

## RESEARCH ARTICLE

# Using machine learning to predict judgments on Western visual art along content-representational and formal-perceptual attributes

Blanca T. M. Spee<sup>1,2,3\*</sup>, Helmut Leder<sup>1,3</sup>, Jan Mikuni<sup>1</sup>, Frank Scharnowski<sup>3</sup>, Matthew Pelowski<sup>1,3</sup>, David Steyrl<sup>3</sup>

**1** Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria, **2** Center of Expertise for Parkinson & Movement Disorders, Department of Neurology, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands, **3** Department of Cognition, Emotion and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

\* [blanca.spee@univie.ac.at](mailto:blanca.spee@univie.ac.at)



## OPEN ACCESS

**Citation:** Spee BTM, Leder H, Mikuni J, Scharnowski F, Pelowski M, Steyrl D (2024) Using machine learning to predict judgments on Western visual art along content-representational and formal-perceptual attributes. *PLoS ONE* 19(9): e0304285. <https://doi.org/10.1371/journal.pone.0304285>

**Editor:** Maja Vukadinovic, Novi Sad School of Business, SERBIA

**Received:** September 27, 2023

**Accepted:** May 9, 2024

**Published:** September 6, 2024

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0304285>

**Copyright:** © 2024 Spee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The datasets generated and/or analyzed during the current study, along with detailed results on how each

## Abstract

Art research has long aimed to unravel the complex associations between specific attributes, such as color, complexity, and emotional expressiveness, and art judgments, including beauty, creativity, and liking. However, the fundamental distinction between attributes as inherent characteristics or features of the artwork and judgments as subjective evaluations remains an exciting topic. This paper reviews the literature of the last half century, to identify key attributes, and employs machine learning, specifically Gradient Boosted Decision Trees (GBDT), to predict 13 art judgments along 17 attributes. Ratings from 78 art novice participants were collected for 54 Western artworks. Our GBDT models successfully predicted 13 judgments significantly. Notably, judged creativity and disturbing/irritating judgments showed the highest predictability, with the models explaining 31% and 32% of the variance, respectively. The attributes emotional expressiveness, valence, symbolism, as well as complexity emerged as consistent and significant contributors to the models' performance. Content-representational attributes played a more prominent role than formal-perceptual attributes. Moreover, we found in some cases non-linear relationships between attributes and judgments with sudden inclines or declines around medium levels of the rating scales. By uncovering these underlying patterns and dynamics in art judgment behavior, our research provides valuable insights to advance the understanding of aesthetic experiences considering visual art, inform cultural practices, and inspire future research in the field of art appreciation.

## Introduction

The way we assign value to a work of art is, to a great extent, a matter of judgments [1]. It begins with the question of whether a work qualifies as art, marking the first step in a

predictor influences the prediction outcomes, are available in the GitHub repository at <https://github.com/univiemops/art-rating-prediction>. Artwork images used from the Vienna Art Picture System (VAPS, see for full list of stimulus-set rated including VAPS-identification code S3 Table) are not publicly available due to copyright restrictions concerning the artists. However, these can be requested at the Faculty of Psychology, University of Vienna, Vienna, Austria ([dekanat.psychologie@univie.ac.at](mailto:dekanat.psychologie@univie.ac.at)), see also publication on VAPS at <https://doi.org/10.1037/aca0000460>.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

continuum of assessment [2]. Once a piece is recognized as art, individuals express their engagement through various judgments associated with visual art and aesthetic experiences [3, 4]. These judgments encompass concepts such as beauty, liking, interest, and thought-provoking, representing a means to articulate subjective evaluations [1, 3, 5]; evaluations that are deeply influenced by the individual history of encounters with art within social and cultural contexts [6–9].

Attributes, on the other hand, play a pivotal role in imbuing these judgments with subjective characteristics. Specifically, someone's preference for an artwork or the recognition of a masterpiece by certain cultural groups, laymen, or experts go beyond mere artist recognition or historical references [10, 11]. Attributes are specific qualities individuals employ to substantiate their evaluations, forming the foundation for articulating and validating subjective opinions. For instance, one might attribute their appreciation of a work of art to its colorfulness, find beauty in emotional expressiveness, associate imaginativeness with creativity [12], or experience fascination due to unique stylistic elements [3–5, 13, 14 see for further examples, 15]. These attributes serve as criteria for evaluating and categorizing artworks, including factors such as complexity, color usage, emotional expressiveness, stylistic elements, and other observable features. By considering these attributes, judgments acquire personalization and nuance, adding subjective qualities to the overall evaluation of artworks.

Our journey into understanding art judgments and their connection to art attributes began with a focused study on creativity judgments. In this prior work [12], we employed interpretable machine learning using Random Forest ensemble models [16] to probe the complex associations between 17 subjective art attributes and creativity judgments across a diverse range of artworks. While attributes like symbolism, emotional expressiveness, and imaginativeness played significant roles in predicting creativity judgments, other factors like abstractness, valence, and complexity also made an impact, albeit to a lesser degree. This investigation presented the first attribute-integrating quantitative model of factors contributing to creativity judgments in visual art among novice raters. It marked a significant stride forward in building the groundwork for introducing machine learning as an innovative approach, by focusing specifically on subjective attributes ratings predicting judgments.

Building upon this foundation, our current study extends the scope beyond creativity judgments to encompass a broader spectrum of 13 different art judgments. We used the same set of 17 art attributes [12]. Our approach remains rooted in machine learning, with the aim of unveiling intricate patterns, regularities, and potential inconsistencies within art evaluation. However, we updated our model, specifically applying Gradient Boosted Decision Trees [GBDT, 17]. By leveraging data-driven analyses, we seek to enhance our understanding of art judgment behavior and the appreciation of art, shedding light on the multifaceted interplay between judgments and attributes.

## The societal significance of understanding art judgments and attributes

Understanding art judgments and their connection to attributes is of profound importance in both contemporary society and the realm of art and aesthetic research [1, 18]. When people talk about art and their preferences, judgments, values, and explanatory attributes are inherently tied to the attempt to communicate, might it be recognizing some information about people themselves—such as social reputation or sparking intellectual discourse [7, 9, 10]—or about societal issues, fostering critique, provocation, and progress within societies [8, 19]. The connection between attributes and art judgment has also been used to test universal assumptions of art historical claims, which largely rely on the coherence of these connections [4]. In Specker and colleagues, for example, they explored the assumption of universality in the

perception of aesthetic effects associated with visual objective elements such as colors and lines. They investigated whether people universally associate certain qualities with key visual forms, a notion fundamental to art discourse. Their findings challenged this assumption, revealing significantly lower agreement than expected on several aesthetic effect dimensions. While participants agreed on the effects of warm-cold, heavy-light, and happy-sad, there was hardly any consensus on 11 other dimensions. Intriguingly, consensus was higher regarding the whole artwork, rather than for single elements within the artwork, potentially addressing the importance of more holistic content-representational attributes rather than objective, or formal perceptual features [18]. Notably, our prior research into creativity judgments revealed that certain attributes, particularly content-representational features, significantly predict creativity judgments, offering a deeper understanding of this intricate relationship [12]. This highlights the complexity of art judgments and the need to explore the role of attributes further.

Artistic creations possess the remarkable capacity to challenge established norms, ignite intellectual discourse, and inspire change. Consequently, understanding art judgments allows us to gain insights into how individuals perceive and respond to artistic interventions that seek to question, disrupt, or redefine societal narratives (Becker, 1982). Through this lens, we can comprehensively analyze the diverse range of responses and interpretations evoked by artworks, thereby contributing to a deeper comprehension of the cultural and social impact of art [20].

Furthermore, the relevance of studying art judgments is magnified in today's digital worlds [21]. The proliferation of online platforms and social media has democratized art consumption, making artworks more accessible than ever before. Platforms like Pinterest and Instagram have fundamentally transformed the way art is discovered, shared, and evaluated in digital spaces [22]. This heightened accessibility underscores the necessity of a comprehensive understanding of art judgments. By delving into the relationship between art judgments and attributes, we can gain profound insights into the mechanisms that underpin the appreciation and dissemination of art in the digital realm as well. Our research, building on our prior work [12], strives to uncover how attributes, particularly those linked to subjective ratings, influence the judgments placed on artworks, extending our knowledge in the field of art and creativity research.

## The study of art judgments and machine learning: A historical perspective and future prospects

The exploration of art judgments and their association with art attributes has a rich history that stretches back nearly a century, although philosophical discussions on the topic likely trace back to the first expressions of human creativity [23–25]. Among the early pioneers in this field, Daniel Berlyne stands out for introducing multidimensional scales and focusing on hedonic judgments and attributes related to complexity and arousal [5, 26–29]. However, Berlyne's approach, which included the measurement of "cortical arousal" (albeit in the form of ratings), faced criticism for its reductionistic view of aesthetic experience and often encountered challenges in replication [30–33]. These limitations led to further research in the field. Despite ongoing research on art judgments and art appreciation models [3], there remains a significant gap in the analysis of the predictability of factor clusters and the intricate relationship between art judgments and attributes [2, 34]. Current studies often focus on a limited number of commonly used ratings, such as liking, familiarity, understanding, and beauty, without thoroughly investigating the validity of these judgments or the attributes that underlie them [35, 36]. Additionally, there is often a conflation of art judgments and attributes, with attributes frequently not explicitly tested for their role in shaping art judgments [18].

Existing research on art judgments and their connection to art attributes faces two primary limitations. First, the field lacks a comprehensive and systematic analysis of the variables and dimensions involved in art judgments and their interrelationships. While there have been extensive studies on individual judgments and attributes, such as beauty, complexity, liking, and others, there is a pressing need to explore a broader spectrum of factors and how they interact in the evaluation of artworks. Understanding the complex dynamics of art judgments and the influence of attributes can provide valuable insights into the multifaceted nature of aesthetic experiences [37].

Second, the majority of studies in this area have traditionally relied on linear models for analysis [38]. These models assume linear relationships between judgments and attributes, which may not adequately capture the intricate and nonlinear nature of art evaluation as a human behavior. To address these limitations, our research embraces machine learning techniques and non-linear models to explore the complex and nuanced relationships between judgments and attributes [39]. Machine learning offers flexible and robust modeling approaches capable of handling high-dimensional data and capturing nonlinear patterns, allowing us to uncover hidden relationships and identify novel factors that influence art judgments [16, 40–44].

The integration of computational models into the realm of aesthetics began in the 1960s when Max Bense and G. D. Birkhoff laid the foundations for computational aesthetics and the quantification of aesthetics using mathematical formulas [45, 46]. Bense aimed to derive scalar measurements of artwork aesthetics, merging information theories and generative language theories. Intriguingly, Bense's approach was influenced by Berlyne's studies on attribute interrelations for preference and aesthetics [5, 28, 47].

While machine learning has gained prominence in various fields exploring human behavior, psychology, and decision processes [16, 40–44], its application in art research has been relatively limited. Studies by Li and colleagues [48, 49] have used machine learning to investigate the predictability of subjective aesthetic preferences based on visually objective properties, such as formal-perceptual features, of artworks. Their findings suggest that objective artwork features can inform predictions of preference ratings, hinting at a degree of universality in the evaluation of visual art objects, at least on a perceptual level (although note the discussions presented by Specker et al., 2020).

Research conducted by Iigaya and colleagues [50–52] has delved into the neural mechanisms underpinning preference computations in the brain. They have revealed that preference computations involve a graduated hierarchical representation of attribute structures in the visual system, with attribute information influencing subjective evaluation in higher cognitive brain areas. This aligns with the argument that information processing in art judgments may involve shared brain networks [35], hereby supporting Berlyne's starting viewpoint on art psychobiology [26] and subsequent work in neuroaesthetics [53–55]. Iigaya's and colleagues research indicates interpersonally shared attribute-specific processing, aligning with the notion that judgments rely on factors patterned along socially and culturally learned attributes [6, 56, 57].

In summary, all these findings encourage the use of machine learning to model art judgments using attributes as features derived from artworks. Leveraging machine learning algorithms allows to construct data-driven models capable of capturing the nuanced dynamics of art evaluation. Notably, our prior research on that matter [12] shows that using non-linear machine learning models is most informative also for subjective ratings.

## Methods literature review: Procedure and identification of factors

A systematic literature review was conducted, in line with the guidelines for systematic reviews, to identify potential attributes and art judgments [58]. Based on the review, we developed two sets of scales (1) art judgments and (2) art attributes. These scales were used in the empirical study for the ratings of the artworks. Additionally, the summary of publications (S1 Appendix) might itself serve as a comprehensive summary of the current state of research in the field and inspire future studies, such as determining which measures should be used for specific purposes and identifying aspects that have yet to be investigated.

## Identification of art judgments and art attributes

Results of the systematic review adhered to established guidelines for systematic reviews [58], and the process is detailed in S1 Fig. We conducted a thorough search using various electronic databases and search engines, including Google Scholar, PEDro, Scopus, and PubMed, in addition to academic libraries accessible to University of Vienna staff. Initially, we employed keywords such as “art,” “art judgment/judgement,” “art evaluation,” “art scales,” “art assessment,” “artistic,” and “aesthetic.” In a subsequent search, we included sub-keywords such as “attributes,” “features,” “dimensions,” “criteria,” “factor,” “factor-analyses,” “machine learning,” and “statistic learning approach.” For each keyword, we examined the first five pages of results, yielding 76 full-text publications. After applying relevance criteria related to art judgments and attributes, 46 publications were selected for further analysis. By scrutinizing the references within these publications, we identified an additional 14 relevant publications. In total, our review encompassed 33 art assessment publications and three studies at the intersection of art and machine learning. The development of two sets of scales, one for art judgments and one for art attributes, drew from a comprehensive pool of 36 publications. For an overview of the approaches and reviews, including relevant papers not explicitly discussed in this study, please refer to S1 Appendix under section 3.0.

To maintain a specific focus on the evaluation of artworks and their attributes while avoiding the inclusion of scale groupings found in previous studies, we excluded ratings that considered participants' elicited emotions [e.g., 59]. However, it is important to note that a clear demarcation between emotional responses and judgments is not always possible. Therefore, judgments such as “aesthetically moving” or attributes like “emotional expressiveness” were considered in our analysis. Nevertheless, we explicitly instructed participants to concentrate on the evaluation of the artwork's representations rather than their personal emotional experiences. Our study acknowledges that our selection of publications is not exhaustive. We specifically focused on studies that discussed scales related to judgments or attributes, recognizing that other research explores variables such as liking or beauty with different research objectives. In essence, our review aimed to encompass studies that delved into the intricacies of these scales rather than using ratings solely as an outcome measure for alternative research foci.

## Art judgements scale

Our review of art judgments resulted in 13 items, which we sorted along six aspects of art-judgements (see Table 1, I-VI): aesthetic, qualitative, epistemic, adverse, semantic, and preference aspects. In S1 Appendix, we provide more detail on each aspect and the items representing the 13 judgments included within each aspect.

**Table 1. Scale-set of art-judgements (targets in machine learning analysis) used in the empirical study, English version (see for German version S1 Table).**

	Instruction	Please provide your rating on the artwork:
Aspects	Scale items of art-judgments	Questions
I. Aesthetic aspects	1. Aesthetically moving	How aesthetically moving do you find the artwork?
	2. Beauty	How beautiful do you find the artwork?
II. Qualitative aspects	3. Good work of art	Do you find the painting a good artwork?
	4. Creativity	How creative do you find the artwork?
	5. Technical skill	How highly do you rate the artist's technical skill?
III. Epistemic aspects	6. Fascinating (intellectual)	Do you find the artwork fascinating/intellectually stimulating?
	7. Interesting	How interesting do you find the artwork?
	8. Thought-provoking	How much does the artwork make you thoughtful?
IV. Adverse aspects	9. Boring	How boring do you find the artwork?
	10. Disturbing, irritating	How disturbing or irritating do you find the artwork?
V. Semantic aspects	11. Familiarity	How familiar does the artwork seem to you?
	12. Understanding	How much do you understand the artwork?
VI. Preference aspects	13. Liking	How much do you personally like the artwork?

<https://doi.org/10.1371/journal.pone.0304285.t001>

## Art-attributes scale

For the selection of art attributes, we primarily followed the framework provided by ‘The Assessment of Art Attributes’ (AAA) by Chatterjee and colleagues [18]. Although the AAA specifically aims to advance studies in the field of neuropsychology of art, it serves as a versatile instrument for quantifying attributes of artworks and assessing differences in any context. Considering the comprehensive range of art attributes covered by the AAA, we decided to focus on this list and adapt it accordingly. The AAA includes 12 art attributes, divided into two perspectives: formal-perceptual attributes and content-representational attributes. This division we also kept in our study. The formal-perceptual attributes of the AAA consist of balance, color saturation, color temperature, depth, complexity, and brushstroke. On the other hand, the content-representational attributes encompass abstractness, animacy, emotion, realism, objective accuracy, and symbolism.

In our attribute scale set, we introduced three additional scales: (e) color variety, (g) color world, and (i) utilization of drawing area. The inclusion of these scales was motivated by the frequency with which color was addressed in the literature [13, 48–51, 60–64]. Various studies examined aspects such as color saturation, color temperature, color combination, color contrast, and specific color values. To capture the comprehensive nature of color aspects, we added (e) color variety (ranging from few to many colors) and (g) color world (ranging from dark to light colors) to our scales. Furthermore, the AAA does not include questions regarding the spatial structure or composition of artworks. To address this aspect, we introduced the scale (i) utilization of drawing area, which ranges from little utilization to full utilization of the painting area [65].

Regarding the content-representational attributes, we revised the animacy item. The decision to revise the animacy scale was based on its limited usage in other literature compared to the more frequently used attribute of (n) liveliness with poles still to dynamic.

Based on our literature review, we also introduced two additional items: (p) valence (ranging from positive to negative). This attribute has been recognized as a major dimension in various art models and as a specific factor influencing preferences for art [1, 3, 66–68]. We also

added (q) focus. The inclusion of the focus scale (ranging from much contextual information to focused content) aimed to account for potential cultural differences observed in previous research, where Asian individuals may rely more on contextual information and Western individuals focus on specific content when making judgments [69–71]. Although our participant sample is European, in outlook of future studies, we decided to include this item.

One major change was that we utilized the scale from the AAA, employing semantic differentials (opposite word pairs) for all items to represent each attribute's poles as it has been applied in past studies [5, 27, 72]. The list of art attributes can be found in Table 2.

## Methods: Rating assessment utilizing machine learning

Based on the review and established list of art judgments and attributes, we developed a study design to investigate our research question using data-driven machine learning.

### Participants

The ratings were collected from 78 psychology students at the University of Vienna (55 women,  $M_{age} = 24.23$ ,  $SD = 3.45$ , ranging from 19 to 35, data collection took place 07.11.2019 to 13.11.2020). Each participant rated the full set of artworks (see below). All participants signed informed consent, were informed about the purpose of the study, and participated for course credit. We also asked them for their art education, their self-reported art-making experience as a hobby or professional artist, and their self-identified knowledge and interest in art [based on previously published methods by, e.g., 73]. Ethical approval was not required for the studies involving humans because ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements.

**Table 2. Scales of attributes (predictors in machine learning analysis) used in the empirical study, English version (see for German version S2 Table).**

Attributes	Instruction Scale items of attribute dimensions	<i>Please evaluate the artwork based on the different art attributes:</i>	
		Negative pole (minimum)	Positive pole (maximum)
i. Formal-perceptual attributes	a. Visual harmony (balanced)	visual harmony, proportional	peculiar, strange shapes
	b. Depth	two-dimensional	three-dimensional
	c. Complexity	simple	Complex
	d. Color saturation	soft, pastel	intense, strong
	e. Color variety	few colors	many colors
	f. Color temperature	warm colors	cold colors
	g. Color world	dark color world	light color world
	h. Brushwork	fine brushwork	rough brushwork
	i. Utilization of drawing area	little utilization of the painting area	full utilization of the painting area
ii. Content-representational attributes	j. Abstractness	representative	Abstract
	k. Imaginativeness	realistic content/topic	imaginary, unreal, fantastic
	l. Symbolism (ambiguity)	distinct/clear (unambiguous interpretation)	symbolic (more space for interpretation, ambiguous)
	m. Accurate object representation	photorealistic	painterly
	n. Liveliness/animation	dynamic	Still
	o. Emotional expressiveness	emotionless	emotionally loaded
	p. Valence	negative valence	positive valence
	q. Focus	much context/environment in the image	focused content

<https://doi.org/10.1371/journal.pone.0304285.t002>

All participants provided their written informed consent to participate in this study. The study corresponds with the ethical standards of the Declaration of Helsinki and the ethical regulations at the University of Vienna. No retrospective data was included in the study.

The number of participants followed best practice in artwork rating studies [36, sample of at least 10 judges provide stable scoring with good reliability, e.g., 74–76] and best practice for high feature predictability using machine learning. Because precise sample size justifications for complex, high-dimensional, multivariable models from the machine learning field have not yet been standardized, our power analysis follows the most recent suggestions [77, 78]: (1) A minimum of 50 samples are required to start any meaningful machine learning based data analysis [79]. (2) 10 to 20 samples per degree of freedom (predictor) is reasonable, which would lead to a total number of 170 to 340 ratings (samples) required in the present case [80, 81]. (3) A power analysis using the two-tailed student's *t*-test, with an alpha of 0.05 and a power of 0.8 suggests that 620 ratings (samples) would be required to detect small effects of Cohen's *d* 0.2 [82]. Generally, as in other statistical data analysis methods, more samples allow finding smaller effects. Based on these considerations and available resources, we collected a total number of 4206 ratings (samples, 78 participants evaluating 54 images, 6 picture ratings were missing due to recording issues).

## Materials

We selected 54 images of Western paintings from the Vienna Art Picture System [VAPS, 83] as stimuli for our study (see for full list of artworks S3 Table). The VAPS provided a suitable platform for stimulus selection, offering a diverse range of 999 fine art paintings with various motives, styles, and time epochs, along with pre-existing rating scores (liking, valence, arousal, complexity, familiarity) for each stimulus.

To ensure variation in the artworks, we included three subsets of depicted motifs (portrait, landscape, still life) and three style categories (representative, impressionistic, abstract art). The number of images was balanced across the depicted motifs and styles. We conducted a Wilk-Shapiro test (see S4 Table) to assess the normal distribution of the provided ratings for the selected stimuli. Overall, we observed normal distributions in most categories, except for familiarity ratings, which showed significant differences across all categories due to the inclusion of famous and lesser-known artworks and considering pre-ratings were made by art novices in the VAPS. Additionally, more representative art was rated significantly more positively in valence and arousal, possibly due to the higher familiarity with realistic artworks.

Despite these limitations, we deemed them acceptable given the normal distributions observed in other categories and the unavailability of other stimulus databases with pre-existing ratings.

## Procedure

The testing took place in a laboratory room equipped with four workstations for simultaneous participant testing. Participants provided written informed consent and were instructed to rate visual artworks using a series of scales. They were given the freedom to take as much time as needed, with an average completion time of 90 to 120 minutes. The artworks were displayed on a 19" monitor (Iiyama ProLite B1906S, 1280x1024, 60Hz), with a maximum dimension of 500 pixels and participants seated approximately 50 cm away from the screen. The artworks and ratings were presented using the program SoSci-Survey on a University of Vienna server [84]. Each artwork was shown individually centered on the screen with a white background, while the scales were displayed below. The order of artwork presentation was randomized



between participants. The scales and items were randomized between each trial (within participant) to avoid rating sequence effects.

Two types of scales were used: (1) items of art judgments rated on a 100-point Likert-type scale from “not at all” to “very much” using a slider, and (2) items of art attributes presented as opposite pairs (semantic differentials) on a 100-point Likert scale. The slider had a middle position at the 50-point mark. The full list of scales in English can be found in Tables 1 and 2 (for German versions see S1 and S2 Tables).

### Data analysis: Multivariable regression

The multivariable statistical regression analyses were performed using Python v3.10.5 and the machine-learning framework scikit-learn library v1.1.1 [79]. Each analysis consisted of three parts: (1) prediction model training, (2) generalizability testing, and (3) model analysis.

1. Gradient Boosted Decision Tree (GBTD) models were used for the regression task because they are computationally efficient and highly accurate [17]. In addition, GBTD models can model non-linear associations and interactions between predictors and target variables. They are also robust to multicollinearity and outliers in the data. The predictors used in the models included a total of 17 attributes. The 13 art judgments were the prediction targets. Our model selection was primarily driven by the findings of recent research, such as the work by Grinsztajn and colleagues [85], which suggests that tree-based models often outperform deep learning on typical tabular data like ours.
2. To assess the performance of the models on unknown data (i.e., how well the predictions generalize), a nested cross-validation (CV) procedure was employed [86]. CV implements repeated splits of the data into training and testing set. A group-controlled shuffle-split scheme was used in the main (outer) CV loop to control for participant-related clusters (20% of participants in the testing set, 80% of participants in the training set, 50 repetitions). In each repetition, the training set was used for data scaling (standardization) and model complexity tuning. A nested (inner) CV procedure (20% of participants in the testing set, 80% of participants in the training set, 14 repetitions) was used to tune the complexity parameters of the model. To find the best performing parameters, a sequential Bayesian optimization procedure in combination with a shuffle-split scheme was used (400 repetitions, 200 initial points; BayesSearchCV, scikit-optimize, v0.9, learning\_rate 1e-3 to 1e0, n\_estimators 1 to 3000, num\_leaves 2 to 1000, min\_child\_samples 1 to 1000, colsample\_by\_tree 1e-3 to 1e0, reg\_alpha 1e-3 to 1e3, extra\_trees True/False). These parameters were used in to train regressor models in the main CV loop, all other parameters were left to default. The models were subsequently tested on the respective testing set of the main CV loop. The testing set was explicitly not used in the inner CV loop. Regression performance was evaluated using two metrics: (i) the prediction coefficient of determination ( $R^2$ ), and (ii) the mean absolute error (MAE) [87].  $R^2$  can have values between minus infinity and 1, with a value of 0 indicating a performance as good as using the average target value as a predictor (the trivial predictor), and a value of 1 indicating no error at all. It is worth noting that the prediction  $R^2$  will be smaller than  $R^2$  values of conventional statistical models because the prediction  $R^2$  measures prediction performance for unknown data, rather than post-hoc model fit [87]. The MAE reflects the average error made at each prediction and is not normalized; thus, the error is measured with the scale of the dependent variable, the prediction targets.
3. To assess the importance of individual predictors for the model's performance, we used SHAP (SHapley Additive exPlanations) [88, 89]. SHAP is a method from interpretable

machine learning that is based on Shapley values, a method from cooperative game theory. It measures the contributions of each predictor to the model's performance. Pooling those contributions over many predictions allows for a comprehensive analysis of the importance of individual predictors for the regression task [88, 89].

### Data analysis: Statistical tests

Statistical significance of the prediction  $R^2$  and MAE metrics as well as of the predictor's importance's was assessed using a modified  $t$ -test that takes the sample dependence due to CV into account [90, 91].  $T$ -test results with Bonferroni corrections are presented together with the ones without correction.

We note that detailed results on how each predictor influences the prediction results shown below are available from this link: <https://github.com/univiemops/art-rating-prediction>.

### Results

We applied machine learning to predict 13 art judgments based on ratings of visual art attributes, in order to identify which attributes (if any) contribute to the predictions. We computed the multivariate models and evaluated their performance. The GBDT models provided a good fit within the range of conventional cutoff values (as can be seen in Table 3). In total, all 13 judgments could be significantly predicted by the attributes. The results showed average absolute errors of 14.85 (SD = 1.20) for the lowest *disturbing/irritating* and 21.48 (SD = 0.83) for the highest, *familiarity*, within a range of 1 to 101. The prediction coefficient of determination  $R^2$  ranged between 0.02 for *familiarity*, 0.14 for *good work of art* as the lowest  $R^2$  value, and 0.31 for *creativity* and 0.32 for *disturbing/irritating* as the highest  $R^2$  value. This indicates that on average, the model explains 31% of the variance in the *creativity* ratings and 32% of the variance in *disturbing/irritating* ratings.

**Table 3. GBDT regressor prediction results of all art judgment ratings.** 50 repetitions from cross-validation.

Scales of art-judgments	Mean Absolute Error (MAE) $\pm$ standard deviation	$p$ value of obtained MAE bigger or equal to MAE from shuffled labels data	Coefficient of determination ( $R^2$ ) $\pm$ standard deviation	$p$ value of obtained $R^2$ smaller or equal to $R^2$ from shuffled labels data
Aesthetically moving	19.79 $\pm$ 1.33	$p = 0.009$	0.15 $\pm$ 0.04	$p < 0.001$
Beauty	20.09 $\pm$ 1.10	$p < 0.001$	0.20 $\pm$ 0.06	$p < 0.001$
Good work of art	19.02 $\pm$ 1.21	$p < 0.001$	0.14 $\pm$ 0.06	$p < 0.001$
Creativity	17.34 $\pm$ 1.20	$p < 0.001$	0.31 $\pm$ 0.04	$p < 0.001$
Technical skill	16.66 $\pm$ 1.48	$p < 0.001$	0.16 $\pm$ 0.09	$p < 0.001$
Fascinating (intellectual)	18.62 $\pm$ 1.24	$p < 0.001$	0.22 $\pm$ 0.04	$p < 0.001$
Interesting	19.55 $\pm$ 1.38	$p < 0.001$	0.25 $\pm$ 0.05	$p < 0.001$
Thought provoking	18.18 $\pm$ 1.14	$p < 0.001$	0.22 $\pm$ 0.05	$p < 0.001$
Boring	20.77 $\pm$ 1.30	$p < 0.001$	0.22 $\pm$ 0.05	$p < 0.001$
Disturbing, irritating	14.85 $\pm$ 0.83	$p < 0.001$	0.32 $\pm$ 0.06	$p < 0.001$
Familiarity	21.48 $\pm$ 1.18	$p = 0.002$	0.02 $\pm$ 0.04	$p = 0.001$
Understanding	19.23 $\pm$ 1.40	$p < 0.001$	0.24 $\pm$ 0.08	$p < 0.001$
Liking	21.15 $\pm$ 1.26	$p < 0.001$	0.18 $\pm$ 0.05	$p < 0.001$

<https://doi.org/10.1371/journal.pone.0304285.t003>

The results of our machine learning predictions for aesthetic aspects showed prediction-based coefficients  $R^2$  in the range of 0.15 for *aesthetically moving* to 0.20 for *beauty*. Qualitative aspects vary strongly, with 14% explained variance for *good work of art*, *creativity* showing the highest explained variance at 31% and with 16% variance for *technical skill*. Similarly, general aspects showed low explanation strength for *familiarity* ( $R^2 = 0.02$ ) but relatively high explanation strength for *understanding* (24% variance). Epistemic aspects, such as *fascinating*, *interesting*, and *thought-provoking*, provide a relatively similar explanation of variance (>22%). Adverse aspects also show a higher variance explanation, with 22% for *boring* and 32% for *disturbing/irritating*. Last, *liking* had an average variance explanation of 18%.

Furthermore, we analyzed the contributions of individual factors to the model's performance, i.e., which factors are important to predict the art-judgments ratings. In Figs 1 and 2 we show two examples of judgments, aesthetically moving and liking, representing two examples of lower degree of attribute explanation (lower  $R^2$ ); Figs 3 and 4 show the results of the judgments creativity and disturbing/irritating as examples for high explained variance attributes.

Our analysis revealed several key findings regarding the importance of different attributes in art judgments (see Figs 1 to 2 as well as <https://github.com/univiemops/art-rating-prediction> for full list of plots). See for overview of attribute's average rank Fig 5.

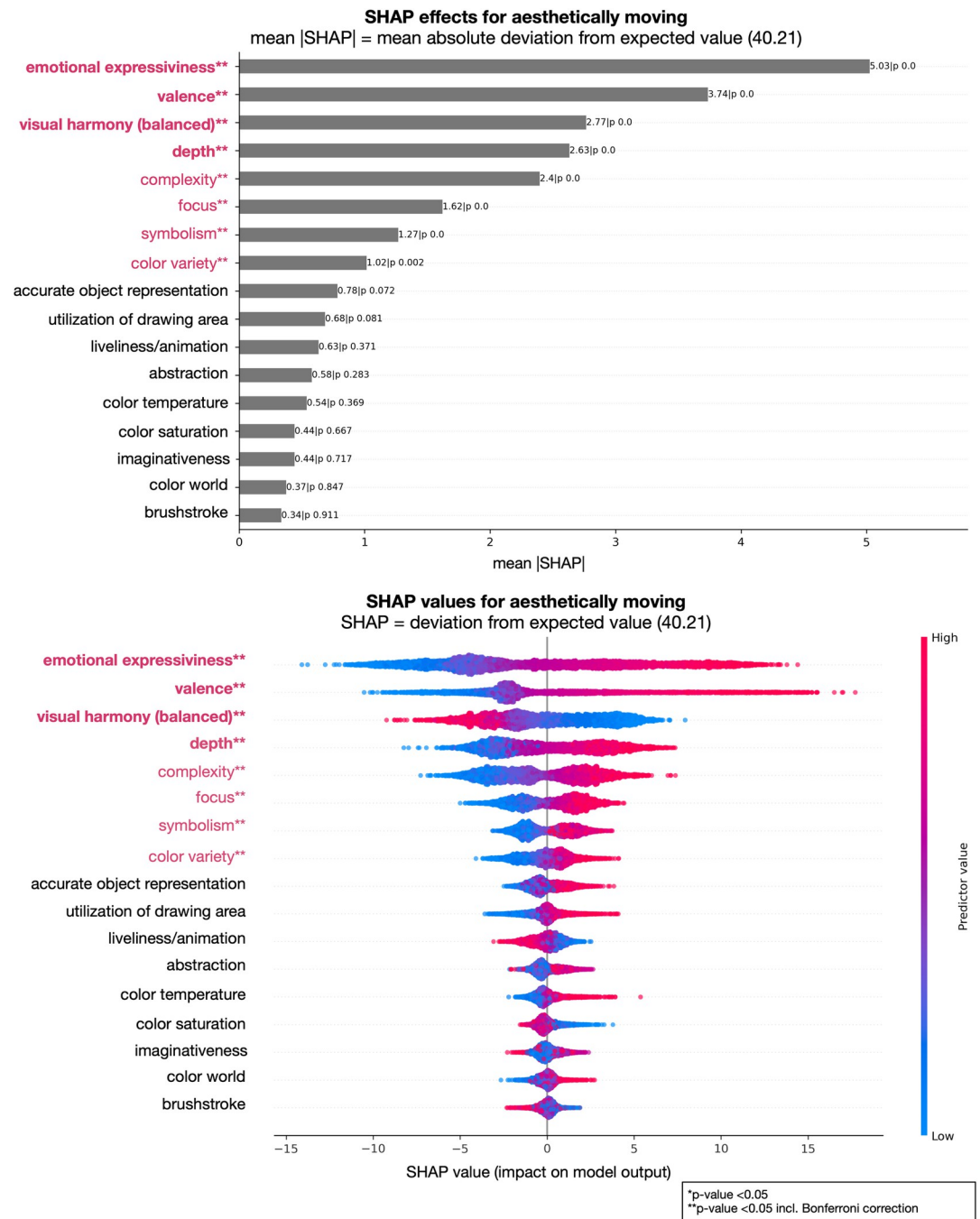
Firstly, emotional expressiveness and valence emerged as the most significant attributes across multiple judgments. Emotional expressiveness exhibited high importance for judgments related to aesthetically moving, beauty, good work of art, creativity, fascinating, interesting, thought-provoking, boring, and liking, with mean absolute values ranging from 3.37 to 7.25. Similarly, valence played a crucial role in judgments of aesthetically moving, beauty, good work of art, disturbing/irritating, familiarity, and liking, with mean values ranging from 3.35 to 6.24. Additionally, symbolism was a prominent attribute for creativity, epistemic aspects (fascinating, interesting, thought-provoking), adverse aspects (boring, disturbing/irritating), and the semantic aspect understanding, with mean values ranging from 3.81 to 7.81. Furthermore, within the content-representational attribute group, abstractness was found to be particularly important for the judgment of creativity and, to a lower extent, disturbing/irritating. Imaginativeness contributed to creativity and understanding, and to a lesser degree to adverse and epistemic aspects.

In the formal-perceptual features group, visual harmony (balance) demonstrated relevance for multiple judgments. It made notable contributions to aesthetically moving (moderate mean SHAP value: 2.77) and beauty (mean SHAP value: 5.26), as well as good work of art (mean SHAP value: 3.32) and liking (mean SHAP value: 4.14) among others.

Depth contributes to every judgment, however considering rank was often on the fifth or sixth ranks. The attribute was found to be significant for aesthetically moving, although with a relatively moderate mean SHAP value of 2.63. It also showed some relevance to technical skill, albeit with a modest mean SHAP value of 2.84. Interestingly, brushstroke was mainly contributing to technical skill (mean SHAP value: 3.77).

Complexity exhibited relevance for qualitative aspects, specifically good work of art, creativity, and technical skill. Further, complexity contributed with the lowest mean SHAP value of 2.79 for liking and the highest 4.68 for boring. Considering color attribute in general, hardly any showed strong impact on the judgments. In sum the results indicate that content-representational attributes were the dominant contributors often gaining ranks one to three with higher mean SHAP value compared to those in the formal-perceptual group.

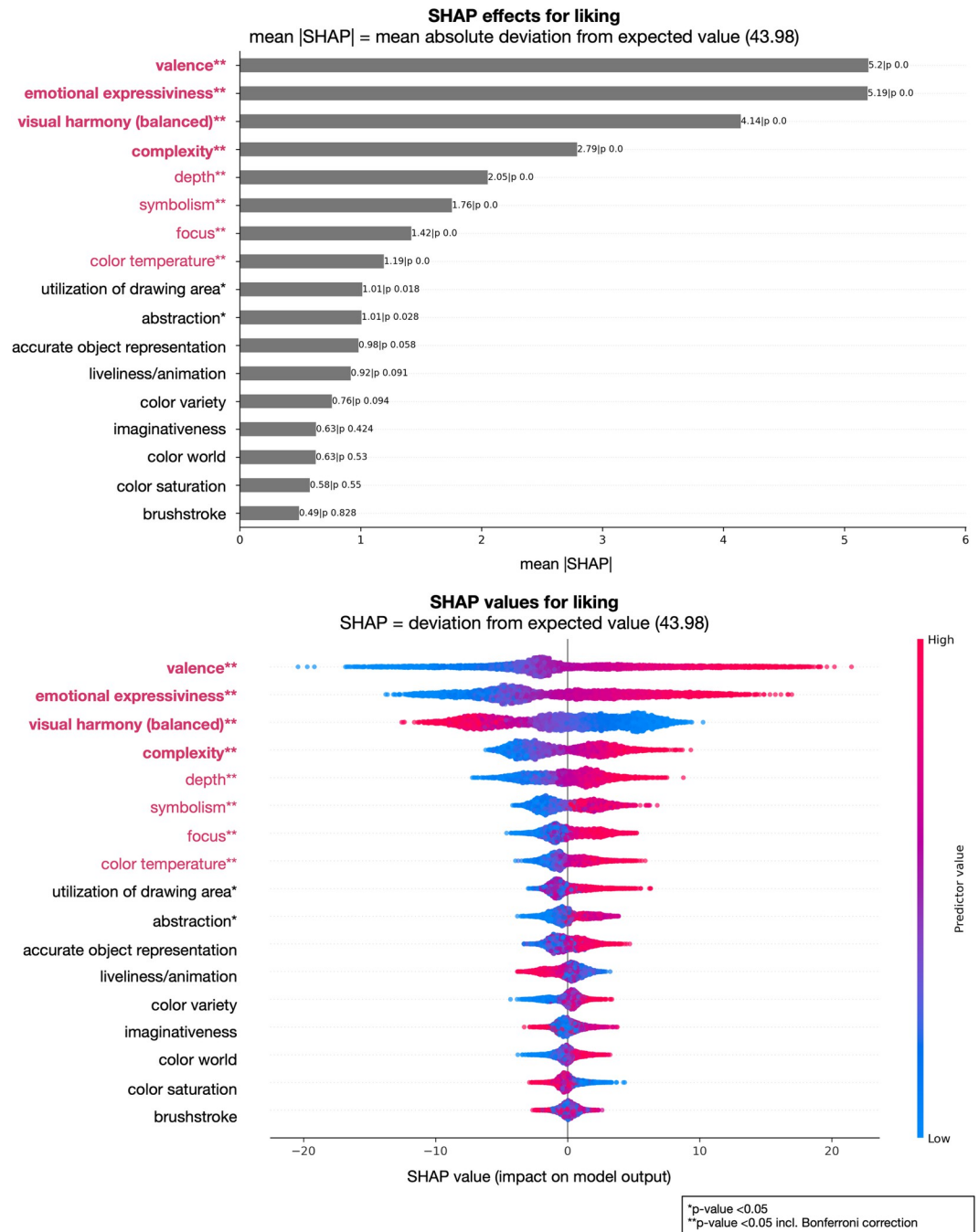
Furthermore, we examined the specific attributes for judgments depicted in Figs 1 to 4. For the judgments of aesthetically moving and liking, the order of importance for attributes was largely similar, with emotional expressiveness, valence, and visual harmony consistently



**Fig 1.** Art-attributes importance's for predicting aesthetically moving: Top plot shows mean absolute SHAP (SHapley Additive exPlanations) values for all 17 attributes in the model. Bottom plot displays SHAP values directly (color code: low = blue, red = high, and showing non-linearity of associations) and offers a detailed view of the individual impacts of predictors on specific predictions. **\*\*Bold and red attributes represent the most important attributes considering mean SHAP values; \*\*red attributes significant with Bonferroni correction, however with a lesser influence considering mean SHAP values; \*significant attributes before Bonferroni correction.**

<https://doi.org/10.1371/journal.pone.0304285.g001>

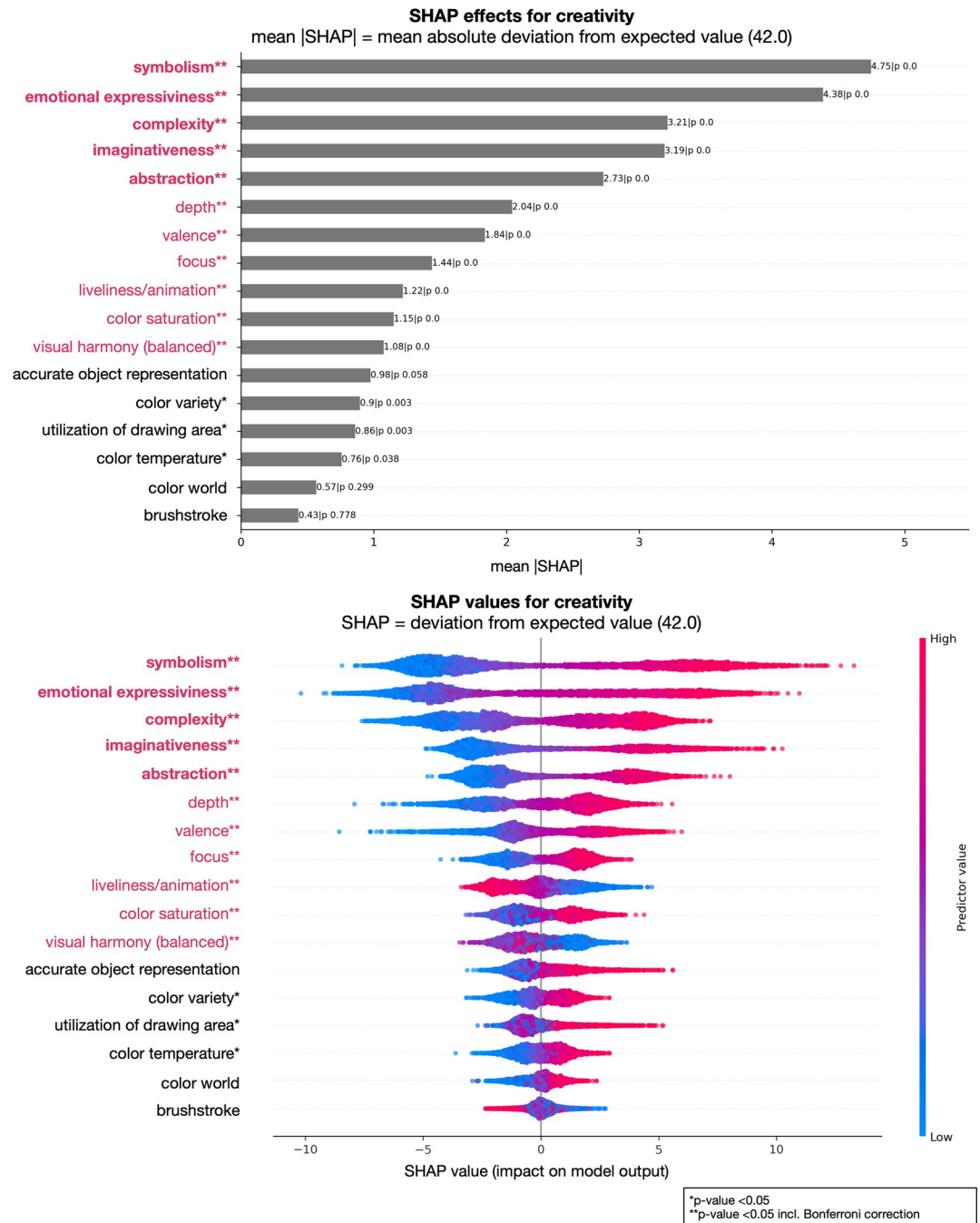
identified as the top three attributes. However, some attributes that remained significant after Bonferroni correction, but had relatively low mean SHAP values (< 2.70, see for detailed results <https://github.com/univiemops/art-rating-prediction>), did not play a substantial role in



**Fig 2.** Art-attributes importance's for predicting liking: Top plot shows mean absolute SHAP (SHapley Additive exPlanations) values for all 17 attributes in the model. Bottom plot displays SHAP values directly (color code: low = blue, red = high, and showing non-linearity of associations) and offers a detailed view of the individual impacts of predictors on specific predictions. \*\*Bold and red attributes represent the most important attributes considering mean SHAP values; \*\*red attributes significant with Bonferroni correction, however with a lesser influence considering mean SHAP values; \*significant attributes before Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0304285.g002>

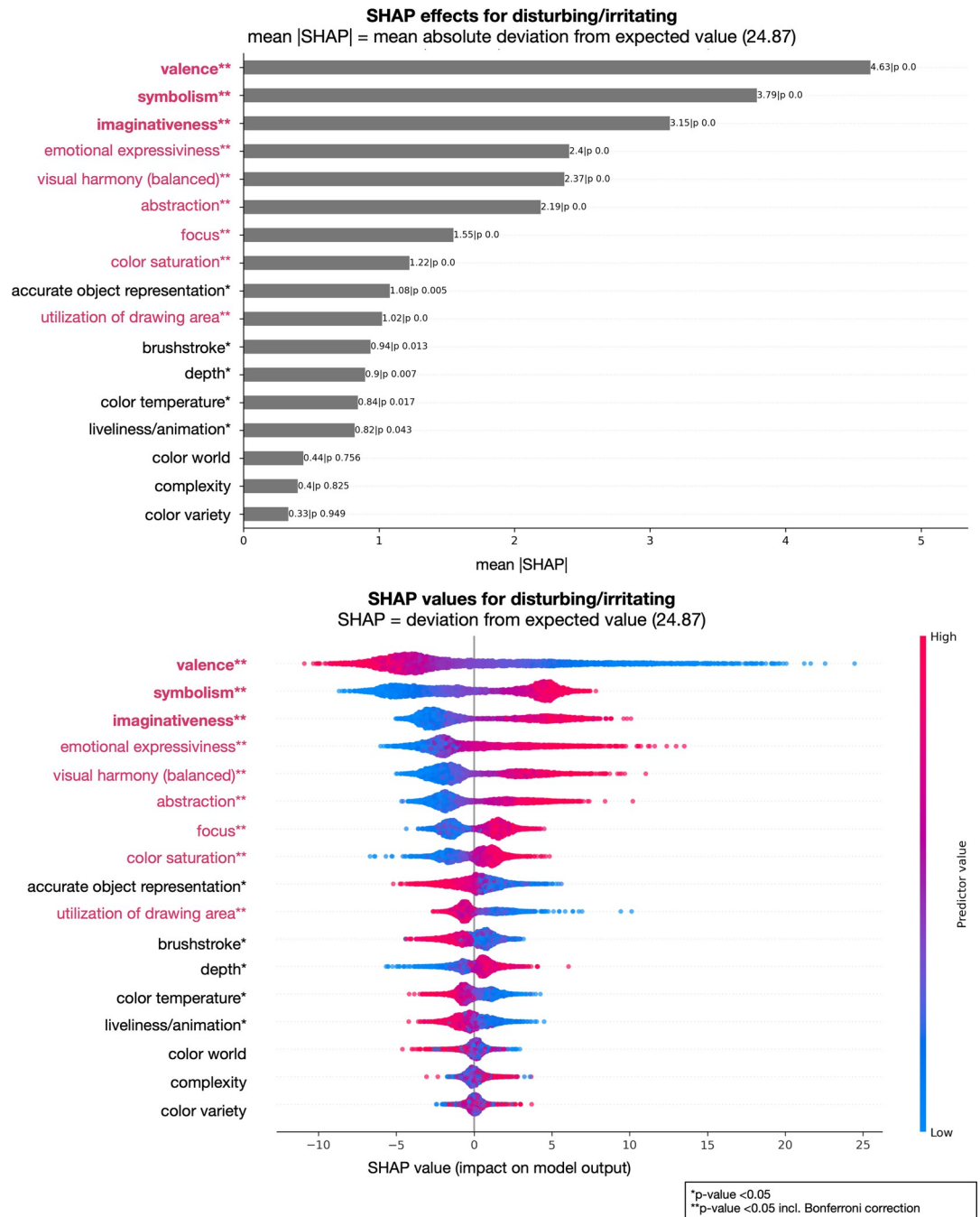
these judgments. Herein, specifically, depth and complexity exhibited opposite orders of importance in the two judgments.



**Fig 3.** Art-attributes importance's for predicting creativity: Top plot shows mean absolute SHAP (SHapley Additive exPlanations) values for all 17 attributes in the model. Bottom plot displays SHAP values directly (color code: low = blue, red = high, and showing non-linearity of associations) and offers a detailed view of the individual impacts of predictors on specific predictions. \*\*Bold and red attributes represent the most important attributes considering mean SHAP values; \*\*red attributes significant with Bonferroni correction, however with a lesser influence considering mean SHAP values; \*red attributes significant before Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0304285.g003>

When considering the attributes for aesthetically moving and liking with Bonferroni correction, additional attributes emerged as important factors. For aesthetically moving,



**Fig 4.** Art-attributes importance's for predicting disturbing/irritating: Top plot shows mean absolute SHAP (SHapley Additive exPlanations) values for all 17 attributes in the model. Bottom plot displays SHAP values directly (color code: low = blue, red = high, and showing non-linearity of associations) and offers a detailed view of the individual impacts of predictors on specific predictions. \*\*Bold and red attributes represent the most important attributes considering mean SHAP values; \*\*red attributes significant with Bonferroni correction, however with a lesser influence considering mean SHAP values; \*significant attributes before Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0304285.g004>

symbolism and color variety were notable attributes; while symbolism, focus, color temperature, utilization of drawing area, and abstractness were notable further attributes for liking.

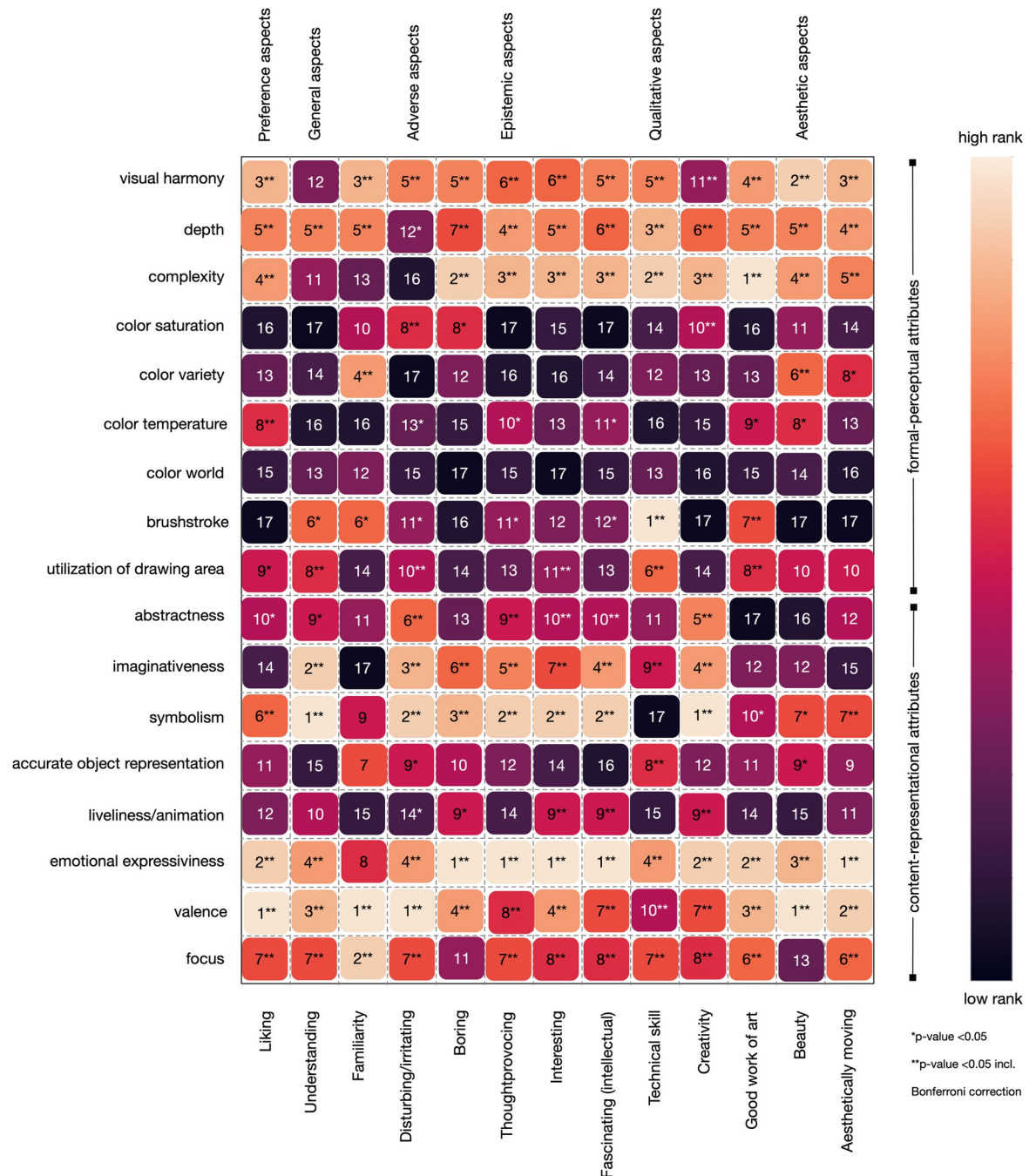


Fig 5. Rank based on mean SHAP-values for each attribute in association to the art judgments; \*p-value <0.05; \*\*p-value <0.05 incl. Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0304285.g005>

The judgment of creativity presented a different pattern, with symbolism identified as the most important attribute, followed by emotional expressiveness, complexity, imaginativeness, and abstractness. Depth, valence, focus, liveliness/animation, color saturation, and visual harmony demonstrated lesser importance based on mean SHAP values (see for detailed results <https://github.com/univiemops/art-rating-prediction>). Without Bonferroni correction, utilization of drawing area and color variety were found to contribute to creativity.



In the case of the judgment disturbing/irritating, valence emerged as the most important attribute, followed by symbolism and imaginativeness. Emotional expressiveness, visual harmony, abstractness, focus, color saturation, and utilization of drawing area showed lesser importance based on mean SHAP values (see Fig 4). Without Bonferroni correction, accurate object representation, depth, color temperature, and animation/liveliness were found to be contributing factors.

Additionally, in Figs 1 to 4 show on the bottom a visual representation of the impact of attributes on each judgment. This representation not only indicates the directionality of the impact (positive or negative) but also demonstrates the degree of influence, with attributes having varying levels of impact from high to low (color coding red to blue, respectively). Moreover, the non-linear associations between attributes and judgments are evident for single cases, further highlighting the complex and nuanced nature of the relationships; a relationship between predictors and target, which might have been overseen when using linear models.

In addition to the primary analyses, we conducted further model comparisons and an examination of predictor importance, the details of which are presented in S2 Appendix. Notably, we compared the performance of Gradient Boosting Decision Trees (GBDT) with various Generalized Linear Models (GLMs), including models without regularization (GLM-OLS), with regularization (GLM-Elastic-Net), and enhanced with polynomial features (GLM-Elastic-Net-poly-feat). Our findings highlighted GBDT's superior predictive accuracy, underscored the value of incorporating non-linear terms and regularization in modeling, and confirmed GBDT's leading position in terms of average  $R^2$  values across 13 judgments. Additionally, we analyzed the most significant predictors identified by both GBDT and GLM-Elastic-Net linear regression, revealing a remarkable consensus on the top predictors despite the intrinsic differences between linear and non-linear modeling approaches. For those interested in a deeper exploration of all predictors and their comprehensive rankings across different prediction targets and methods, we have made the data available on our GitHub repository (<https://github.com/univiemops/art-rating-prediction>).

Last, we provide a correlation heatmap (see S2 Fig). These heatmap reveal the following pattern: although there are correlations among the art attributes themselves and the art judgments themselves, there are minimal correlations between the art judgments and the attributes. In other words, while the attributes may exhibit relationships with each other, as do the judgments, the attributes do not strongly correlate with any judgments. These finding support our machine learning approach and point to the complex and multifaceted nature of art evaluation.

## Discussion

The present study aimed to investigate the relationship between art judgments and art attributes using a machine learning framework. Through non-linear modeling techniques, we aimed to overcome the limitations of traditional linear models and gain a deeper understanding of the complex dynamics underlying art evaluation [16, 17, 38, 40–43]. Our findings are based on prior research focusing on creativity judgments only [12]. We now broadened and updated the analysis to in total 13 art judgements using 17 attributes as predictors.

Our study revealed that specific art attributes significantly predicted art judgments, with content-representational attributes playing a more prominent role than formal-perceptual attributes addressing individual judgments. This finding aligns with previous research suggesting that the subjective interpretation and evaluation of artworks rely heavily on the content and meaning conveyed by the artwork [3, 5, 12, 14, 92]. Emotional expressiveness, valence, and symbolism emerged as particularly important attributes, appearing in the

predictions for multiple judgments. These attributes capture the strength of the artist to express emotion, the positivity/negativity, and the level of ambiguity/symbolistic meaning associated with artworks [1, 35, 66, 83].

The non-linear relationships observed between art attributes and judgments further highlight the complexity of art evaluation. Some cases exhibited sudden inclines or declines around the midpoint of the scales, suggesting that the impact of an attribute on a judgment may vary depending on its magnitude. This finding underscores the need to consider the multidimensional nature of art evaluation and an interplay between different attributes [37, 38, 42, 43, 93]. This provides a fascinating implication for future studies suggesting that judgments are not solely determined by the presence or absence of specific attributes but rather by the dynamic interaction between various attributes.

Our findings considering the non-linear associations with sudden in- or declines hint to an intriguing pattern, which might best be understood as an aesthetic threshold [see also discussion in, 12]. This phenomenon might hint to a critical point at which the perception and evaluation of an artwork shift significantly, a concept also suggested in psychophysics where perceptual changes do regularly scale non-linear with stimulus variation. Such aesthetic thresholds have been discussed to occur around the midpoint of the rating scales for several art attributes, where Berlyne's studies and later work suggest with some attributes, such as complexity, a non-linear relationship [5, 27, 32, 37, 94, 95].

Importantly, our study contributes to addressing the gap in the literature regarding the relationship between art judgments and attributes. First, previous research often focused on a limited number of judgments and attributes without systematically examining their connections. Second, one key distinction that past literature has often overlooked is the differentiation between art judgments and art attributes. Art judgments refer to the evaluative opinions and assessments individuals make about artworks, encompassing concepts such as beauty, liking, interestingness, and thought-provoking. On the other hand, art attributes are specific qualities or features of artworks that individuals use to substantiate their evaluative arguments. These attributes provide a foundation for articulating and validating subjective evaluations, such as colorfulness, emotional expressiveness, imaginativeness, and complexity. The correlation heatmap (see [S2 Fig](#)) support this notion, showing very low correlations between judgment and attributes. By failing to dissociate between judgments and attributes, previous research has not fully explored the underlying factors and mechanisms that shape art evaluations.

Third, future research should also consider the difference between evaluating an artworks representation and the focus of elicited emotion within the person. Although, naturally, emotion and evaluation cannot be departed, the focus of judgments to be provided can be framed, which we tried to respect in our study. The task given should be clearly stated in the participant instructions. However, we acknowledge that also we cannot prove if we were fully successful within our study; future studies could investigate differences in task instructions (focus on artwork evaluation or on own emotional experience).

Considering the attributes importance we found a dominance of higher importance of content-representational attributes in predicting art judgments. This suggests the importance of meaning and emotional resonance in aesthetic experiences. Artworks are a medium that can represent and mirror emotions, convey symbolic messages, and resonate with personal experiences. This is a fundamental aspect of art appreciation supporting also new frameworks in the field considering this interplay of knowledge, emotion, and the dynamic interaction predicting the value of an experience [67, 96, 97]. Therefore, our research approach should also be extended to investigate this interplay, starting with the association between the strength of emotional expressiveness and the elicitation of emotions and connected to meaning making

based on experience [98]. Our results also support further study of aesthetic experiences in visual art research specifically, such as awe and wonder [99–102].

Our findings show that art judgments are not solely based on formal aspects such as color or composition but are strongly influenced by the content and meaning portrayed by the artworks, however, it is worth noting that some formal-perceptual attributes also made significant contributions to the predictions, albeit to a lesser extent than content-representational attributes.

Considering formal-perceptual features, especially visual harmony, depth, but also complexity were significant contributors (see for example, Figs 1 and 2). Complexity emerged as the most relevant formal-perceptual attribute, particularly in judgments related to boring (simple images were more boring), interestingness (low complex artworks were less interesting), and all judgments from the qualitative aspects (good work of art, creativity, and technical skill) showing that high complex art images received higher ratings. Note, for good work of art complexity was the most important contributor. These finding is consistent with previous research highlighting the role of complexity in art evaluation and the influence of cognitive processes on aesthetic experiences [37, 94, 103]. Visual harmony also appeared as a predictor for the judgments aesthetically moving, beauty, and liking indicating its relevance to the very specific evaluation of appreciating aesthetic elements in art. Here, we should highlight our cohort, where laymen might orient their liking more on aesthetic appeal; while art experts, for example, might use other attributes for their personal preferences.

The low contributions from certain attributes, such as color saturation, color variety, and other color aspects, suggests that these attributes may have limited impact on art judgments in our art novice study cohort. These finding challenges previous assumptions about the universality and centrality of these attributes in art evaluation [65, 104], although still may have relevance in cross-cultural comparison studies [69, 105] or comparisons with art experts [106–108]. The absence of some attributes also highlights the need for a more nuanced understanding of the contextual factors that influence the importance and relevance of different attributes in art appreciation.

In sum, our research reveals a complex interplay of attributes contributing to art judgments, where sometimes the same attributes have high impact for different judgments. Apparently, it lies in the nuances and interplay of the other attributes, that determine the judgments differently. Another assumption would be to assume that art novices might treat some judgments as being the same (for example, beauty and liking). To support this interpretation more research is necessary in diverse cohorts and certainly also employing more within subject comparison study designs.

Our study also expands on previous research by incorporating machine learning methodologies and non-linear modeling techniques [12, 48–52, 109]. By employing machine learning techniques, we were able to identify a comprehensive set of attributes that significantly contribute to the predictions of art judgments [16, 17, 40–43]. This data-driven approach provides a more nuanced understanding of the factors that shape art evaluations and highlights the interplay between different attributes.

This approach allowed us to capture complex relationships and patterns that may have been overlooked by traditional linear models [38, 40]. By leveraging machine learning algorithms, we were able to construct data-driven models that provide a more comprehensive understanding of art evaluation. This approach has the potential to uncover hidden patterns, regularities, and potential inconsistencies within art judgments, contributing to the advancement of knowledge in aesthetics.

The findings of our study have implications for various domains, including art education, curatorial practices, and the development of computational models of aesthetics. By gaining a

deeper understanding of the factors that shape art judgments, educators, curators, and online platforms can design more engaging and meaningful art experiences, adapted to the target group. They can emphasize the importance of emotional expressiveness, symbolism, and other content-representational attributes in fostering aesthetic engagement and appreciation, addressing for laymen rather the whole artwork than individual elements [4]. Furthermore, our findings provide valuable insights for the development of computational models that simulate human-like art evaluations. By incorporating the identified attributes and their non-linear relationships, these models can generate more realistic and nuanced judgments of artworks.

Despite the valuable insights provided by our study, there are also limitations. First, we only tested art novices, which limits the generalizability, especially to expert art evaluators or individuals with different cultural backgrounds. Future research could explore the potential differences in the relationship between art judgments and attributes across diverse populations, employing the same procedures. Second, the selection of artworks in our study was limited to artworks of Western culture, which restricts the generalizability of the results to other artistic traditions and cultures. Future studies could incorporate artworks from a wider range of cultural contexts to examine the cross-cultural variations in art judgments and their associated attributes. Lastly, the use of machine learning techniques introduces the challenge of interpretability. While these models provide accurate predictions, understanding the underlying mechanisms and causal relationships between judgments and attributes may require additional analysis and experimentation.

In conclusion, our study provides valuable insights into the complex interplay between art judgments and attributes. By employing machine learning techniques and non-linear modeling, we uncovered the significance of content-representational mainly and a view formal-perceptual attributes in shaping art judgments. Among these attributes, complexity and visual harmony are certainly the two most interesting ones. Our findings highlight the non-linear nature of the relationships between attributes and judgments and emphasize the importance of emotional expressiveness, valence, symbolism, and complexity in art appreciation. These insights contribute to a deeper understanding of the multidimensional nature of aesthetic experiences and have implications for art education, art practices, and the development of studying non-linear human behavior in art research. In addition, we also want to highlight the apparent relevance of level of symbolism (interpretability), meaning, and emotional expressivity. Future research can build upon these findings by exploring the cross-cultural variations in art judgments, focus on one judgement or attributes solely, investigate the role of individual differences, and further refining the understanding of the underlying mechanisms of art evaluation.

## Supporting information

**S1 Table. Scales of art-judgements (targets in machine learning analysis) used in the study, German version.**

(PDF)

**S2 Table. Scales of art attributes (predictors in machine learning analysis) used in the study, German version.**

(PDF)

**S3 Table. Full list of stimulus-set rated including name of artist, title, style, and depicted motif. ID represents the VAPS-identification code.**

(PDF)

**S4 Table. Stimulus selection of VAPS using Wilk-Shapiro test for ensuring best possible normal distribution of VAPS ratings while including images.**

(PDF)

**S1 Fig. Procedures of systematic literature review.**

(TIF)

**S2 Fig. Correlation heatmap of art judgments and attributes.**

(TIF)

**S1 Appendix. Review of art related judgments and attributes, including list of judgements descriptions [1, 3, 5, 12, 13, 15, 18, 23, 24, 27–32, 34–37, 46–50, 59–65, 67–69, 72, 73, 94, 95, 98–102, 105, 108, 110–179].**

(DOCX)

**S2 Appendix. Model comparisons and an examination of predictor importance.**

(DOCX)

## Author Contributions

**Conceptualization:** Blanca T. M. Spee, Helmut Leder, Jan Mikuni, Matthew Pelowski.

**Data curation:** Blanca T. M. Spee, Jan Mikuni.

**Formal analysis:** Blanca T. M. Spee, Jan Mikuni, Frank Scharnowski, David Steyrl.

**Investigation:** Blanca T. M. Spee.

**Methodology:** Blanca T. M. Spee.

**Project administration:** Blanca T. M. Spee.

**Software:** David Steyrl.

**Supervision:** Helmut Leder, Frank Scharnowski.

**Validation:** Blanca T. M. Spee, David Steyrl.

**Visualization:** Blanca T. M. Spee.

**Writing – original draft:** Blanca T. M. Spee, Helmut Leder, David Steyrl.

**Writing – review & editing:** Blanca T. M. Spee, Helmut Leder, Jan Mikuni, Frank Scharnowski, Matthew Pelowski, David Steyrl.

## References

1. Leder H, Nadal M. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—Developments and challenges in empirical aesthetics. *British Journal of Psychology*. 2014; 105: 443–464. <https://doi.org/10.1111/bjop.12084> PMID: 25280118
2. Pelowski M, Gerger G, Chetouani Y, Markey PS, Leder H. But Is It really Art? The classification of images as “Art”/“Not Art” and correlation with appraisal and viewer interpersonal differences. *Front Psychol*. 2017; 8. <https://doi.org/10.3389/fpsyg.2017.01729> PMID: 29062292
3. Pelowski M, Markey PS, Luring JO, Leder H. Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Front Hum Neurosci*. 2016; 10: 1–21. <https://doi.org/10.3389/fnhum.2016.00160> PMID: 27199697
4. Specker E, Forster M, Brinkmann H, Boddy J, Immelmann B, Goller J, et al. Warm, lively, rough? Assessing agreement on aesthetic effects of artworks. *PLoS One*. 2020; 15: e0232083. <https://doi.org/10.1371/journal.pone.0232083> PMID: 32401777
5. Cupchik GC, Berlyne DE. The perception of collative properties in visual stimuli. *Scand J Psychol*. 1979; 20: 93–104. <https://doi.org/10.1111/j.1467-9450.1979.tb00688.x> PMID: 451496

6. Fingerhut J. Habits and the Enculturated Mind: Pervasive Artifacts, Predictive Processing, and Expansive Habits. *Habits: Pragmatist Approaches from Cognitive Neuroscience to Social Science*. 2020. pp. 352–375.
7. Becker HS. Art as collective action. *Am Sociol Rev*. 1974; 39: 767–776.
8. Stamkou E, Keltner D. Aesthetic Revolution: The Role of Art in Culture and Social Change. *SSRN Electronic Journal*. 2020. <https://doi.org/10.2139/ssrn.3578575>
9. Spee BTM, Pelowski M, Arato J, Mikuni J, Tran US, Eisenegger C, et al. Social reputation influences on liking and willingness-to-pay for artworks: A multimethod design investigating choice behavior along with physiological measures and motivational factors. *PLoS One*. 2022; 17: e0266020. <https://doi.org/10.1371/journal.pone.0266020> PMID: 35442966
10. Stamkou E, van Kleef GA, Homan AC. The art of influence: When and why deviant artists gain impact. *J Pers Soc Psychol*. 2018;115. <https://doi.org/10.1037/pspi0000131> PMID: 30024244
11. Becker HS. *Art worlds*. Berkeley: University of California Press; 1982.
12. Spee BTM, Mikuni J, Leder H, Scharnowski F, Pelowski M, Steyerl D. Machine learning revealed symbolism, emotionality, and imaginativeness as primary predictors of creativity evaluations of western art paintings. *Sci Rep*. 2023; 13: 12966. <https://doi.org/10.1038/s41598-023-39865-1> PMID: 37563194
13. Sidhu DM, McDougall KH, Jalava ST, Bodner GE. Prediction of beauty and liking ratings for abstract and representational paintings using subjective and objective measures. *PLoS One*. 2018; 13: 1–15. <https://doi.org/10.1371/journal.pone.0200431> PMID: 29979779
14. Leder H, Gerger G, Dressler SG, Schabmann A. How art is appreciated. *Psychol Aesthet Creat Arts*. 2012; 6: 2–10. <https://doi.org/10.1037/a0026396>
15. Cupchik GC, Vartanian O, Crawley A, Mikulis DJ. Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain Cogn*. 2009; 70: 84–91. <https://doi.org/10.1016/j.bandc.2009.01.003> PMID: 19223099
16. Breiman L. Random Forests. *Mach Learn*. 2001; 45: 5–32.
17. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*. *Adv Neural Inf Process Syst*. 2017; 30: 3146–3154.
18. Chatterjee A, Widick P, Sternschein R, Smith W, Bromberger B. The assessment of art attributes. *Empirical Studies of the Arts*. 2010; 28: 207–222. <https://doi.org/10.2190/EM.28.2.f>
19. Belfiore E, Bennett O. *The Social Impact of the Arts*. London: Palgrave Macmillan UK; 2008. <https://doi.org/10.1057/9780230227774>
20. Alexander VD, Bowler AE. Art at the crossroads: The arts in society and the sociology of art. *Poetics*. 2014; 43: 1–19. <https://doi.org/10.1016/j.poetic.2014.02.003>
21. Wands B. *Art of The Digital Age*. USA: Thames and Hudson; 2007.
22. Huang Y-T, Su S-F. Motives for Instagram Use and Topics of Interest among Young Adults. *Future Internet*. 2018; 10. <https://doi.org/10.3390/fi10080077>
23. Kant I. *Critique of Judgment*. 1st Ed. 1790. Indianapolis, Cambridge: Hackett Publishing Company; 1987.
24. Baumgarten AG. *Aesthetica*. 1st Ed. 1750. Hildesheim: Georg Olms Verlag; 1970.
25. Dissanayake E. The Artification Hypothesis and Its Relevance to Cognitive Science, Evolutionary Aesthetics, and Neuroaesthetics. *Cognitive Semiotics*. 2009; 5: 136–191.
26. Berlyne DE. *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts; 1971.
27. Berlyne DE. *Conflict, arousal, and curiosity*. New York: McGraw-Hill; 1960.
28. Berlyne DE. *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation*. Oxford: Hemisphere; 1974.
29. Nadal M, Ureña E. *One hundred years of Empirical Aesthetics: Fechner to Berlyne (1876–1976)*. 2021.
30. Martindale C, Moore K, Borkum J. Aesthetic Preference: Anomalous Findings for Berlyne's Psychobiological Theory. *Am J Psychol*. 1990; 103: 53–80. Available: <https://www.jstor.org/stable/1423259>
31. Martindale C. Recent Trends in the Psychological Study of Aesthetics, Creativity, and the Arts. *Empirical Studies of the Arts*. 2007; 25: 121–141. <https://doi.org/10.2190/B637-1041-2635-16NN>
32. Marin MM, Lampatz A, Wandl M, Leder H. Berlyne revisited: Evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Front Hum Neurosci*. 2016; 10: 1–20. <https://doi.org/10.3389/fnhum.2016.00536> PMID: 27867350

33. Commare L, Rosenberg R, Leder H. More than the sum of its parts: Perceiving complexity in painting. *Psychol Aesthet Creat Arts*. 2018; 12: 380–391. <https://doi.org/10.1037/aca0000186>
34. Leder H, Belke B, Oeberst A, Augustin D. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*. 2004. Available: [www.bps.org.uk](http://www.bps.org.uk)
35. Pelowski M, Markey PS, Forster M, Gerger G, Leder H. Move me, astonish me. . . delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*. Elsevier B.V.; 2017. pp. 80–125. <https://doi.org/10.1016/j.plrev.2017.02.003> PMID: 28347673
36. Leder H, Ring A, Dressler SG. See me, feel me! Aesthetic evaluations of art portraits. *Psychol Aesthet Creat Arts*. 2013; 7: 358–369. <https://doi.org/10.1037/a0033311>
37. Van Geert E, Wagemans J. Order, complexity, and aesthetic appreciation. *Psychol Aesthet Creat Arts*. 2020; 14. <https://doi.org/10.1037/aca0000224>
38. Glaser JI, Benjamin AS, Farhoodi R, Kording KP. The Roles of Supervised Machine Learning in Systems Neuroscience. *Prog Neurobiol*. 2019; 175: 126–137. <https://doi.org/10.1016/j.pneurobio.2019.01.008> PMID: 30738835
39. Alpaydin E. *Maschinelles Lernen*. Walter de Gruyter GmbH & Co KG; 2022.
40. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol*. 2018; 14: 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037> PMID: 29401044
41. Breiman L. Statistical Modeling: The Two Cultures. *Statistical Science*. 2001; 16. <https://doi.org/10.1214/ss/1009213726>
42. Jolly E, Chang LJ. The Flatland Fallacy: Moving Beyond Low-Dimensional Thinking. *Top Cogn Sci*. 2019; 11: 433–454. <https://doi.org/10.1111/tops.12404> PMID: 30576066
43. Rocca R, Yarkoni T. Putting Psychology to the Test: Rethinking Model Evaluation Through Benchmarking and Prediction. *Adv Methods Pract Psychol Sci*. 2021; 4: 251524592110268. <https://doi.org/10.1177/25152459211026864> PMID: 38737598
44. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *Am J Epidemiol*. 2013; 177: 292–298. <https://doi.org/10.1093/aje/kws412> PMID: 23371353
45. Bense M. *Aesthetica: Einführung in die neue Aesthetik*. Baden-Baden: AGIS-Verlag; 1965.
46. Birkhoff GD. *Aesthetic Measure*. Cambridge: Massachusetts Harvard; 1933.
47. Berlyne DE. Interrelations of verbal and nonverbal measures used in experimental aesthetics. *Scand J Psychol*. 1973; 14: 177–184. <https://doi.org/10.1111/j.1467-9450.1973.tb00107.x> PMID: 4759280
48. Li C, Chen T. Aesthetic Visual Quality Assessment of Paintings. *IEEE J Sel Top Signal Process*. 2009; 3. <https://doi.org/10.1109/JSTSP.2009.2015077>
49. Li Y, Hu C, Minku LL, Zuo H. Learning aesthetic judgements in evolutionary art systems. *Genet Program Evolvable Mach*. 2013; 14: 315–337. <https://doi.org/10.1007/s10710-013-9188-7>
50. Iigaya K, Yi S, Wahle IA, Tanwisuth K, O'Doherty JP. Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.02.09.940353>
51. Iigaya K, Yi S, Wahle IA, Tanwisuth K, O'Doherty JP. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nat Hum Behav*. 2021; 5. <https://doi.org/10.1038/s41562-021-01124-6> PMID: 34017097
52. Iigaya K, Yi S, Wahle IA, Tanwisuth S, Cross L, O'Doherty JP. Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nat Commun*. 2023; 14: 127. <https://doi.org/10.1038/s41467-022-35654-y> PMID: 36693833
53. Chatterjee A, Vartanian O. Neuroaesthetics. *Trends in Cognitive Sciences*. Elsevier Ltd; 2014. pp. 370–375. <https://doi.org/10.1016/j.tics.2014.03.003> PMID: 24768244
54. Pearce MT, Zaidel DW, Vartanian O, Skov M, Leder H, Chatterjee A, et al. Neuroaesthetics: The Cognitive Neuroscience of Aesthetic Experience. *Perspectives on Psychological Science*. 2016; 11: 265–279. <https://doi.org/10.1177/1745691615621274> PMID: 26993278
55. Leder H. Next steps in neuroaesthetics: Which processes and processing stages to study? *Psychol Aesthet Creat Arts*. 2013; 7: 27–37. <https://doi.org/10.1037/a0031585>
56. Fingerhut J. Enacting Media. An Embodied Account of Enculturation Between Neuromedia and New Cognitive Media Theory. *Front Psychol*. 2021; 12. <https://doi.org/10.3389/fpsyg.2021.635993> PMID: 34113285
57. Van de Cruys S, Wagemans J. Gestalts as predictions: Some reflections and an application to art. *Gestalt Theory*. 2011; 33: 325–344.

58. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* 2009; 6. <https://doi.org/10.1371/journal.pmed.1000097> PMID: 19621072
59. Schindler I, Hosoya G, Menninghaus W, Beermann U, Wagner V, Eid M, et al. Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLoS ONE. Public Library of Science*; 2017. <https://doi.org/10.1371/journal.pone.0178899> PMID: 28582467
60. Eysenck HJ. A Critical and Experimental Study of Colour Preferences. *Am J Psychol.* 1941; 54. <https://doi.org/10.2307/1417683>
61. Jonauskaitė D, Wicker J, Mohr C, Dael N, Havelka J, Papadatou-Pastou M, et al. A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *R Soc Open Sci.* 2019; 6. <https://doi.org/10.1098/rsos.190741> PMID: 31598303
62. Mayer S, Landwehr JR. Quantifying visual aesthetics based on processing fluency theory: Four algorithmic measures for antecedents of aesthetic preferences. *Psychol Aesthet Creat Arts.* 2018; 12. <https://doi.org/10.1037/aca0000187>
63. Palmer SE, Schloss KB, Sammartino J. Visual aesthetics and human preference. *Annu Rev Psychol.* 2013; 64: 77–107. <https://doi.org/10.1146/annurev-psych-120710-100504> PMID: 23020642
64. Stamatopoulou D. Integrating the Philosophy and Psychology of Aesthetic Experience: Development of the Aesthetic Experience Scale. *Psychol Rep.* 2004; 95. <https://doi.org/10.2466/pr0.95.2.673-695> PMID: 15587237
65. Palmer SE, Schloss KB, Sammartino J. Hidden Knowledge in Aesthetic Judgments: Preference for Color and Spatial Composition. In: Shimamura AP, Palmer SE, editors. *Aesthetic Science Connecting minds, brains and experience.* New York: Oxford University Press; 2012. pp. 189–222.
66. Schepman A, Rodway P, Pullen SJ, Kirkham J. Shared liking and association valence for representational art but not abstract art. *J Vis.* 2015; 15. <https://doi.org/10.1167/15.5.11> PMID: 26067529
67. Van de Cruys S, Chamberlain R, Wagemans J. Tuning in to art: A predictive processing account of negative emotion in art. *Behav Brain Sci.* 2017; 40: e377. <https://doi.org/10.1017/S0140525X17001868> PMID: 29342804
68. Menninghaus W, Wagner V, Hanich J, Wassiliwizky E, Jacobsen T, Koelsch S. The Distancing-Embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences.* 2017; 40. <https://doi.org/10.1017/S0140525X17000309> PMID: 28215214
69. Che J, Sun X, Gallardo V, Nadal M. Cross-cultural empirical aesthetics. *Progress in Brain Research.* Elsevier B.V.; 2018. pp. 77–103. <https://doi.org/10.1016/bs.pbr.2018.03.002> PMID: 29779752
70. Chua HF, Boland JE, Nisbett RE. Cultural variation in eye movements during scene perception. *PNAS* August. 2005. Available: [www.pnas.org/cgi/doi/10.1073/pnas.0506162102](http://www.pnas.org/cgi/doi/10.1073/pnas.0506162102) <https://doi.org/10.1073/pnas.0506162102> PMID: 16116075
71. Redies C. Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front Hum Neurosci.* 2015; 9: 1–20. <https://doi.org/10.3389/fnhum.2015.00218> PMID: 25972799
72. Osgood CE. Semantic Differential Technique in the Comparative Study of Cultures<sup>1</sup>. *Am Anthropol.* 1964; 66: 171–200. <https://doi.org/10.1525/aa.1964.66.3.02a00880>
73. Pelowski M, Markey PS, Goller J, Förster EL, Leder H. But, How Can We Make “Art?” Artistic Production Versus Realistic Copying and Perceptual Advantages of Artists. *Psychol Aesthet Creat Arts.* 2018. <https://doi.org/10.1037/aca0000178>
74. Baer J, McKool SS. Assessing Creativity Using the Consensual Assessment Technique. *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education.* IGI Global; <https://doi.org/10.4018/978-1-60566-667-9.ch004>
75. Pelowski M, Förster M, Tinio PPL, Scholl M, Leder H. Beyond the lab: An examination of key factors influencing interaction with “Real” and museum-based art. *Psychol Aesthet Creat Arts.* 2017; 11: 245–264. <https://doi.org/10.1037/aca0000141>
76. Pugach C, Leder H, Graham DJ. How stable are human aesthetic preferences across the lifespan? *Front Hum Neurosci.* 2017; 11. <https://doi.org/10.3389/fnhum.2017.00289> PMID: 28620290
77. Lakens D. Sample Size Justification. *Collabra Psychol.* 2022; 8. <https://doi.org/10.1525/collabra.33267>
78. Rosenbusch H, Soldner F, Evans AM, Zeelenberg M. Supervised machine learning methods in psychology: A practical introduction with annotated R code. *Soc Personal Psychol Compass.* 2021; 15. <https://doi.org/10.1111/spc3.12579>
79. Pedregosa F, Varoquaux G, Gramfort A, Vincent M, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.* 2011; 12: 2825–2830. Available: <http://scikit-learn.sourceforge.net>.



80. Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *Am J Epidemiol*. 2007; 165. <https://doi.org/10.1093/aje/kwk052> PMID: 17182981
81. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II—binary and time-to-event outcomes. *Stat Med*. 2019; 38: 1276–1296. <https://doi.org/10.1002/sim.7992> PMID: 30357870
82. Cohen J. *Statistical power analysis for the behavioural sciences*. Hillsdale, New York: Laurence Erlbaum Associates; 1988.
83. Fekete A, Pelowski M, Specker E, Brieber D, Rosenberg R, Leder H. The Vienna Art Picture System (VAPS): A data set of 999 paintings and subjective ratings for art and aesthetics research. *Psychol Aesthet Creat Arts*. 2022. <https://doi.org/10.1037/aca0000460>
84. Leiner DJ. SoSci Survey (Version 3.2.43) [Computer Software]. In: <http://www.sosicisurvey.com>. 2018.
85. Grinsztajn L, Oyallon E, Varoquaux G. Advances in Neural Information Processing Systems. 2022; 35: 507–520.
86. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*. 2010; 11: 2079–2107.
87. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics; 2009.
88. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 2017.
89. Molnar C. Interpretable machine learning. A Guide for Making Black Box Models Explainable. In: <https://christophm.github.io/interpretable-ml-book/>. 2019.
90. Bouckaert RR, Frank E. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. 2004. pp. 3–12. [https://doi.org/10.1007/978-3-540-24775-3\\_3](https://doi.org/10.1007/978-3-540-24775-3_3)
91. Nadeau C. Inference for the Generalization Error. *Machine Learning*. *Mach Learn*. 2003; 52: 239–281. <https://doi.org/10.1023/A:1024068626366>
92. Augustin MD, Leder H, Hutzler F, Carbon CC. Style follows content: On the microgenesis of art perception. *Acta Psychol (Amst)*. 2008; 128: 127–138. <https://doi.org/10.1016/j.actpsy.2007.11.006> PMID: 18164689
93. Augustin MD, Wagemans J. Empirical aesthetics, the beautiful challenge: An introduction to the special issue on art & perception. *Iperception*. 2012; 3: 455–458. <https://doi.org/10.1068/i0541aap> PMID: 23145296
94. Jakesch M, Leder H. The qualitative side of complexity: Testing effects of ambiguity on complexity judgments. *Psychol Aesthet Creat Arts*. 2015; 9: 200–205. <https://doi.org/10.1037/a0039350>
95. Markey PS, Jakesch M, Leder H. Art looks different—Semantic and syntactic processing of paintings and associated neurophysiological brain responses. *Brain Cogn*. 2019; 134: 58–66. <https://doi.org/10.1016/j.bandc.2019.05.008> PMID: 31151085
96. Van de Cruys S. Affective Value in the Predictive Mind. *Open MIND*. 2017; 1–21. <https://doi.org/10.15502/9783958573253>
97. Van de Cruys S, Wagemans J. Putting reward in art: A tentative prediction error account of visual art. *Iperception*. 2011; 2: 1035–1062. <https://doi.org/10.1068/i0466aap> PMID: 23145260
98. Sarasso P, Neppi-Modona M, Sacco K, Ronga I. “Stopping for knowledge”: The sense of beauty in the perception-action cycle. *Neurosci Biobehav Rev*. 2020; 118: 723–738. <https://doi.org/10.1016/j.neubiorev.2020.09.004> PMID: 32926914
99. Shiota MN, Keltner D, Mossman A. The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cogn Emot*. 2007; 21: 944–963. <https://doi.org/10.1080/02699930600923668>
100. Silvia PJ, Fayn K, Nusbaum EC, Beaty RE. Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychol Aesthet Creat Arts*. 2015; 9. <https://doi.org/10.1037/aca0000028>
101. Keltner D, Haidt J. Approaching awe, a moral, spiritual, and aesthetic emotion. *Cogn Emot*. 2003; 17: 297–314. <https://doi.org/10.1080/02699930302297> PMID: 29715721
102. Fingerhut J, Prinz JJ. Wonder, appreciation, and the value of art. 2018. pp. 107–128. <https://doi.org/10.1016/bs.pbr.2018.03.004> PMID: 29779731
103. Hekkert P, Wieringen PCW. Complexity and prototypicality as determinants of the appraisal of cubist paintings. *British Journal of Psychology*. 1990; 81: 483–495. <https://doi.org/10.1111/j.2044-8295.1990.tb02374.x>
104. Brinkmann H, Boddy J, Immelmann B, Specker E, Pelowski M, Leder H, et al. Ferocious colors and peaceful lines describing and measuring aesthetic effects. *Wiener Jahrbuch für Kunstgeschichte*. 2018; 65: 7–26. <https://doi.org/10.7767/wjk.2018.65.1.7>

105. Adams FM, Osgood CE. A Cross-Cultural Study of the Affective Meanings of Color. *J Cross Cult Psychol.* 1973; 4. <https://doi.org/10.1177/002202217300400201>
106. Kirk U, Skov M, Christensen MS, Nygaard N. Brain correlates of aesthetic expertise: A parametric fMRI study. *Brain Cogn.* 2009; 69. <https://doi.org/10.1016/j.bandc.2008.08.004> PMID: 18783864
107. Leder H, Gerger G, Brieber D, Schwarz N. What makes an art expert? Emotion and evaluation in art appreciation. *Cogn Emot.* 2014; 28: 1137–1147. <https://doi.org/10.1080/02699931.2013.870132> PMID: 24383619
108. Haanstra F, Damen ML, Van Hoorn M. Interestingness and pleasingness of drawings from different age and expertise groups. *Empirical Studies of the Arts.* 2013; 31: 173–194. <https://doi.org/10.2190/EM.31.2.EOV.2>
109. Iigaya K, Fonseca MS, Murakami M, Mainen ZF, Dayan P. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nat Commun.* 2018; 9. <https://doi.org/10.1038/s41467-018-04840-2> PMID: 29946069
110. Bachmann T, Vipper K. Perceptual rating of paintings from different artistic styles as a function of semantic differential scales and exposure time. *Archiv fuer Psychologie.* 1983; 135: 149–161. PMID: 6670931
111. Barron F, Welsh GS. Artistic Perception as a Possible Factor in Personality Style: Its Measurement By a Figure Preference Test. *J Psychol.* 1952; 33. <https://doi.org/10.1080/00223980.1952.9712830>
112. Barron F. Complexity-Simplicity as a Personality Dimension. *The Journal of Abnormal and Social Psychology.* 1953; 48: 163–172. <https://doi.org/10.1037/h0054907> PMID: 13052336
113. Rosen JC. The Barron-Welsh Art Scale as a Predictor of Originality and Level of Ability Among Artists. *The Journal of Applied Psychology.* 1955; 39: 366–367. <https://doi.org/10.1037/h0042340>
114. Berlyne DE. Dimensions of perception of exotic and pre-Renaissance paintings. *Canadian Journal of Psychology/Revue canadienne de psychologie.* 1975; 29. <https://doi.org/10.1037/h0101682> PMID: 1182607
115. Biaggio MK, Supplee KA. Dimensions of aesthetic perception. *Journal of Psychology: Interdisciplinary and Applied.* 1983; 114: 29–35. <https://doi.org/10.1080/00223980.1983.9915392>
116. Eysenck HJ. The General Factor in Aesthetic Judgements. *British Journal of Psychology General Section.* 1940; 31: 94–102. <https://doi.org/10.1111/j.2044-8295.1940.tb00977.x>
117. Eysenck HJ. “Type” Factors in Aesthetic Judgements. *British Journal of Psychology General Section.* 1941; 31: 262–270. <https://doi.org/10.1111/j.2044-8295.1941.tb00992.x>
118. Hager M, Hagemann D, Danner D, Schankin A. Assessing aesthetic appreciation of visual artworks—The construction of the Art Reception Survey (ARS). *Psychol Aesthet Creat Arts.* 2012; 6. <https://doi.org/10.1037/a0028776>
119. Hagtvædt H, Patrick VM, Hagtvædt R. The Perception and Evaluation of Visual Art. *Empirical Studies of the Arts.* 2008; 26: 197–218. <https://doi.org/10.2190/em.26.2.d>
120. Hawley-Dolan A, Young L. Whose mind matters more—the agent or the artist? An investigation of ethical and aesthetic evaluations. *PLoS One.* 2013; 8. <https://doi.org/10.1371/journal.pone.0070759> PMID: 24039707
121. Kim S, Burr D, Alais D. Attraction to the recent past in aesthetic judgments: A positive serial dependence for rating artwork. *J Vis.* 2019; 19. <https://doi.org/10.1167/19.12.19> PMID: 31627213
122. Miller CA, Hübner R. Two Routes to Aesthetic Preference, One Route to Aesthetic Inference. *Psychol Aesthet Creat Arts.* 2019. <https://doi.org/10.1037/aca0000241>
123. O’Hare DPA, Gordon IE. Dimensions of the perception of art: verbal scales and similarity judgements. *Scand J Psychol.* 1977; 18. <https://doi.org/10.1111/j.1467-9450.1977.tb00257.x> PMID: 841288
124. Ortner KS. *Die emotionale Wirkung moderner Kunst.* University of Vienna. 2010.
125. Panagl Mariella. *Moderne Kunst als Auslöser von Emotion.* University of Vienna. 2011.
126. Silvia PJ. Cognitive Appraisals and Interest in Visual Art: Exploring an Appraisal Theory of Aesthetic Emotions. *Empirical Studies of the Arts.* 2005; 23. <https://doi.org/10.2190/12AV-AH2P-MCEH-289E>
127. Silvia PJ. Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology.* 2005; 9: 342–357. <https://doi.org/10.1037/1089-2680.9.4.342>
128. Specker E, Fried EI, Rosenberg R, Leder H. Associating With Art: A Network Model of Aesthetic Effects. *Collabra Psychol.* 2021; 7. <https://doi.org/10.1525/collabra.24085>
129. Tröndle M, Tschacher W. The physiology of phenomenology: The effects of artworks. *Empirical Studies of the Arts.* 2012; 30: 75–113. <https://doi.org/10.2190/EM.30.1.g>

130. Tschacher W, Greenwood S, Kirchberg V, Wintzerith S, van den Berg K, Tröndle M. Physiological correlates of aesthetic perception of artworks in a museum. *Psychol Aesthet Creat Arts*. 2012; 6. <https://doi.org/10.1037/a0023845>
131. Van Paasschen J, Bacci F, Melcher DP. The influence of art expertise and training on emotion and preference ratings for representational and abstract artworks. *PLoS One*. 2015; 10: 1–21. <https://doi.org/10.1371/journal.pone.0134241> PMID: 26244368
132. Wanzer DL, Finley KP, Zarian S, Cortez N. Experiencing Flow While Viewing Art: Development of the Aesthetic Experience Questionnaire. *Psychol Aesthet Creat Arts*. 2018. <https://doi.org/10.1037/aca0000203>
133. Bahrami-Ehsan H, Mohammadi-Zarghan S, Atari M. Aesthetic Judgment Style: Conceptualization and Scale Development. *International Journal of Arts*. 2015; 5: 33–39. <https://doi.org/10.5923/j.arts.20150502.01>
134. Afhami R, Mohammadi-Zarghan S. The big five, aesthetic judgment styles, and art interest. *Eur J Psychol*. 2018; 14: 764–775. <https://doi.org/10.5964/ejop.v14i4.1479> PMID: 30555584
135. Corradi G, Chuquichambi EG, Barrada JR, Clemente A, Nadal M. A new conception of visual aesthetic sensitivity. *British Journal of Psychology*. 2020; 111: 630–658. <https://doi.org/10.1111/bjop.12427> PMID: 31587262
136. Chamorro-Premuzic T, Furnham A. Art Judgment: A Measure Related to Both Personality and Intelligence? *Imagin Cogn Pers*. 2004; 24: 3–24. <https://doi.org/10.2190/u4lw-th9x-80m3-nj54>
137. Csikszentmihalyi M, Robison RE. *The art of seeing: An interpretation of the aesthetic encounter*. Getty Publications; 1990.
138. Cupchik GC. Emotion in aesthetics: Reactive and reflective models. *Poetics*. 1994; 23: 177–188. [https://doi.org/10.1016/0304-422X\(94\)00014-W](https://doi.org/10.1016/0304-422X(94)00014-W)
139. Cupchik GC. Aesthetics and Emotion in Entertainment Media. *Media Psychol*. 2001; 69–89.
140. Cupchik GC, Hilscher MC. *Holistic Perspectives on the Design of Experience*. 2008.
141. Ekárt A, Joó A, Sharma D, Chalakov S. Modelling the underlying principles of human aesthetic preference in evolutionary art. *Journal of Mathematics and the Arts* output *Journal of Mathematics and the Arts*. 2012; 10: 1–21. <https://doi.org/10.1080/1751347YYxxxxxxx>
142. Israeli N. Affective reactions to painting reproductions: A study in the psychology of esthetics. *Journal of Applied Psychology*. 1928;12. <https://doi.org/10.1037/h0070445>
143. Lundy D, Schenkel M, Akrie T, Walker A. How important is beauty to you? The development of the desire for aesthetics scale. *Empirical Studies of the Arts*. 2010; 28: 73–92. <https://doi.org/10.2190/EM.28.1.e>
144. Experience Marković S. and the emotional content of paintings. *Psihologija*. 2010; 43: 47–64. <https://doi.org/10.2298/PSI1001047M>
145. Marković S, Radonjić A. Implicit and explicit features of paintings. *Spat Vis*. 2008; 21: 229–259. <https://doi.org/10.1163/156856808784532563> PMID: 18534101
146. Menninghaus W, Wagner V, Hanich J, Wassiliwizky E, Kuehnast M, Jacobsen T. Towards a psychological construct of being moved. *PLoS One*. 2015;10. <https://doi.org/10.1371/journal.pone.0128451> PMID: 26042816
147. Murray N, Marchesotti L, Perronnin F. AVA: A Large-Scale Database for Aesthetic Visual Analysis. *IEEE Conference on Computer Vision and Pattern Recognition*. 2012. pp. 2408–2415. Available: [www.dpchallenge.com](http://www.dpchallenge.com)
148. Lu X, Lin Z, Jin H, Yang J, Wang JZ. Rapid: Rating pictorial aesthetics using deep learning. *MM 2014—Proceedings of the 2014 ACM Conference on Multimedia*. Association for Computing Machinery, Inc; 2014. pp. 457–466. <https://doi.org/10.1145/2647868.2654927>
149. Lu X, Lin Z, Jin H, Yang J, Wang JZ. Rating Image Aesthetics Using Deep Learning. *IEEE Trans Multimedia*. 2015; 17: 2021–2034. <https://doi.org/10.1109/TMM.2015.2477040>
150. Osgood CE. The nature and measurement of meaning. *Psychol Bull*. 1952;49. <https://doi.org/10.1037/h0055737> PMID: 14930159
151. Osgood CE, Suci GJ, Tannenbaum PH. *The measurement of meaning*. University of Illinois press; 1957.
152. Pelowski M. Tears and transformation: feeling like crying as an indicator of insightful or “aesthetic” experience with art. *Front Psychol*. 2015;6. <https://doi.org/10.3389/fpsyg.2015.01006> PMID: 26257671
153. Pelowski M, Akiba F. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas Psychol*. 2011; 29: 80–97. <https://doi.org/10.1016/j.newideapsych.2010.04.001>
154. Rowold J. Instrument Development for Esthetic Perception Assessment. *J Media Psychol*. 2008; 20: 35–40. <https://doi.org/10.1027/1864-1105.20.1.35>

155. Silvia PJ, Nusbaum EC. On personality and piloerection: Individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychol Aesthet Creat Arts*. 2011;5. <https://doi.org/10.1037/a0021914>
156. Smith JK, Smith LF. The Nature and Growth of Aesthetic Fluency. *New directions in aesthetics, creativity, and the arts*. 2006; 47–58. Available: <https://www.researchgate.net/publication/232595805>
157. Silvia PJ. Knowledge-Based Assessment of Expertise in the Arts: Exploring Aesthetic Fluency. *Psychol Aesthet Creat Arts*. 2007; 1: 247–249. <https://doi.org/10.1037/1931-3896.1.4.247>
158. Specker E, Forster M, Brinkmann H, Boddy J, Pelowski M, Rosenberg R, et al. The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A Unified and Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and the Arts*. 2018. <https://doi.org/10.1037/aca0000205>
159. Virtanen T, Nuutinen M, Häkkinen J. Underlying Elements of Image Quality Assessment: Preference and Terminology for Communicating Image Quality Characteristics. *Psychol Aesthet Creat Arts*. 2020. <https://doi.org/10.1037/aca0000312>
160. Fechner GT. *Vorschule der Aesthetik*. Leipzig, Germany: Breikopf & Haertel; 1876.
161. Zeki S. Art and the brain. *Journal of Consciousness Studies*. 1999; 6: 76–96.
162. Kirk U, Skov M, Hulme O, Christensen MS, Zeki S. Modulation of aesthetic value by semantic context: An fMRI study. *Neuroimage*. 2009;44. <https://doi.org/10.1016/j.neuroimage.2008.10.009> PMID: 19010423
163. Cela-Conde CJ, Ayala FJ. Brain keys in the appreciation of beauty: A tale of two worlds. *Rendiconti Lincei*. 2014; 25: 277–284. <https://doi.org/10.1007/s12210-014-0299-8>
164. Reber R, Schwarz N, Winkielman P. Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*. 2004;8. [https://doi.org/10.1207/s15327957pspr0804\\_3](https://doi.org/10.1207/s15327957pspr0804_3) PMID: 15582859
165. Vessel EA, Starr GG, Rubin N. Art reaches within: Aesthetic experience, the self and the default mode network. *Frontiers in Neuroscience*. Frontiers Media SA; 2013. <https://doi.org/10.3389/fnins.2013.00258> PMID: 24415994
166. Pelowski M, Leder H, Tinio PPL. Creativity in the Visual Arts. *The Cambridge Handbook of Creativity across Domains*. 2017; 80–109.
167. Campbell DT. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychol Rev*. 1960;67. <https://doi.org/10.1037/h0040373> PMID: 13690223
168. Sternberg RJ. *Handbook of creativity*. Cambridge University Press; 1999.
169. Newman GE, Bloom P. Art and authenticity: The importance of originals in judgments of value. *J Exp Psychol Gen*. 2012;141. <https://doi.org/10.1037/a0026035> PMID: 22082113
170. Lauring JO, Pelowski M, Forster M, Gondan M, Ptito M, Kupers R. Well, if they like it. ... effects of social groups' ratings and price information on the appreciation of art. *Psychol Aesthet Creat Arts*. 2016; 10: 344–359. <https://doi.org/10.1037/aca0000063>
171. Hagtvedt H, Patrick V, Bradshaw A, Hackley C, Maclaran P. The Manner of Art: An Aesthetic Influence on Evaluation. 2011. Available: <http://www.acrwebsite.org/volumes/1006887/eacr/vol9/E-09http://www.copyright.com/>.
172. Leder H. Determinants of Preference: When do we like What we Know? *Empirical Studies of the Arts*. 2001;19. <https://doi.org/10.2190/5TAE-E5CV-XJAL-3885>
173. Ishizu T, Zeki S. The experience of beauty derived from sorrow. *Hum Brain Mapp*. 2017; 38: 4185–4200. <https://doi.org/10.1002/hbm.23657> PMID: 28544456
174. Reber R. Processing fluency, aesthetic pleasure, and culturally shared taste. In: Shimamura AP, Palmer SE, editors. *Aesthetic science: Connecting minds, brains, and experience*. New York: Oxford University Press; 2012. pp. 223–249.
175. Belke B, Leder H, Strobach T, Carbon CC. Cognitive Fluency: High-Level Processing Dynamics in Art Appreciation. *Psychol Aesthet Creat Arts*. 2010; 4: 214–222. <https://doi.org/10.1037/a0019648>
176. Belke B. *Cognitive Mastering phenomena in the appreciation of visual art: Theoretical outlines and experimental probes..* University of Vienna. 2020.
177. Westerman DL, Lanska M, Olds JM. The effect of processing fluency on impressions of familiarity and liking. *J Exp Psychol Learn Mem Cogn*. 2015;41. <https://doi.org/10.1037/a0038356> PMID: 25528088
178. Augustin D, Leder H. Art expertise: A study of concepts and conceptual spaces. *Psychol Sci*. 2006; 48: 135.
179. Hekkert P, van Wieringen PCW. The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychol (Amst)*. 1996; 94: 117–131. [https://doi.org/10.1016/0001-6918\(95\)00055-0](https://doi.org/10.1016/0001-6918(95)00055-0)