.

Against this backdrop, the present study proposes a formal mapping between cognitive averaging and vector centroids, arguing that both infant perception and LLM generation can be modeled as movements toward cluster centers in high-dimensional representation space under constraints of entropy minimization. The subsequent sections develop this mapping by moving from infant aesthetics to processing fluency and predictive coding, then

to latent space geometry, attention mechanisms, and generative mode collapse, culminating in a set of empirical predictions and design implications for human–AI systems that treat beauty as a mathematically tractable consequence of intelligent information processing.

## 2. Background: Infant Aesthetics, Processing Fluency, and Predictive Brains

### 2.1 Infant Preferences for Symmetry and Averageness

Research on early visual perception indicates that infants display systematic preferences for faces that adults also judge as attractive, suggesting that sensitivity to symmetry and averageness emerges before extensive cultural learning. Preferential-looking paradigms, in which relative gaze duration is used as an index of perceptual interest, show that infants allocate more attention to faces with balanced structural relations and regular feature configurations than to more distorted alternatives (Jeffery, Rhodes, & Bui, 2023; Walton, Bower, & Bower, 1998). Habituation designs further demonstrate that after repeated exposure to multiple individual faces, infants respond most strongly to averaged composites, implying that they build category-level representations anchored on central tendencies rather than on single exemplars (Kobayashi, Sato, & Minemoto, 2024; Quinn et al., 2001). Across these experiments, infants appear attuned to configural regularities such as bilateral symmetry, proportionality, and smooth statistical averaging across features, all of which reduce the representational complexity of the input space (Lee, Tanaka, & Burris, 2024). Developmental trajectories suggest that these preferences are modulated rather than replaced by experience, as older infants show more refined discrimination between prototypical faces, novel variants, and clearly atypical outliers (Slater et al., 2000). This pattern supports the interpretation that early preferences reflect a perceptual system biased toward representational efficiency, privileging central exemplars that minimize processing demands while still supporting exploratory engagement with deviations from the norm. Such findings position infant preferences not as epiphenomenal curiosities, but as early expressions of a predictive architecture that encodes and exploits prototypes in order to stabilize perception in noisy environments. In this way, infant sensitivity to symmetry and averageness provides a developmental starting point for theorizing beauty as a marker of high-fluency, prototype-centered representations across the lifespan.

### 2.2 Processing Fluency as a Basis for Aesthetic Preference

Processing fluency offers a powerful explanatory construct for understanding why prototypical or symmetric stimuli tend to be evaluated as more beautiful than their irregular counterparts. Experimental manipulations of symmetry, prototypicality, contrast, and repetition consistently show that stimuli which can be encoded and categorized with fewer computational operations elicit more positive aesthetic judgments, even when observers remain unaware of the underlying manipulations (Gao & Liu, 2024; Reber, Schwarz, & Winkielman, 2004). Contemporary studies further demonstrate that fluency operates across modalities, influencing the evaluation of images, music, and language, and that subtle changes in clarity or structural redundancy can shift judgments from pleasing to disfluent with no change in semantic content (Popescu & Holman, 2024; Yoo, Jasko, & Winkielman, 2024). Importantly, fluency can be deliberately destabilized through controlled deviations, such as irregular contours or unexpected

pattern breaks, allowing researchers to map nonlinear relations between predictability, attention, and pleasure. These results suggest that aesthetic experience often emerges near a boundary condition where structure is sufficiently predictable to be processed efficiently yet sufficiently rich to avoid triviality, a balance that mirrors optimization goals in other cognitive domains. Because fluency can be quantified using measures such as response time, error rates, and subjective ease, it lends itself naturally to computational formalization and to comparison with machine-learning metrics that index model confidence or surprise (Gao & Liu, 2024). This convergence supports an account in which aesthetic pleasure indexes representational economy: observers experience beauty when internal models compress incoming data effectively without sacrificing meaning. Such an interpretation positions fluency as a critical bridge between infant preferences for prototypical exemplars and the behavior of generative systems trained to minimize predictive error.

### 2.3 Predictive Coding, Efficient Coding, and Entropy Minimization

Predictive coding and efficient coding frameworks conceptualize perception as an inferential process in which the brain continuously predicts sensory input and updates its generative models to minimize prediction error or free energy. Within these accounts, hierarchical neural circuits send predictions downward and propagate residual errors upward, iteratively adjusting internal representations until they capture the statistical structure of the environment with maximal parsimony (Frascaroli et al., 2024; Sarasso et al., 2020). Central concepts such as prediction error, precision weighting, and hierarchical inference describe how sensory signals are selectively amplified or attenuated according to their expected informativeness, thereby reducing uncertainty and stabilizing perceptual experience (Van de Cruys et al., 2024). Neuroaesthetic applications of these models propose that aesthetic pleasure arises when prediction errors are reduced in an especially efficient manner, as when an initially complex stimulus yields to a satisfying, low-entropy interpretation that integrates multiple levels of structure (Redies et al., 2025). Empirical work supports this view by showing that artworks and scenes with intermediate complexity and well-organized redundancy, which is neither fully random nor trivially simple, are most consistently judged as beautiful, and that such stimuli occupy constrained bands of entropy and statistical regularity (Popescu & Holman, 2024). These results suggest that beauty can be modeled as an emergent property of entropy minimization under representational constraints, rather than as a purely subjective overlay on perceptual processing. Because generative models in ML are likewise trained to approximate probability distributions and reduce surprise in their outputs, predictive and efficient coding provide a shared conceptual framework for comparing biological and artificial systems. In both cases, aesthetically valued patterns appear to be those that best reconcile richness of structure with the drive to reduce uncertainty, a synthesis that will inform the subsequent mapping between cognitive prototypes and latent space centroids.

## 3. Large Language Models and Latent Space Geometry

### 3.1 Transformer Architectures and Training Objectives

Transformer-based language models operate through a layered sequence of computational stages that collectively enable the prediction of ordered linguistic structures, beginning with the

conversion of raw text into discrete units through tokenization. Once tokens are established, they are projected into continuous embedding vectors whose dimensional structure allows semantic and syntactic properties to be encoded in ways that support downstream contextual integration (Devlin et al., 2019; Radford et al., 2019; Vaswani et al., 2017). Within each Transformer block, the self-attention mechanism computes weighted interactions among all tokens in a sequence, enabling the model to track long-range dependencies without the recurrent constraints that characterized earlier architectures. Feedforward layers then refine these contextually enriched representations, while positional encodings maintain information about sequence order that would otherwise be lost due to the model's fully parallel computation. During training, Transformers optimize the next-token prediction objective by minimizing cross-entropy loss, thereby aligning internal probability estimates with observed linguistic distributions in large-scale corpora (Brown et al., 2020). This objective is tightly linked to the information-theoretic concept of surprise, with perplexity functioning as an exponential transformation of cross-entropy that indexes average uncertainty across predictions. Because models are trained on corpora spanning diverse genres, registers, and knowledge domains, their representational geometry adapts to broad statistical regularities within natural language, reinforcing attractor regions associated with high-frequency and contextually stable patterns. Through this iterative optimization, Transformer-based models develop internal structures that reflect the probabilistic contours of language itself, preparing the conceptual ground for analyzing latent clustering and geometric prototypicality.

### **3.2 Latent Representations, Clusters, and Vector Centroids**

Latent representations in these models emerge through the repeated transformation of embeddings across attention and feedforward layers, producing high-dimensional vector spaces in which semantically or functionally similar tokens and sentences cluster together. Studies using probing tasks, representational similarity analysis, and dimensionality reduction techniques consistently show that Transformer layers encode hierarchical linguistic distinctions, including parts of speech, semantic fields, and conceptual relations (Ethayarajh, 2019; Manning et al., 2020; Tenney, Das, & Pavlick, 2019). These representations exhibit graded structure, with dense regions corresponding to frequently co-occurring or conceptually central items that function as geometric prototypes within the embedding space. The centroid of a cluster, which is the mean vector representing its central tendency, captures the statistical average of items within a semantic category and serves as a reference point for measuring typicality or conceptual proximity (Li, Yin, Han, & Ji, 2020). Empirical work demonstrates that distances to these centroids

correlate with human judgments of category membership, conceptual centrality, and semantic coherence, indicating that model representations often mirror cognitive structures documented in psychological prototype theory (Tenney, Das, & Pavlick, 2019; Li, Yin, Han, & Ji, 2020). When visualized through t-SNE or UMAP projections, these high-dimensional structures reveal well-defined boundaries and attractor regions that resemble human conceptual clusters, with outliers occupying sparser regions associated with low-frequency or contextually specialized usage. Prototypicality in this geometric context thus becomes definable as local density or minimal distance to a centroid, providing a quantitative framework for linking computational averaging to human cognitive processes. This shared geometry lays the foundation for mapping emergent model prototypes onto the mechanisms of cognitive averaging found in human perception.

### **3.3 Attention as a Computational Filtering Mechanism**

Self-attention in Transformer architectures functions as a dynamic filtering process that selectively amplifies or suppresses specific aspects of the input sequence to optimize contextual inference. Each attention head computes interactions among tokens using learned query, key, and value vectors, producing weights that reflect the relative informativeness or relevance of each token for interpreting others in the sequence (Vaswani et al., 2017). Multi-head attention extends this process by allowing the model to compute multiple, parallel relevance structures that capture distinct relational patterns, such as syntactic dependencies or semantic alignment. This selective weighting can be understood as reducing effective entropy within the representational space, because attention narrows the set of plausible continuations or interpretations by concentrating probability mass on higher-precision signals (**Table 1**) (Clark et al., 2019; Manning et al., 2020). Neuroscientific models of attention describe a similar process, wherein precision weighting regulates the gain on prediction errors to privilege signals expected to be informative for hierarchical inference (Feldman & Friston, 2010; Parr & Friston, 2019). By analogy, self-attention modulates representational dynamics by assigning greater influence to features consistent with emerging contextual predictions while diminishing features that introduce noise or uncertainty. This mechanism supports the emergence of coherent, aesthetically fluent representations by smoothing idiosyncratic variation and amplifying structured regularities, thereby preparing the conceptual ground for framing attention as an aesthetic filter in later sections. Through this parallel between computational and neural mechanisms, attention becomes a central bridge between predictive coding, representational coherence, and the geometric organization of latent space.

**Table 1. Construct Mapping Between Biological Aesthetics and Generative Modeling**

Biological construct	Computational analogue	Quantitative metric	Predicted relation to aesthetic rating	Measurement paradigm
<b>Prototype / Typicality</b>	Vector centroid in latent space	Distance-to-centroid; local density	Lower distance → higher ratings (monotonic)	Compare human ratings with centroid distances in image or text embeddings
<b>Cognitive averaging</b>	Statistical mean representation learned through optimization	Cluster variance; embedding compactness	Lower variance → greater perceived coherence	Evaluate responses to averaged vs. non-averaged morphs; compare to model-averaged embeddings
<b>Processing fluency</b>	Perplexity minimization; prediction regularity	Perplexity; cross-entropy	Lower perplexity → higher fluency and liking (until over-fluency region)	Manipulate textual or visual predictability and collect fluency ratings
<b>Attention as precision weighting</b>	Self-attention weight distribution	Attention entropy; head-specific sparsity	Lower entropy → higher perceived clarity or coherence	Map attention-weight entropy to ratings of clarity or readability
<b>Prediction error</b>	Cross-entropy loss; token-level surprise	Surprise per token; gradient magnitude	Moderate error reduction → aesthetic peak (inverted-U)	Compare neural or behavioral prediction-error signatures with model cross-entropy
<b>Entropy / Uncertainty</b>	Latent-space entropy; log-probability landscape	Local entropy; output-entropy; diversity indices	Mid-range entropy → highest ratings (inverted-U); too low → sterility	Test aesthetic ratings across images/text sampled at varying entropy levels
<b>Aesthetic preference / Pleasure</b>	Reward signal linked to low entropy and structured predictability	Representational similarity; centroid proximity; stability indices	Greater structural coherence → higher ratings	Align human liking ratings with embedding similarity measures
<b>Falsifiable predictions</b>	—	—	Prototype-distance monotonicity for faces/scenes; inverted-U for language and complex art	Joint human–model experiments comparing centroid distance, entropy, and preference curves

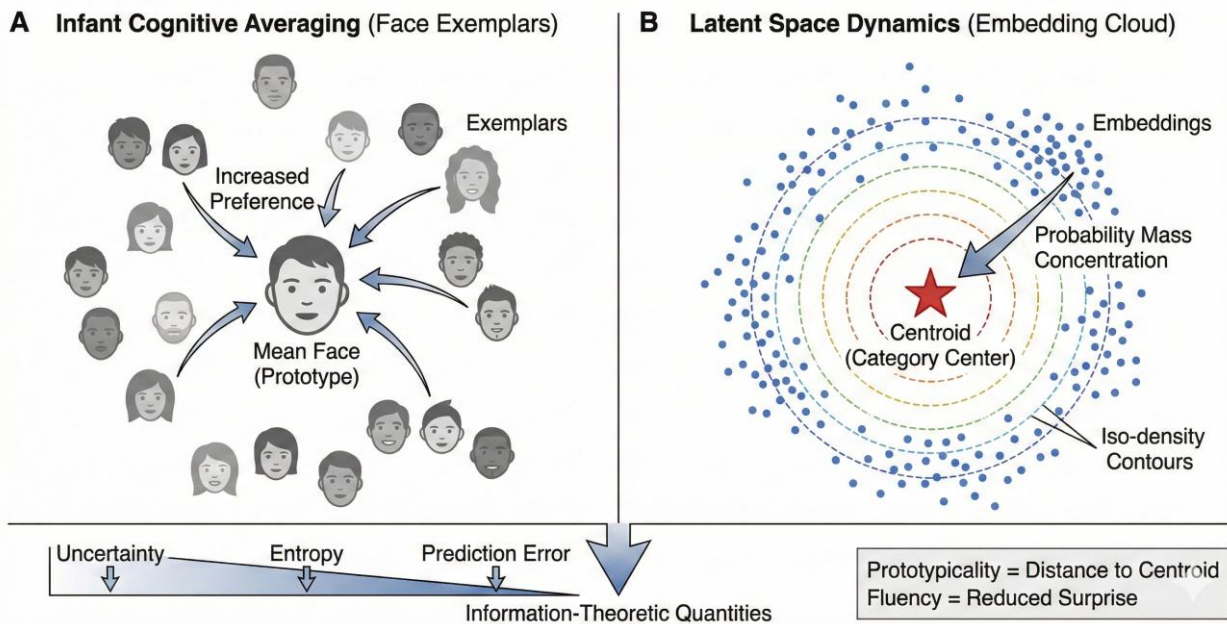
## 4. Isomorphism of Averaging in Biological and Artificial Systems

### 4.1 Psychological Prototypes and Vector Centroids

Prototype theory in cognitive psychology and centroid-based representations in machine learning both conceptualize categories as organized around central tendencies in a high-dimensional space. Classic work showed that category members differ in graded typicality, with “good” exemplars sharing many features with other members and few with out-group categories, a structure Rosch described in terms of family resemblance and internal cue validity (**Figure 1**) (Hampton, 1998; Rosch & Mervis, 1975). Subsequent experiments demonstrated that people are faster and more accurate when verifying typical category instances, and that averaged faces or gaits are judged more representative and aesthetically pleasing than atypical exemplars, indicating that human cognition privileges statistical centrality in perceptual and semantic domains (Minda & Smith, 2001; Rosch, 1975). In parallel, LLMs and related deep architectures construct embedding spaces in which semantically related words and concepts occupy dense clusters, and central

embeddings behave as distributional prototypes for these categories (Peters et al., 2018). Recent studies show that distances within these spaces systematically track human judgments of similarity and conceptual relatedness, and that multimodal models develop object-concept embeddings whose geometry partially aligns with neural activity in high-level visual cortices (Du et al., 2024; Xu et al., 2025). From a geometric standpoint, the vector centroid of a cluster is formally equivalent to a prototype: it represents the mean of all instances and defines a reference point against which typicality can be computed as distance or local density (Hampton, 1998; Minda & Smith, 2001). At the same time, both psychological and computational models must contend with multimodal and anisotropic distributions, where single centroids may fail to capture substructure or skewed variance within categories, prompting hybrid accounts that combine prototype and exemplar information. Despite these caveats, the recurrent emergence of central-tendency representations in behavioral data, neural coding, and model embeddings supports the claim that brains and machines converge on similar geometric solutions when organizing diverse exemplars.

**Figure 1. Prototype and Centroid Isomorphism**



**4.2 Processing Fluency and Perplexity Minimization**

Processing fluency and perplexity can be understood as parallel measures of how expected or easy a sequence is relative to an internal model, one implemented biologically and the other algorithmically. In cognitive psychology, fluency refers to the subjective and objective ease with which stimuli are encoded and integrated, with higher fluency reliably predicting positive evaluations and aesthetic liking across domains such as faces, patterns, and texts (Gao & Liu, 2024; Reber, Schwarz, & Winkielman, 2004). Prototype-theoretic manipulations of typicality, symmetry, and structural predictability increase fluency by aligning stimuli with well-entrenched category representations, thereby lowering effective processing demands and prediction error. In language modeling, perplexity is defined as the exponentiated cross-entropy between a model’s predicted distribution and the true distribution, and it provides a direct measure of average surprise per token: low perplexity indicates that the expectations of a model closely match observed sequences (Brown et al., 2020; Cover & Thomas, 2006). Under this lens, an output with low perplexity is one that adheres tightly to dominant statistical patterns in the training corpus, effectively occupying high-density regions of latent space and mirroring the fluency advantages of prototypical or highly predictable stimuli in human perception. This parallel suggests that beauty can be formalized as a state of reduced representational entropy or prediction error, in which incoming data are efficiently compressed by an internal generative model (Frascaroli et al., 2024; Reber, Schwarz, & Winkielman, 2004). At the same time, extremely low perplexity can yield banal or clichéd language that fails to provide sufficient novelty for sustained engagement, echoing psychological findings that excessive fluency diminishes interest. These limitations motivate a more nuanced view in which aesthetically satisfying outputs, whether human or machine-generated, lie in regions of state space that balance relatively low perplexity or high fluency with controlled deviation from prototypes.

**4.3 Attention Mechanisms as Aesthetic Filters**

Attention in both neural and computational systems can be interpreted as an aesthetic filter that selectively enhances coherent structure while suppressing noise. In predictive coding accounts, attention is modeled as the modulation of precision on prediction errors, effectively increasing the gain on signals that are expected to be reliable and informative for hierarchical inference (Feldman & Friston, 2010; Frascaroli et al., 2024). This precision weighting allows the brain to privilege features that conform to higher-level expectations, such as the symmetric configuration of a face or the regular metric of a musical phrase, while down-weighting idiosyncratic perturbations that would otherwise increase entropy without improving predictive accuracy. Transformer self-attention implements a mathematically analogous operation by computing relevance-weighted combinations of token representations via query–key–value interactions, enabling the model to focus on contextually diagnostic elements (e.g., syntactic heads or coreferent nouns) and to deemphasize tokens that contribute little to semantic coherence (Brown et al., 2020; Vaswani et al., 2017). Multi-head attention further distributes this filtering across multiple representational subspaces, capturing distinct structural relations in parallel and reinforcing central, high-probability patterns in the process. From an aesthetic standpoint, these attentional operations smooth away local irregularities and amplify higher-order regularities, producing representations and outputs that appear “polished” or internally consistent, in a manner reminiscent of how the visual system abstracts “face-ness” by ignoring blemishes and incidental detail. Commentary within predictive processing frameworks has emphasized that attention cannot be reduced to precision alone, but that precision modulation remains a core mechanism by which systems manage uncertainty and sculpt their internal experience of structure (Bowman et al., 2013; Feldman & Friston, 2010). This convergence suggests that empirical work linking attention dynamics, prediction-error weighting, and aesthetic judgments in humans could be complemented by analyses of attention patterns in large language models, clarifying when and how attentional filtering yields outputs experienced as coherent, fluent, or beautiful.

## 5. Empirical Bridges: Linking Human Judgments to Model Geometry

### 5.1 Prototypicality in Faces, Scenes, and Deep Vision Models

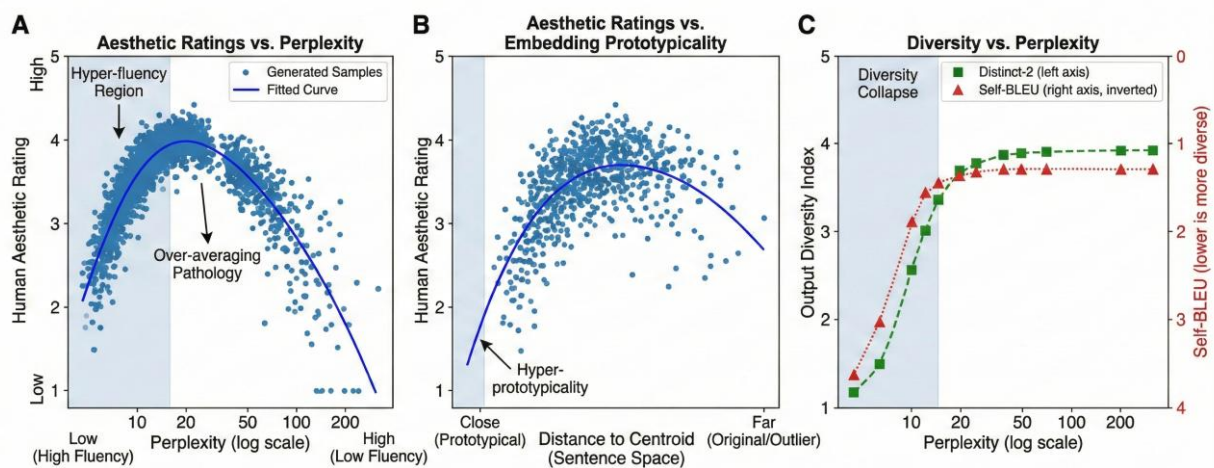
Research in computational aesthetics increasingly demonstrates that deep vision models capture prototypicality structures that align with human judgments of attractiveness and perceptual coherence. Convolutional and transformer-based networks trained on large-scale image datasets produce feature spaces in which images cluster according to high-level properties such as facial structure, scene layout, or stylistic composition, allowing researchers to quantify how close any exemplar lies to a learned category centroid (Redies et al., 2025; Soydaner & Wagemans, 2023; Talebi & Milanfar, 2018). Studies analyzing facial attractiveness have shown that embeddings derived from deep networks not only replicate known perceptual biases toward symmetry and averageness but also predict human beauty ratings with high accuracy when distance-to-centroid is used as a measure of prototypicality (Redies et al., 2025; Shi et al., 2024). Similar results have been demonstrated for natural scenes, where images located near dense regions of feature space—indicative of statistically typical structures—are judged as more aesthetically pleasing across cultures (Mikuni et al., 2024; Celona et al., 2022). These findings parallel infant studies showing early-emerging preferences for average faces and adult work documenting smoother and faster processing for prototypical exemplars. Together, they indicate that deep models and human observers share a representational geometry favoring central, low-entropy configurations. This convergence supports the broader claim that biological and artificial systems both rely on prototype-centered inference to stabilize perception in the face of noisy variability. Consequently, vision-model feature spaces can be treated as computational approximations of perceptual spaces, enabling a

unified analysis of typicality across neural and algorithmic systems.

### 5.2 Text Prototypicality and Aesthetic Qualities in Language

A growing body of research suggests that textual prototypicality, defined here as through embedding proximity, syntactic regularity, and predictability, correlates with human judgments of fluency, elegance, and stylistic appeal. Studies in computational linguistics show that sentences positioned near cluster centers in embedding space exhibit higher perceived coherence and readability, reflecting linguistic norms encoded in large corpora (Figure 2) (Peters et al., 2018; Manning et al., 2020). Parallel research in poetics and psycholinguistics identifies predictable rhythmic structures, moderate redundancy, and controlled variation as key contributors to aesthetic response, with overly irregular or overly predictable sequences rated as less aesthetically compelling (Jacobs, 2015; Obermeier et al., 2016). These results suggest that text with intermediate perplexity (neither too surprising nor trivially predictable) may map onto regions of embedding space that balance centrality with novelty. Computationally, this opens the possibility of assessing aesthetic quality by calculating the distance between generated sentences and corpus-derived centroids or by measuring entropy within syntactic and semantic patterns. Cross-linguistic work further reveals that fluency and stylistic preference vary across languages and traditions, indicating that prototypicality must be contextualized within culturally specific embedding geometries shaped by multilingual training data (Manning et al., 2020). Such analyses enable a comparative framework in which linguistic beauty can be studied as the interaction between structural expectations and probabilistic norms encoded in latent space. Consequently, text prototypicality provides a direct bridge between cognitive theories of aesthetic fluency and the statistical architectures underpinning the models.

**Figure 3: The Fluency-Originality Trade-off in Language Generation**



**Figure 3. Quantifying the trade-off between fluency and originality.** Data points represent thousands of continuations generated from fixed prompts under varying decoding regimes (temperature, top-p) across multilingual subsets. Panel A: Human aesthetic ratings peak at intermediate perplexity, with a drop in the ‘hyper-fluent’ regime. Panel B: Ratings are maximized at an optimal distance from the embedding centroid, not at the centroid itself. Panel C: Output diversity (Distinct-2, Self-BLEU) collapses dramatically at low perplexity, coinciding with the aesthetic drop in Panel A, illustrating the pathology of over-averaging.

### 5.3 Proposed Experiments and Testable Predictions

The theoretical convergence between biological and computational prototypes invites a set of empirical studies designed to jointly examine human aesthetic judgments and model-based geometric measures. One proposed paradigm uses morph continua that vary systematically between highly prototypical and atypical faces, collecting human attractiveness ratings while computing model-derived measures such as distance to the vision-model centroid, local entropy, or gradients in attention maps. Another approach involves presenting participants with textual stimuli generated at varying perplexity levels and measuring perceived elegance, coherence, or emotional impact while quantifying embedding distances and syntactic entropy (Jacobs, 2015; Obermeier et al., 2016). These studies would test predictions that human liking functions exhibit monotonic or inverted U-shaped relationships with prototypicality metrics, depending on domain and context: faces and simple scenes might show monotonic increases, while poetry or narrative may require moderate deviation from the centroid to achieve maximal engagement. Additional experiments could use cross-cultural sampling to determine whether centroid positions shift with linguistic or perceptual norms and whether model geometries can predict population-level variation in aesthetic preference (Mikuni et al., 2024). Methodologically, such research requires careful selection of pretrained models, transparent dimensionality-reduction pipelines, and curated datasets that adequately represent the diversity of human perceptual and linguistic experience. Through these designs, aesthetic evaluation becomes an empirical meeting point between human cognition and model geometry, allowing researchers to test whether beauty, as predicted, emerges from shared principles of entropy reduction and prototype-centered inference.

## 6. Generative Models, Mode Collapse, and Aesthetic Pathologies

### 6.1 The Midjourney Effect and Hyper-Prototypical Outputs

Observers and practitioners increasingly report that AI-generated faces and scenes often appear excessively smooth, symmetrical, and homogeneous, producing an impression of technical polish that many describe as aesthetically sterile. Qualitative studies with digital artists document recurring concerns about plasticity, hyperreal surface quality, and an absence of perceived “authorship,” especially in images produced by mainstream text-to-image systems that default to flawless skin, idealized proportions, and highly symmetrical compositions (Bianchi et al., 2025; Zhang, Wilson, & Amos, 2025). User-experience research on AI-based painting platforms further indicates that while non-expert users appreciate the ease with which visually impressive results can be achieved, both art majors and non-majors express ambivalence about images that feel too perfect or generic, highlighting tensions between efficiency and perceived authenticity (Wu et al., 2020; Xu et al., 2023). This hyper-prototypical aesthetic can be interpreted as the visible consequence of sampling from high-density regions near category centroids in latent space, where generative models concentrate probability mass in order to optimize reconstruction fidelity and likelihood. Because training objectives reward close adherence to the empirical distribution, generators are implicitly encouraged to reproduce the most frequent, “safe” combinations of features, thereby amplifying the prototype bias already documented in human perception of faces and scenes. The resulting images

exemplify extreme aesthetic fluency: they are easy to parse and often initially pleasing, yet they lack the small irregularities and idiosyncrasies that human observers frequently associate with expressive depth or individuality. Empirical work on attitudes toward AI-generated artworks and performances suggests that such hyper-fluent outputs can trigger subtle aversions or reduced liking, particularly when audiences are aware of the artificial origin and expect distinctively human traces of effort or imperfection (Ansani et al., 2025; Wu et al., 2020). The Midjourney effect can thus be framed as a limiting case of prototype optimization, where the pursuit of maximal fluency and prototypicality undermines the variability that sustains long-term aesthetic interest.

### 6.2 Mode Collapse, Embedding Drift, and Loss of Idiosyncrasy

Beyond the visible tendency toward hyper-prototypical samples, generative models are vulnerable to more formal pathologies that further erode diversity, including mode collapse and model or embedding drift. Mode collapse occurs when a model fails to represent all modes of a target distribution, instead overproducing a narrow subset of prototypes and ignoring rarer configurations; recent work has quantified this behavior with specialized metrics such as the Mode Collapse Entropy score, which captures reductions in class or feature diversity across generated samples (Duym, Oramas Mogrovejo, & Anwar, 2025). Complementary research introduces architectures such as DivGAN, which explicitly enforce output diversity by penalizing generators that concentrate on a small number of modes, highlighting how standard adversarial training objectives implicitly favor majority patterns at the expense of minority features (Allahyani et al., 2023). At a broader scale, studies on recursively trained models demonstrate that when synthetic outputs are repeatedly recycled as training data, distributions gradually drift toward oversimplified, homogenized representations that mischaracterize the complexity of the original data, a phenomenon labeled model collapse (Doshi & Hauser, 2024; Shumailov et al., 2024). These formal degeneracies have direct aesthetic consequences, manifesting as loss of individual style, repetition of similar motifs, and convergence of distinct character identities or visual signatures toward generic templates in both image and text generation. Behavioral work on human–AI creativity underscores this risk: although generative tools can increase individual idea volume and perceived creativity, controlled experiments show a simultaneous reduction in the collective diversity of novel content when many creators rely on the same systems and defaults (Doshi & Hauser, 2024; Zhang, Wilson, & Amos, 2025). Taken together, mode collapse and model collapse can be read as aesthetic analogues of over-averaging, where the benefits of prototypicality and fluency are pushed beyond a useful regime, resulting in outputs that are statistically efficient yet impoverished in variation and nuance.

### 6.3 Designing for Diversity: Resisting Over-Averaging

Addressing these aesthetic pathologies requires design strategies that explicitly counteract over-averaging and promote sustained diversity in generative outputs. At the algorithmic level, diversity-enforcing frameworks such as DivGAN modify loss functions and introduce auxiliary components to ensure that generators cover a wider range of modes with comparable fidelity, rather than concentrating probability mass exclusively on dominant patterns (Allahyani et al., 2023). Mode-balancing approaches extend this logic by reweighting training examples according to their contribution to dataset diversity, thereby encouraging models to

sample more frequently from underrepresented regions of the data manifold and to treat rare features as first-class citizens rather than noise (Berns, Colton, & Guckelsberger, 2023). Evaluation metrics like the Mode Collapse Entropy score and related diversity indices provide quantitative tools for monitoring these interventions, enabling systematic trade-offs between typicality and variation (Duym, Oramas Mogrovejo, & Anwar, 2025). At the interaction level, studies of human–AI co-creation emphasize the importance of interface affordances that steer users away from default prompts and sampling settings, instead supporting exploratory workflows that probe the tails of the distribution and cultivate idiosyncratic styles (Bianchi, Branchini, Uricchio, & Bongelli, 2025; Xu, Zhang, Li, Yoo, & Pan, 2023). These design choices resonate with aesthetic theories that foreground uniqueness, surprise, and deliberate departures from prototypes as central to artistic value, suggesting that well-designed systems should balance prototype alignment with structured opportunities for deviation. Ultimately, managing the tension between prototypicality and diversity is essential if generative models are to function not only as efficient pattern compressors but also as collaborators capable of producing work that remains engaging, distinctive, and responsive to human aesthetic expectations.

## 7. Conclusion: Beauty as an Entropic Signature of Intelligence

Beauty emerges in this account as an entropic signature of intelligent information processing, unifying developmental evidence, perceptual theory, and computational modeling under a single explanatory framework. Research on infant preferences demonstrates that human observers are attuned from the earliest months of life to patterns characterized by symmetry, averageness, and representational coherence, suggesting that prototype formation is a foundational strategy for reducing perceptual uncertainty and stabilizing early cognition (Quinn et al., 2001; Jeffery, Rhodes, & Bui, 2023). Predictive coding accounts extend this logic by modeling perception as a hierarchical process of minimizing prediction error or free energy, where aesthetically fluent stimuli achieve their appeal by enabling efficient compression of structured information and by satisfying expectations at multiple levels of inference (Frascaroli, Leder, Brattico, & Van de Cruys, 2024; Sarasso, Neppi-Modona, Sacco, & Ronga, 2020). Latent space analyses in deep neural networks reveal a structurally analogous phenomenon: large language models and vision models organize complex input distributions into cluster geometries centered on statistical prototypes, with high-density regions corresponding to the most predictable and easily encoded patterns encountered during training (Manning, Clark, Hewitt, Khandelwal, & Levy, 2020; Peters et al., 2018). Together, these lines of evidence support the view that beauty reflects the capacity of an intelligent system to navigate a high-dimensional space of possibilities by identifying, privileging, and compressing the central tendencies that afford the greatest reduction in entropy.

This synthesis not only clarifies why brains and artificial systems converge on prototype-centered structures but also highlights conceptual payoffs for multiple disciplines. For developmental psychology, it suggests that early-emerging aesthetic preferences are not epiphenomenal but reflect general-purpose mechanisms for structuring perceptual experience. For neuroaesthetics, it reframes beauty as a measurable outcome of predictive stabilization, rooted in dynamic interactions between expectation, precision weighting,

and representational efficiency. For AI research, it positions prototypicality, perplexity minimization, and attention-weighting as computational correlates of aesthetic fluency, offering a new vocabulary for evaluating model behavior and designing architectures that cultivate—not suppress—informative variation. These convergences invite empirical studies that examine human and model responses side by side, testing whether distances to centroids, entropy levels, or attention-weight patterns reliably predict aesthetic judgments across modalities, cultures, and levels of expertise.

At the philosophical level, treating beauty as an entropic phenomenon underscores its systematicity and tractability without negating the cultural, historical, and individual dimensions that enrich aesthetic life. While entropic signatures may explain why certain configurations reliably feel fluent or coherent, cultural norms and personal histories shape where individuals seek novelty, depth, and emotional resonance beyond statistical averages. Recognizing this duality encourages a pluralistic framework in which computational principles are necessary but not sufficient for understanding aesthetic meaning. Finally, the interdisciplinary scope of this argument highlights the need for collaborative inquiry across psychology, neuroscience, machine learning, philosophy, and the arts. By refining, challenging, and extending the proposed model, such collaboration can deepen our understanding of how intelligent systems—biological and artificial—construct worlds of meaning, and why beauty so often emerges at the point where structure, expectation, and information converge.

### Data Availability

Data available upon request.

### Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

### Funding Statement

NA

### Authors' Contributions

Conceptualization, D. Plate; Methodology, D. Plate; Validation, J. Hutson; Investigation, J. Hutson – Original Draft Preparation, J. Hutson; Writing – Review & Editing, J. Hutson.; Visualization, J. Hutson.

### References

1. Allahyani, M., Alsulami, R., Alwafi, T., Alafif, T., Ammar, H., Sabban, S., & Chen, X. (2023). DivGAN: A diversity enforcing generative adversarial network for mode collapse reduction. *Artificial Intelligence*, 317, 103863. <https://doi.org/10.1016/j.artint.2023.103863>
2. Ansani, A., Koehler, F., Giombini, L., Hämäläinen, M., Meng, C., Marini, M., & Saarikallio, S. (2025). AI performer bias: Listeners like music less when they think it was performed by an AI. *Empirical Studies of the Arts*. Advance online publication. <https://doi.org/10.1177/02762374241308807>
3. Berns, S., Colton, S., & Guckelsberger, C. (2023). Towards mode balancing of generative models via diversity weights. In *Proceedings of the 14th International Conference on Computational Creativity (ICCC 2023)*.

4. Bianchi, I., Branchini, E., Uricchio, T., & Bongelli, R. (2025). Creativity and aesthetic evaluation of AI-generated artworks: Bridging problems and methods from psychology to AI. *Frontiers in Psychology*, *16*, 1648480. <https://doi.org/10.3389/fpsyg.2025.1648480>
5. Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013). Attention is more than prediction precision. *Behavioral and Brain Sciences*, *36*(3), 206–208. <https://doi.org/10.1017/S0140525X12002324>
6. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.
7. Celona, L., Leonardi, M., Napoletano, P., & Rozza, A. (2022). Composition and style attributes guided image aesthetic assessment. *IEEE Transactions on Image Processing*, *31*, 5009–5024. <https://doi.org/10.1109/TIP.2022.3187590>
8. Chávez-Autor, J. C. (2025). Artificial creativity: From predictive AI to Generative System 3. *Frontiers in Artificial Intelligence*, *8*, 1654716. <https://doi.org/10.3389/frai.2025.1654716>
9. Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What does BERT look at? An analysis of BERT's attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP* (pp. 276–286). <https://doi.org/10.48550/arXiv.1906.04341>
10. Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley.
11. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186). <https://doi.org/10.48550/arXiv.1810.04805>
12. Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
13. Du, C., Fu, K., Wen, B., Sun, Y., Peng, J., Wei, W., ... He, H. (2025). Human-like object concept representations emerge naturally in multimodal large language models. *Nature Machine Intelligence*. Advance online publication. <https://doi.org/10.1038/s42256-025-01049-z>
14. Duym, J., Oramas Mogrovejo, J. A., & Anwar, A. (2025). Quantifying generative stability: Mode Collapse Entropy score for mode diversity evaluation. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops* (pp. 187–196).
15. Ethayarajh, K. (2019). How contextual are contextualized word representations? In *Proceedings of EMNLP-IJCNLP 2019* (pp. 55–65). <https://doi.org/10.48550/arXiv.1909.00512>
16. Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free energy. *Frontiers in Human Neuroscience*, *4*, 215. <https://doi.org/10.3389/fnhum.2010.00215>
17. Frascaroli, J., Leder, H., Brattico, E., & Van de Cruys, S. (2024). Aesthetics and predictive processing: Grounds and prospects of a fruitful encounter. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1895), 20220410. <https://doi.org/10.1098/rstb.2022.0410>
18. Gao, C., & Liu, C. (2024). How processing fluency, expectation, and epistemic goals influence aesthetic judgment: A perspective of multi-model integration. *Advances in Psychological Science*, *32*(11), 1872–1881. <https://doi.org/10.3724/SP.J.1042.2024.01872>
19. Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*(2–3), 137–165. [https://doi.org/10.1016/S0010-0277\(97\)00042-5](https://doi.org/10.1016/S0010-0277(97)00042-5)
20. Hwang, Y. J., & Cho, T. (2025). How familiarity and fluency shape perceived naturalness and aesthetic pleasure in digital spatial images. *Archives of Design Research*, *38*(3), 131–147. <https://doi.org/10.15187/adr.2025.08.38.3.131>
21. Jacobs, A. M. (2015). Neurocognitive poetics: A computational theory of aesthetic reading. *Frontiers in Human Neuroscience*, *9*, 186. <https://doi.org/10.3389/fnhum.2015.00186>
22. Jeffery, L., Rhodes, G., & Bui, Q. (2023). Developmental sensitivity to facial structure and attractiveness. *Developmental Science*, *26*(3), e13372. <https://doi.org/10.1111/desc.13372>
23. Jones, C. R., et al. (2024). Do multimodal large language models and humans reason similarly? *Computational Linguistics*, *50*(4), 1415–1447.
24. Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing* (3rd ed.). Prentice Hall.
25. Kobayashi, R., Sato, M., & Minemoto, K. (2024). Prototype formation in infant face perception. *Infancy*. Advance online publication. <https://doi.org/10.1111/infa.12539>
26. Lee, S., Tanaka, J., & Burris, A. (2024). Early configural processing and symmetry detection in infancy. *Journal of Experimental Child Psychology*, *236*, 105695. <https://doi.org/10.1016/j.jecp.2024.105695>
27. Li, B. Z., Yin, K., Han, X., & Ji, H. (2020). Dice: Deep contextualized clustering for entity types. In *Proceedings of ACL 2020* (pp. 7187–7198). <https://doi.org/10.18653/v1/2020.acl-main.642>
28. Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks. *Annual Review of Linguistics*, *6*(1), 195–212. <https://doi.org/10.1146/annurev-linguistics-011619-030326>
29. Mikuni, J., Spee, B. T. M., Forlani, G., et al. (2024). Cross-cultural comparison of beauty judgments in visual art using machine learning analysis of art attribute predictors among Japanese and German speakers. *Scientific Reports*, *14*, 15948. <https://doi.org/10.1038/s41598-024-65088-z>
30. Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799. <https://doi.org/10.1037/0278-7393.27.3.775>
31. Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., & Kotz, S. A. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(2), 362–373. <https://doi.org/10.3758/s13415-015-0405-x>
32. Parr, T., & Friston, K. J. (2019). Attention or salience? *Frontiers in Human Neuroscience*, *13*, 28. <https://doi.org/10.3389/fnhum.2019.00028>
33. Pedrotti, A., et al. (2025). How humans and LLMs organize conceptual knowledge. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (pp. 1415–1435). Association for Computational Linguistics.
34. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized

- word representations. In *Proceedings of NAACL-HLT 2018* (pp. 2227–2237). <https://doi.org/10.18653/v1/N18-1202>
35. Popescu, A., & Holman, A. C. (2024). Loop and enjoy: A scoping review of the research on the effects of processing fluency on aesthetic reactions to auditory stimuli. *Psychological Reports*. Advance online publication. <https://doi.org/10.1177/00332941241277474>
  36. Qi, J., Su, C., Hu, X., Chen, M., Sun, Y., Dong, Z., ... Luo, J. (2025). AMFMER: A multimodal full transformer for unifying aesthetic assessment tasks. *Signal Processing: Image Communication*, 138, 117320. <https://doi.org/10.1016/j.image.2025.117320>
  37. Quinn, P. C., Yahr, J., Kuhn, A., Slater, A., & Pascalis, O. (2001). Representation of facial prototypes in infants. *Psychological Science*, 12(5), 452–456. <https://doi.org/10.1111/1467-9280.00384>
  38. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*. <https://doi.org/10.48550/arXiv.1901.04584>
  39. Redies, C., Bartho, R., Koßmann, L., Spehar, B., Hübner, R., & Wagemans, J. (2025). A toolbox for calculating quantitative image properties in aesthetics research. *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-025-02632-3>
  40. Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382. [https://doi.org/10.1207/s15327957pspr0804\\_3](https://doi.org/10.1207/s15327957pspr0804_3)
  41. Rhodes, G., Geddes, K., Jeffery, L., Dziurawiec, S., & Clark, A. (2002). Are average and symmetric faces attractive to infants? *Cognitive Development*, 17(1), 49–63. [https://doi.org/10.1016/S0885-2014\(02\)00123-1](https://doi.org/10.1016/S0885-2014(02)00123-1)
  42. Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233. <https://doi.org/10.1037/0096-3445.104.3.192>
  43. Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
  44. Rubenstein, A. J., Kalakanis, L., & Langlois, J. H. (1999). Infant preferences for attractive faces: A cognitive explanation. *Developmental Psychology*, 35(3), 848–855. <https://doi.org/10.1037/0012-1649.35.3.848>
  45. Sarasso, P., Neppi-Modona, M., Sacco, K., & Ronga, I. (2020). The sense of beauty in the perception–action cycle: From predictive coding to aesthetic experiences. *Neuroscience & Biobehavioral Reviews*, 118, 723–739. <https://doi.org/10.1016/j.neubiorev.2020.09.001>
  46. Shi, T., Chen, C., Li, X., & Hao, A. (2024). Semantic and style-based multiple reference learning for artistic and general image aesthetic assessment. *Neurocomputing*, 582, 127434. <https://doi.org/10.1016/j.neucom.2024.127434>
  47. Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. <https://doi.org/10.1038/s41586-024-07566-y>
  48. Slater, A., Bremner, J. G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. (2000). Symmetry and attractiveness in infancy. *Symmetry*, 1(3), 345–359. <https://doi.org/10.3390/10030345>
  49. Soydaner, D., & Wagemans, J. (2023). Multi-task convolutional neural network for image aesthetic assessment. *IEEE Access*, 11, 4716–4729. <https://doi.org/10.1109/ACCESS.2023.3246702>
  50. Soydaner, D., & Wagemans, J. (2024). Unveiling the factors of aesthetic preferences with explainable AI. *British Journal of Psychology*. Advance online publication. <https://doi.org/10.1111/bjop.12707>
  51. Talebi, H., & Milanfar, P. (2018). NIMA: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8), 3998–4011. <https://doi.org/10.1109/TIP.2018.2838291>
  52. Tenney, I., Das, D., & Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. In *Proceedings of ACL 2019* (pp. 461–476). <https://doi.org/10.18653/v1/P19-1034>
  53. Van de Cruys, S., Frascaroli, J., & Friston, K. (2024). Order and change in art: Towards an active inference account of aesthetic experience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1895), 20220411. <https://doi.org/10.1098/rstb.2022.0411>
  54. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
  55. Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1998). Recognition of familiar faces by newborns. *Infant Behavior and Development*, 21(2), 233–243. [https://doi.org/10.1016/S0163-6383\(98\)90005-0](https://doi.org/10.1016/S0163-6383(98)90005-0)
  56. Wu, Y., Mou, Y., Li, Z., & Xu, K. (2020). Investigating American and Chinese subjects' explicit and implicit perceptions of AI-generated artistic work. *Computers in Human Behavior*, 104, 106186. <https://doi.org/10.1016/j.chb.2019.106186>
  57. Xu, J., Zhang, X., Li, H., Yoo, C., & Pan, Y. (2023). Is everyone an artist? A study on user experience of AI-based painting system. *Applied Sciences*, 13(11), 6496. <https://doi.org/10.3390/app13116496>
  58. Xu, Q., Peng, Y., Nastase, S. A., Chodorow, M., Wu, M., & Li, P. (2025). Large language models without grounding recover non-sensorimotor but not sensorimotor features of human concepts. *Nature Human Behaviour*, 9(9), 1871–1886. <https://doi.org/10.1038/s41562-025-02203-8>
  59. Yoo, J., Jasko, K., & Winkielman, P. (2024). Fluency, prediction and motivation: How processing dynamics, expectations and epistemic goals shape aesthetic judgments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1895), 20230326. <https://doi.org/10.1098/rstb.2023.0326>
  60. Zhang, L., Wilson, K., & Amos, C. (2025). The rise of AI art: A look through digital artists' eyes. *First Monday*, 30(4), Article 13809. <https://doi.org/10.5210/fm.v30i4.13809>