# The Role of Artwork Origin and Art Experience in Aesthetic Evaluation

Wouter van Dam

s4346203

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 22

Supervisor: prof. dr. R.F.A. (Ralf) Cox

Second evaluator: dr. rer. nat. Stefanie Enriquez-Geppert

In collaboration with: Maria Bendykowska, Anna Dimitriou, Matúš Farkaš, Isa Nammensma, Wouter van Dam, Yuhan Wang

July 4, 2025

**Abstract**

As artificial intelligence (AI) increasingly contributes to the creation of visual art, questions arise about how the perceived origin of an artwork influences its evaluation. This study investigated whether artworks labelled as AI-generated versus human-made are perceived differently in terms of beauty, creativity, and emotional intensity, and whether these effects are moderated by art experience. A total of 199 participants viewed abstract artworks paired with origin-based descriptions and provided ratings using evaluative scales and the Geneva Emotion Wheel. A repeated-measures design was employed with artwork origin as a within-subjects factor and art experience (low vs. high) as a between-subjects factor. Results showed that participants rated human-made artworks as significantly more beautiful and creative than AI-generated ones, particularly among those with higher art experience. However, no significant effects of artwork origin or art experience were found for emotional intensity. These findings might suggest that aesthetic evaluations are strongly influenced by beliefs about authorship, highlighting the enduring cultural and psychological value ascribed to human agency in art, even in the face of increasingly sophisticated generative AI.

*Keywords:* AI-generated art, perceived authorship, aesthetic evaluation, art expertise, beauty, creativity, emotional response

**The Role of Artwork Origin and Art Experience in Aesthetic Evaluation**

As AI becomes increasingly involved in the creation of visual art, it challenges long-standing ideas about what makes art valuable or meaningful. Over the past decade, AI has developed into a general-purpose technology, expanding beyond specialized tasks and becoming widely integrated across sectors such as healthcare, education, transportation, and the creative industries (Wikipedia contributors, 2025). As these technologies evolve, the implications of AI-generated art extend beyond aesthetics into legal, ethical, and cultural domains. Piskopani et al. (2023) argue that AI in the creative industries raises fundamental questions about authorship, authenticity, and the value we place on human artistic labour. The increasing use of AI tools in competitions, museums, and commercial art production has led to growing concern over devaluation of creative work, intellectual property infringement, and a loss of artistic agency. These concerns highlight how the technological disruption introduced by AI challenges not only economic models but also cultural and emotional attachments to human creativity. These broader debates contextualize why understanding how people perceive AI-generated versus human made art is not only timely but necessary.

While AI-generated artworks can be just as visually compelling as those made by humans, people often still show a strong preference for human-made art. Research by Bellaiche et al. (2023) suggests that this preference is not driven by visual quality alone, but also by assumptions about the artist's intentions and emotional involvement. Human-made art is often believed to carry more meaning and effort, which enhances its perceived beauty and emotional resonance. These findings underscore that art appreciation depends not just on what we see, but also on what we believe about an artwork's origin.

Complementing this view Horton et al. (2023) demonstrate that perceptions of authorship significantly affect how viewers evaluate artworks. In their study, artworks believed to be human-made were rated as more intentional, creative, and emotionally

expressive than identical pieces labelled as AI-generated. This suggests that beliefs about the creator's agency shape not only aesthetic preferences but also the interpretation of emotional and conceptual depth, reinforcing the importance of investigating origin effects.

In parallel, recent work by Piskopani et al. (2023) highlights how AI's growing role in the arts has raised public concerns about authorship, authenticity, and fairness. These broader debates, ranging from copyright and job displacement to the erosion of artistic value, reflect a societal unease about whether AI-generated content can or should be valued as genuine artistic expression. Such concerns underscore the importance of studying how people evaluate AI-generated versus human-made art.

Meijer (2024) further demonstrates that the consequences of authorship perceptions extend beyond subjective appreciation into the economic realm. In her study, participants who knew an artwork was AI-generated were not only less likely to perceive it as creative, but also significantly less willing to pay for it, often offering hundreds of euros less than participants who saw the same image without an AI label. This market resistance underscores how beliefs about authorship influence both the symbolic and material value of art, reinforcing the importance of investigating how origin shapes aesthetic and emotional evaluations.

Interestingly, some technical advances in AI art generation suggest that origin-based preferences may not always reflect objective differences in quality. For instance, Elgammal et al. (2017) demonstrated that art produced by Creative Adversarial Networks was frequently rated by human participants as equally or more aesthetically pleasing than works by professional artists. This indicates that, under certain conditions, AI-generated images can match or even surpass human-made art in perceived creativity and appeal. Nevertheless, such findings contrast with the consistent evidence that people often rate identical artworks lower when told they are AI-generated. This discrepancy highlights a key issue: even if AI art

achieves visual parity, the belief about who or what created the piece may still drive evaluation.

Importantly, individual differences in art expertise may influence how strongly people respond to authorship cues. van Hees et al. (2025) found that participants with greater exposure to art were more sensitive to contextual framing when judging visual works, suggesting that expertise amplifies the effect of knowing an artwork's origin. Similarly, Schino et al. (2025) emphasize that art perception is not purely visual but shaped by both emotional and interpretive processes. These findings justify including *emotional intensity* alongside beauty and creativity in the present study, as well as testing whether the impact of artwork origin is moderated by the viewer's level of art experience.

Taken together, these findings raise a central question within psychological and aesthetic research: *How does knowledge of an artwork's origin, whether AI-generated or human-made, influence the way it is perceived in terms of beauty, creativity, and emotional intensity? Moreover, how does this effect vary depending on the viewer's level of art experience?*

Based on previous studies suggesting that people evaluate AI-generated art less favourably than human-made art even when visual quality is held constant (e.g., Bellaiche et al., 2023; Horton et al., 2023), the first hypothesis (H1) predicts that artworks labelled as human-made will receive higher ratings for beauty, creativity, and emotional intensity than those labelled as AI-generated.

The second hypothesis (H2) builds on findings that art expertise may enhance sensitivity to contextual cues such as authorship (van Hees et al., 2025). Specifically, it is expected that the difference in evaluation between human and AI artworks will be more pronounced for participants with high art experience than for those with less experience. This is especially relevant for emotional intensity, as Schino et al. (2025) emphasize that affective

responses to art are shaped not only by visual features but also by personal framing and interpretive context.

To test these hypotheses, participants viewed a set of abstract artworks labelled as either AI-generated or human-made and provided ratings of beauty, creativity, and emotional intensity.

## Methods

### Participants

In total, 258 people participated in the study. From these, 43 participants were excluded because of not completing the survey or giving consent. 16 participants were excluded for not passing the control question designed to check whether they are paying attention. The removal resulted in a final 199 participants included in the analysis (143 female; 47 male; two non-binary; seven preferred not to say/self - described). Most participants ($n$ = 157; 79%) were in the age group 18-24 years old.

Participants took part either in exchange for course credits or voluntarily without reward. Convenience and snowball sampling was used - the link for this study had been sent around in online group-chats, to friends and acquaintances of the researchers, encouraging people to share the link further. The study had also been posted on the SONA study system of the University of Groningen.

The research was approved and conducted in accordance with the ethical codes and regulations of the Faculty of Behavioural and Social Sciences at the University of Groningen (registration code: PSY-2425-S-0337).

### Materials

The study ran on Qualtrics (https://www.qualtrics.com/) and was available in English and Dutch. Participants completed the survey on their own devices without hardware restrictions.

### *Artworks*

From the artworks used, six were human-made, and six AI-generated. See Appendix A for all artworks. Only abstract artworks were selected for reasons discussed in the introduction. Abstract was defined as no figurative elements present. Moreover, images were selected so that human-made and AI-generated artworks were pairwise similar in colour, composition or dynamic.

### *Artwork Descriptions*

Every artwork was preceded by a short textual description that framed the work either neutrally, positively, or negatively and mentioned origin. The descriptions were created by the research team specifically for this paper. There were two AI-generated positive, two AI-generated negative, two AI-generated negative, two human-made positive, two human-made negative and two human-made neutral descriptions. There was no deception on the origin of artworks. The framing was not based on actual facts about the artworks but was made-up. For a list of all descriptions, as well as their translations into Dutch, see Appendix B.

### *Intentionality, Beauty and Creativity*

Participants answered three evaluative statements about intentionality, creativity and beauty using sliders on a scale ranging from 0 to 100, based on Cox et al. 2024. For 'beauty' the statement "I find this work beautiful" was presented. A score of 0 represented "strongly disagree", and a score of 100 represented "strongly agree". For 'intentionality' and 'creativity' the statements "In my opinion, the level of intentionality involved in the creation of this work is…" and "In my opinion, the level of creativity involved in the creation of this work is…" were presented, respectively. A score of 0 indicated "very low" and a score of 100 indicated "very high". For the translations of these statements, see Appendix C.

### *Geneva Emotion Wheel*

Participants were then presented with the Geneva Emotion Wheel (GEW), a tool to label emotions and record emotional intensity and valence in surveys (Scherer, 2005). Participants were asked to select the emotion-label that best captured their emotional response to the artwork, as well as indicate the intensity of that emotion, within one click. They were, for example, able to choose between emotions such as ‘Anger’, ‘Surprise’, ‘Fear’ or ‘Joy’. If no emotion was experienced, they could select the “None” option, and if their emotion was not represented in the wheel, they could select “Other”. Cronbach's α is not specified. An image of the GEW from our study, along with all the emotions and their translations, can be found in Appendix D.

***Aesthetic Fluency Scale***

The aesthetic fluency scale was used as an approximate measure for art knowledge (Cotter et al., 2023). Due to time constraints for the survey, the short version of the questionnaire was used. Participants' familiarity with 10 artists and art-related terms was assessed. Question 11 was a control question, used as an exclusion criterion for analysis. The questionnaire presents a term, for example “Gouache” and gives three response options - “I don’t really know anything about this artist or term”, “I’m familiar with this artist or term”, and “I know a lot about this artist or term”. Cronbach's α was 0.84. For the translations of the response options, see Appendix C.

***General Attitudes toward Artificial Intelligence Scale***

Participants completed an attitude scale assessing their attitude of AI, consisting of five statements adapted from the General Attitudes toward Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2020). “Artificial Intelligence is exciting”, “I am impressed by what Artificial Intelligence can do”, “There are many beneficial applications of Artificial Intelligence”, “I am interested in using artificially intelligent systems in my daily life”, and “Artificial Intelligence can have positive impacts on people's wellbeing” were answered on

sliders ranging from 0 to 100 in steps of 10, from strongly disagree to strongly agree. Cronbach's α was 0.89. For the translations of the response options, see Appendix C.

**Procedure**

After accessing the study environment, participants were informed about the study and its procedure, and then signed an informed consent. Next, participants were asked about age, having five options - 18-24, 25-34, 35-44, 45-54, or 55+ years old, and about gender, also having five options - male, female, non-binary, prefer to self describe, or prefer not to say.

The respondents were distributed equally among six experimental groups. Each group viewed the same 12 artworks and read the same descriptions, but the description–artwork pairings varied across groups, as shown in Table 1. The sequence of description-artwork presentation was randomised per participant.

First, participants evaluated the 12 artworks. Each artwork was preceded by one of the descriptions. There was no time limit to seeing the description. Participants had to click "next" to proceed to the next screen showing an artwork. On this screen descriptions were not visible anymore.

Each artwork appeared for a minimum of five seconds. After that the participant could choose freely when to continue, by clicking "next". The time spent looking at each artwork, between first appearance of the artwork to clicking "next", was measured for every participant, later used as the looking time variable.

In the next step the artwork was not visible anymore. Participants answered the three evaluative statements concerning intentionality, creativity and beauty and continued. On the same page they used the GEW.

After all artworks had been evaluated, participants filled out the two questionnaires. First, the 11-item Aesthetic Fluency scale was presented (Cotter et al., 2023). On the next

page, participants completed the General Attitudes toward Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2020).

The study concluded with a debriefing screen that explained the manipulation of the framings, as well as the looking-time measurements, and informed all participants that they were not deceived of the true origin of each artwork. Participants were thanked for their participation and, if applicable, directed to collect their course credits through the university's SONA system.

**Table 1**

*Artwork x Description Pairings per Group for Human-made and AI-generated artworks*

| Group | Positive Description | | Negative Description | | Neutral Description | |
|---|---|---|---|---|---|---|
| | D_1 | D_2 | D_3 | D_4 | D_5 | D_6 |
| **1** | 1 | 2 | 3 | 4 | 5 | 6 |
| **2** | 6 | 1 | 2 | 3 | 4 | 5 |
| **3** | 5 | 6 | 1 | 2 | 3 | 4 |
| **4** | 4 | 5 | 6 | 1 | 2 | 3 |
| **5** | 3 | 4 | 5 | 6 | 1 | 2 |
| **6** | 2 | 3 | 4 | 5 | 6 | 1 |

*Note:* The numbers under description are representative of the artworks assigned in that group, see Appendix A. D_1, etc. are the human and AI descriptions, see Appendix B.

**Data Analysis**

After data collection, the dataset was cleaned by removing participants who did not complete the survey, did not give consent, or failed the attention check question. This resulted in a final sample of 199 participants, as detailed in the Participants section. The mean completion time of the study was 2,122 seconds (35.4 minutes), while the median time was 726 seconds (12.1 minutes), indicating a positively skewed distribution. This skewness likely reflects that some participants did not complete the survey in a single sitting. The researchers

had pre-registered an exclusion criterion for durations under 300 seconds (5 minutes); however, no cases met this threshold, as the lowest completion time was 302 seconds. All analyses were conducted using JASP version 0.19.3.0 (*JASP - a Fresh Way to Do Statistics*, 2025). Because the scale scores for art experience were not normally distributed (Shapiro-Wilk = 0.927, $p < .001$), and a grouped comparison was necessary for the planned mixed Analysis of Variance (ANOVA), a median split was applied post hoc. The non-normal distribution of the scores is also visually evident in the histogram (Appendix E, Figure E1). This approach, though not initially intended, enabled a clear between-subjects contrast suitable for the analytical design. To analyse the data, repeated-measures ANOVAs were conducted for each of the three dependent variables (beauty, creativity, and emotional intensity). The within-subjects factor was artwork origin (AI vs. human), and the between-subjects factor was art experience group (low vs. high, based on the median split).

Ratings for beauty and creativity were collected using sliders ranging from 0 to 100. For analyses of beauty and creativity, the full sample of 199 participants was used. Based on the median split of art experience scores, 94 participants were classified in the high art experience group and 105 in the low art experience group. Emotional responses were measured using the GEW (Scherer, 2005). Emotional intensity scores were calculated from GEW data using Euclidean distance. Specifically, intensity was computed as the Euclidean distance from the participant's click on the GEW to the centre of the wheel (i.e., the origin of the coordinate system), using the formula $\sqrt{(x^2 + y^2)}$. Clicks within a radius of 75 pixels from the centre were excluded from the analysis, as they corresponded to the "None" and "Other" response options rather than specific emotion labels. Valence, although not analysed in the current study, was computed as the horizontal distance from the participant's click to the vertical axis passing through the GEW's centre. Only participants with non-missing emotional

intensity ratings for both human and AI artworks were included in the analysis, resulting in a final sample of 91 participants ($n$ = 47 high art experience, $n$ = 44 low art experience).

## Results

### Descriptive Statistics and Assumption Checks

Table 2 presents the descriptive statistics for beauty and creativity ratings, and Table 3 shows the descriptive statistics for emotional intensity, each broken down by artwork origin (AI vs. human) and art experience group (low vs. high). The Aesthetic Fluency Scale demonstrated high internal consistency, with Cronbach's α of .875, 95% CI [.837, .902]. These values indicate that the scale was a reliable measure of participants' self-reported art knowledge. Q-Q plots of residuals revealed minor deviations from normality across all three dependent variables (see Figures E2–E4 in Appendix E), but these deviations were not substantial enough to warrant concern. Because the art experience variable was not normally distributed (Shapiro–Wilk = 0.927, $p < .001$).

Levene's test for equality of variances showed mixed results. For beauty, variances were equal for human artwork ratings, $F(1, 197) = 1.62$, $p = .205$, but significantly unequal for AI artwork ratings, $F(1, 197) = 13.60$, $p < .001$, with higher variability observed in the high experience group. For creativity, the assumption of homogeneity of variance was met in both conditions: human artwork ratings, $F(1, 197) = 0.615$, $p = .434$, and AI artwork ratings, $F(1, 197) = 0.916$, $p = .340$, showed no significant differences in variance. In contrast, for emotional intensity, Levene's test indicated unequal variances for both human artwork ratings, $F(1, 89) = 4.500$, $p = .037$, and AI artwork ratings, $F(1, 89) = 9.945$, $p = .002$.

Despite these deviations, repeated-measures ANOVA is generally robust to minor violations of normality and homogeneity of variance, particularly in large and balanced samples (Blanca et al., 2017; Field, 2018; Schmider et al., 2010). Therefore, all planned analyses were retained.

**Table 3**

*Descriptive Statistics for Beauty, Creativity and Emotional Intensity Ratings by Artwork Origin and Art Experience Group*

| Variable | Origin | Art Experience | N | M | SD | SE | CV |
|---|---|---|---|---|---|---|---|
| Beauty | HUM | High | 94 | 35.915 | 21.251 | 2.192 | 0.592 |
| | | Low | 105 | 31.176 | 19.052 | 1.859 | 0.611 |
| | AI | High | 94 | 40.213 | 23.933 | 2.469 | 0.595 |
| | | Low | 105 | 49.062 | 17.650 | 1.722 | 0.360 |
| Creativity | HUM | High | 94 | 49.793 | 22.212 | 2.291 | 0.446 |
| | | Low | 105 | 48.267 | 23.318 | 2.276 | 0.483 |
| | AI | High | 94 | 30.463 | 22.279 | 2.298 | 0.731 |
| | | Low | 105 | 36.738 | 25.041 | 2.444 | 0.682 |
| EM | HUM | High | 47 | 179.034 | 54.389 | 7.933 | 0.304 |
| | | Low | 44 | 194.078 | 69.131 | 10.422 | 0.356 |
| | AI | High | 47 | 181.038 | 50.453 | 7.359 | 0.279 |
| | | Low | 44 | 203.673 | 75.722 | 11.416 | 0.372 |

*Note.* HUM = Human-made artwork; AI = AI–generated artwork. Beauty and Creativity ratings were recorded on a scale from 0 (*not at all*) to 100 (*extremely*).

**Table 2**

*Descriptive Statistics for Emotional Intensity Ratings by Artwork Origin and Art Experience Group*

| Variable | Origin | Art Experience | N | M | SD | SE | CV |
|---|---|---|---|---|---|---|---|
| EI | HUM | High | 47 | 179.034 | 54.389 | 7.933 | 0.304 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Low | 44 | 194.078 | 69.131 | 10.422 | 0.356 |
| AI | High | 47 | 181.038 | 50.453 | 7.359 | 0.279 |
| | Low | 44 | 203.673 | 75.722 | 11.416 | 0.372 |

*Note:* EI = Emotional Intensity. Emotional Intensity scores ranged from 0 to 300 and were assessed in a subset of participants (N = 91).

**Beauty Ratings**

To test (H1), which predicted that human-made artworks would be rated as more beautiful than AI-generated artworks, a repeated-measures ANOVA was conducted with artwork origin (AI vs. human) as a within-subjects factor and art experience group (low vs. high) as a between-subjects factor. Table 2 presents the descriptive statistics for beauty ratings across conditions.

Although AI-generated artworks received higher average ratings overall, largely due to the higher scores given by participants in the low art experience group, the repeated-measures ANOVA revealed a significant main effect of artwork origin, $F(1, 197) = 50.997$, $p < .001$, $\eta^2 = 0.067$. This indicates that, when comparing within individuals, participants tended to rate human-made artworks as more beautiful than AI-generated ones.

A significant interaction was found between artwork origin and art experience, $F(1, 197) = 19.133$, $p < .001$, $\eta^2 = 0.025$, indicating that the effect of origin on beauty ratings differed depending on participants' level of art experience. However, the main effect of art experience itself was not significant, $F(1, 197) = 0.698$, $p = .404$, $\eta^2 = 0.002$.

**Creativity Ratings**

To further test H1, creativity ratings were analysed in the same 2 (origin) × 2 (art experience) repeated-measures ANOVA framework. Table 2 presents the descriptive statistics for creativity ratings across artwork origin (AI vs. human) and art experience group (low vs. high). Human-made artworks were rated as more creative overall. There was a significant

main effect of artwork origin, $F(1, 197) = 68.163$, $p < .001$, $\eta^2 = 0.099$, with human-made artworks rated as more creative than AI-generated ones. A significant interaction was also found between artwork origin and art experience, $F(1, 197) = 4.356$, $p = .038$, $\eta^2 = 0.006$. The main effect of art experience was not significant, $F(1, 197) = 0.757$, $p = .385$, $\eta^2 = 0.002$.

### Emotional Intensity

To test H2, which proposed that emotional intensity ratings would vary more strongly by artwork origin for individuals with high art experience, a repeated-measures ANOVA was again used. Table 3 presents the descriptive statistics for emotional intensity ratings across artwork origin (AI vs. human) and art experience group (low vs. high). Participants in the low experience group showed slightly higher intensity ratings for both types of artworks. No significant main effect of artwork origin was found, $F(1, 89) = 2.94$, $p = .090$, $\eta^2 = 0.002$. There was no significant interaction between origin and art experience, $F(1, 89) = 1.26$, $p = .265$, $\eta^2 < .001$. The main effect of art experience was also not significant, $F(1, 89) = 2.18$, $p = .143$, $\eta^2 = 0.022$.

## Discussion

This study examined whether knowledge of an artwork's origin (AI vs. human) influences its perceived beauty, creativity, and emotional intensity, and whether this effect varies by participants' level of art experience. The results partially supported the hypotheses: Participants rated human-made artworks as significantly more beautiful and more creative than AI-generated ones (supporting H1), and this difference was amplified among those with higher art experience. However, emotional intensity ratings did not significantly differ by artwork origin, and no interaction effect was found with art experience, providing no support for H2 with respect to emotional intensity.

### Interpretation and Theoretical Implications

The observed preference for human-made artworks in beauty and creativity ratings is in line with a growing body of research showing that perceptions of authorship shape aesthetic evaluations. As shown by Bellaiche et al. (2023) and Horton et al. (2023), artworks believed to be created by humans tend to be rated as more intentional, expressive, and creative, even when visually identical to AI-generated pieces. This suggests that perceived agency plays a central role in aesthetic judgments, extending beyond the image itself.

Our results further indicate that this origin effect is magnified in individuals with higher art experience, consistent with findings by van Hees et al. (2025). These participants may be more sensitive to authorship cues or more invested in the traditional values of artistic intention, originality, and human expression.

Although AI-generated artworks received higher mean beauty scores overall (mainly due to the low art experience group), human-made artworks were rated as more beautiful in the within-subjects comparison, reinforcing the psychological weight of authorship beliefs in aesthetic experience.

The absence of significant effects for emotional intensity contrasts with these patterns. This may suggest that emotional engagement with art is less dependent on beliefs about authorship, or that the Geneva Emotion Wheel's measurement approach primarily captures stronger affective responses, possibly omitting subtler emotional nuances. Additionally, the smaller sample size for emotional intensity ($n$ = 91) may have reduced statistical power.

Interestingly, although participants in this study consistently rated human-made artworks more favourably, this does not necessarily imply that AI-generated art lacks aesthetic merit. Elgammal et al. (2017) found that artworks produced by Creative Adversarial Networks were often rated as equally or more appealing than those by professional artists. This suggests that AI is capable of producing high-quality art, and that negative evaluations may stem more from bias than objective difference.

This bias may also manifest in economic contexts. Meijer (2024) found that artworks labelled as AI-generated were perceived as less creative and were assigned significantly lower monetary value than identical works without such labelling. These findings highlight that assumptions about authorship affect not only aesthetic judgment but also market valuation. Together with Piskopani et al. (2023), who emphasize public unease around AI's encroachment into creative fields, these results underscore that preferences for human-made art are grounded not just in perception, but also in deeper cultural, ethical, and economic narratives.

**Limitations and Future Directions**

Several limitations of the current study warrant consideration. First, the decision to apply a median split to the art experience variable, necessitated by the non-normal distribution of the scale scores, may have led to a loss of statistical power and masked meaningful variability among participants (MacCallum et al., 2002). While this approach was analytically useful for the planned mixed ANOVA, future studies could adopt continuous modelling strategies (e.g., ANCOVA or linear mixed models) to better capture gradations in art expertise, as suggested by van Hees et al. (2025), who found that varying levels of art experience differentially affect sensitivity to contextual cues like authorship.

Second, although participants were randomly assigned to one of six groups that varied the artwork-description pairings, the design did not fully counterbalance all combinations of artworks and descriptions. This partial randomization leaves open the possibility that certain artworks may have been systematically paired with particular framings more often in some groups than others, potentially introducing subtle stimulus effects. However, the randomized sequence per participant mitigates the likelihood of major bias.

Third, although the Geneva Emotion Wheel (GEW) allows for a broad mapping of emotional responses, emotional intensity could only be computed when participants selected a

specific emotion label. Responses such as “None” or “Other” did not yield coordinate data and were thus excluded. This led to missing values for a subset of participants, reducing the usable sample size for emotional intensity analyses ($n = 91$). While this exclusion was not due to low intensity per se, it limited statistical power and may have biased the results toward those with more defined emotional reactions. Moreover, Scherer (2005) himself noted that the GEW is better suited for identifying broad emotional patterns than for detecting subtle or low-intensity affective shifts. Future research could combine the GEW with complementary methods, such as verbal descriptions or dimensional rating scales, to better capture nuanced emotional responses across participants.

Finally, although repeated-measures ANOVA is robust to minor violations of assumptions (Blanca et al., 2017; Schmider et al., 2010), small deviations from normality and unequal variances were present in several conditions. While these did not compromise the main analyses, their presence should be kept in mind when interpreting marginal effects.
In light of these findings, future research could explore whether the observed authorship bias persists when participants are presented with artworks of greater emotional or narrative richness, as affective engagement may have been limited by the use of abstract stimuli. Additionally, building on the economic valuation effects found by Meijer (2024), subsequent studies could further examine how aesthetic judgments interact with perceived monetary value, particularly in contexts where authorship is explicitly known versus ambiguously framed.

**Conclusion**

This study examined whether knowledge of an artwork’s origin influences its perceived beauty, creativity, and emotional intensity, and whether these effects vary by level of art experience. Although AI-generated artworks received higher average beauty ratings overall, primarily due to scores from the low art experience group, the within-subjects

analysis revealed that participants, on average, rated human-made artworks as more beautiful than AI-generated ones. Human-made artworks were also rated as more creative, with this effect being stronger among individuals with higher art experience. No significant differences were found for emotional intensity ratings. These results suggest that artwork origin and art experience are associated with differences in how participants evaluate beauty and creativity, while emotional engagement appeared unaffected under the current conditions. Further research is needed to explore the underlying mechanisms and contextual factors that shape these evaluative patterns.

## References

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research Principles and Implications*, *8*(1). https://doi.org/10.1186/s41235-023-00499-6

Blanca, M., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *4*(29), 552–557. https://doi.org/10.7334/psicothema2016.383

Cotter, K. N., Rodriguez-Boerwinkle, R. M., Christensen, A. P., Fekete, A., Smith, J. K., Smith, L. F., Tinio, P. P. L., & Silvia, P. J. (2023). Updating the Aesthetic Fluency Scale: Revised long and short forms for research in the psychology of the arts. *PLoS ONE*, *18*(2), e0281547. https://doi.org/10.1371/journal.pone.0281547

Cox, R. F. A., Schino, G., Van Klaveren, L., & Wittingslow, R. (2024, October 23). *Exploring the Skill-Tech Space: Art-Making Awareness Influences Experiencing Visual Artworks*. https://www.researchgate.net/publication/385186655_Exploring_the_Skill-Tech_Space_Art-Making_Awareness_Influences_Experiencing_Visual_Artworks. https://doi.org/10.31219/osf.io/jxnwc

Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017, June 21). *CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms*. arXiv.org. https://arxiv.org/abs/1706.07068

Field, A. (2017). *Discovering statistics using IBM SPSS Statistics*. SAGE.

Horton, C. B., Jr, White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-45202-3

Meijer, M. (2024). *How AI-generated art affects creativity perception and market value. Rotterdam School of Management.* https://www.rsm.nl/discovery/2024/value-of-ai-generated-art/

*JASP - a fresh way to do statistics. (2025, January 6). JASP - Free and User-Friendly Statistical Software.* https://jasp-stats.org/

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989x.7.1.19

Piskopani, A. M., Chamberlain, A., & Holter, C. T. (2023, July 5). Responsible AI and the Arts: The Ethical and Legal Implications of AI in the Arts and Creative Industries. In *https://www.researchgate.net/*. https://doi.org/10.1145/3597512.3597528

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, *1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Schino, G., Van Klaveren, L., Van Dorsten, T., Pathare, S., Van Heusden, B., & Cox, R. F. A. (2025). Experiencing art together: integrating affect and semiosis. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1544901

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*, *6*(4), 147–151. https://doi.org/10.1027/1614-2241/a000016

Van Hees, J., Grootswagers, T., Quek, G. L., & Varlet, M. (2025). Human perception of art in the age of artificial intelligence. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1497469

Wikipedia contributors. (2025, June 24). *Applications of artificial intelligence*. Wikipedia. https://en.wikipedia.org/wiki/Applications_of_artificial_intelligence

# Appendix A

## Human-made pictures

*Artwork 1*

*Sans titre (Composition brune II) by Youla Chapoval - Artvee*. (o. D.). Artvee.

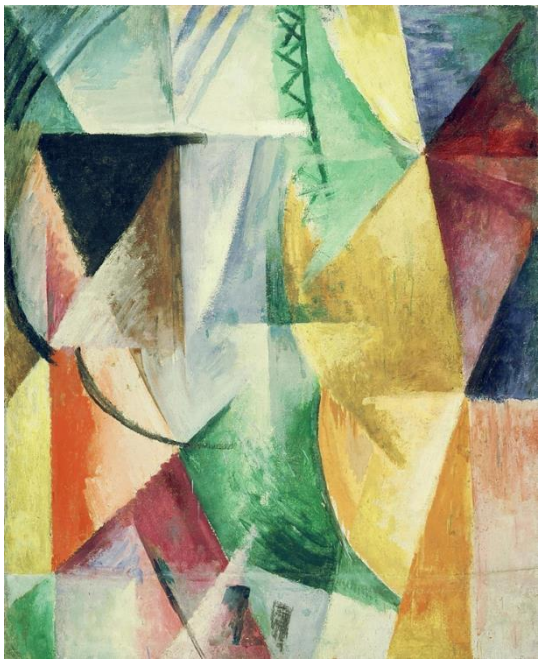
https://artvee.com/dl/sans-titre-composition-brune-ii/#00

*Artwork 2*

*abstract paintings - Abstract paintings Alessandro Tognin*. (2023, 3. September). Abstract Paintings Alessandro Tognin. https://www.dreamsart.it/product/abstract-paintings/

*Artwork 3*

*Une fenêtre by Robert Delaunay - Artvee*. (o. D.). Artvee. https://artvee.com/dl/une-fenetre/#00



*Artwork 4*

*Figürliche Komposition by Adolf Hölzel - Artvee*. (o. D.). Artvee.

https://artvee.com/dl/figurliche-komposition/#00

*Artwork 5*

*Anitra by Anonymous - Artvee*. (o. D.). Artvee. https://artvee.com/dl/anitra/#00

*Artwork 6*

*Komposition by Otto Freundlich - Artvee*. (o. D.). Artvee. https://artvee.com/dl/komposition-14/#0

**AI Artworks**

*Artwork 1*

*Exploring Abstract Art with AI*. (2024, 24. Juli).

https://deepdreamgenerator.com/blog/abstract-art-and-ai

*Artwork 2*

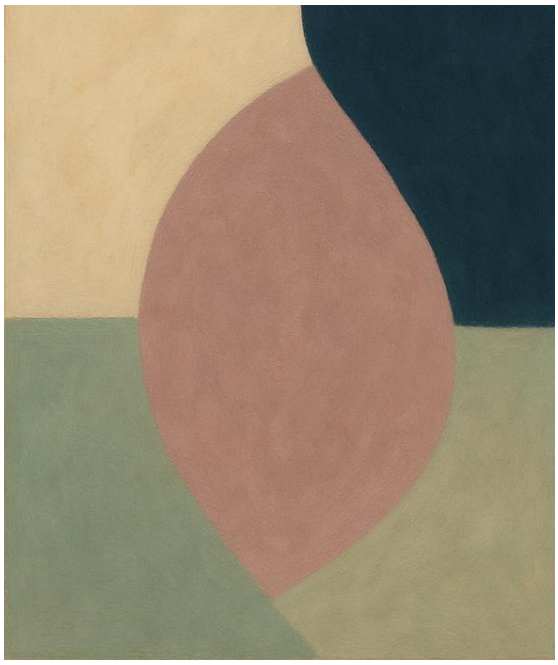
*Abstrakte Erdelemente auf Leinwand*. (2025, 25. April). ChatGPT.

https://chatgpt.com/s/m_680b937d8b548191960f2c69fc085d2b

*Artwork 3*

*Geometrische Abstraktion in Pastellfarben*. (2025, 25. April). ChatGPT.

https://chatgpt.com/s/m_680b9334cf0c819189fd5f1b73c92c39



*Artwork 4*

*Exploring Abstract Art with AI*. (2024, 24. Juli).

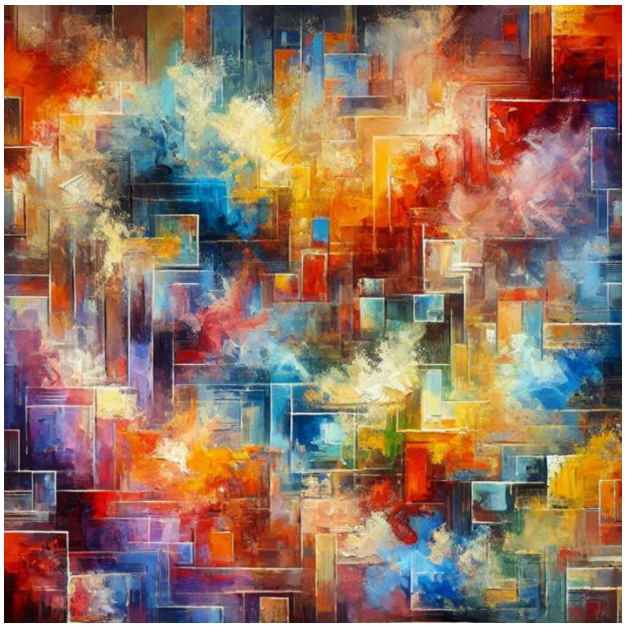
https://deepdreamgenerator.com/blog/abstract-art-and-ai

*Artwork 5*

*AI-generated abstract painting inspired by the sensation of drinking three espressos,.*

( 2025, 7.April). DALL·E via ChatGPT.

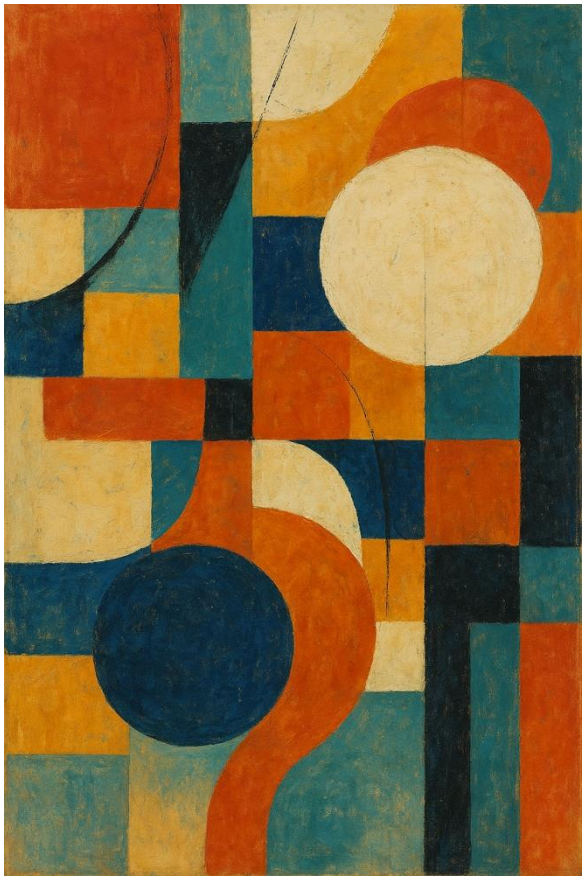
https://chatgpt.com/



*Artwork 6*

*Abstrakte geometrische Komposition mit Farben*. (2025, 25. April). ChatGPT.

https://chatgpt.com/s/m_680b926049d08191a564bb90dbfbf720



# Appendix B

## Artwork Descriptions

### *Positive Human Descriptions*

**HUM Description 1.** "This artwork, created by a skilled artist, demonstrates mastery of technique and showcases years of experience and dedication to the craft."

"Dit kunstwerk, gemaakt door een getalenteerde kunstenaar, getuigt van zijn meesterschap in de techniek en van jarenlange ervaring en toewijding aan het vak."

**HUM Description 2.** "Each detail in this piece reflects the artist's refined expertise and distinctive approach, making it a truly unique expression of artistic vision."

"Elk detail in dit stuk weerspiegelt de verfijnde expertise en de unieke aanpak van de kunstenaar, wat het tot een werkelijk unieke uiting van artistieke visie maakt."

### *Negative Human Descriptions*

**HUM Description 3.** *"Despite being human-made, this artwork reveals the limitations of subjective interpretation, showing how even skilled artists can struggle with expression of their artistic vision."*

*"Hoewel dit kunstwerk door mensen is gemaakt, toont het de beperkingen van subjectieve interpretatie en laat het zien hoe zelfs getalenteerde kunstenaars moeite kunnen hebben met het uiten van hun artistieke visie."*

**HUM Description 4.** *"While created by hand, this piece reflects human biases and imperfections, highlighting how artistic vision is often constrained by personal and cultural influences."*

*"Hoewel dit kunstwerk met de hand is gemaakt, zijn er toch menselijke vooroordelen en onvolkomenheden in het werk te zien. Het laat zien hoe de artistieke visie vaak wordt beperkt door persoonlijke en culturele invloeden."*

### *Neutral Human Descriptions*

**HUM Description 5 and 6.** *"This artwork is made by a human artist."*

*"Dit kunstwerk is gemaakt door een menselijke kunstenaar."*

***Positive AI Descriptions***

**AI Description 1.** *"This AI-generated piece demonstrates how technology is capable of artistic expression, creating intricate and thought-provoking visuals with precision and uniqueness."*

*"Dit door AI gegenereerde kunstwerk laat zien hoe technologie artistieke expressie mogelijk maakt door complexe en tot nadenken stemmende beelden te creëren met precisie en uniciteit."*

**AI Description 2.** *"Generated by advanced AI, this artwork pushes the boundaries, blending complex patterns and ideas beyond human imagination."*

*"Dit kunstwerk is gemaakt met behulp van geavanceerde kunstmatige intelligentie (AI) en verlegt de grenzen door complexe patronen en ideeën te combineren die de menselijke verbeelding te boven gaan."*

***Negative AI Descriptions***

**AI Description 3.** *"This artwork, generated by AI, demonstrates that even the most advanced technology fails to inspire, revealing the mechanical nature of algorithms."*

*"Dit door AI gegenereerde kunstwerk laat zien dat zelfs de meest geavanceerde technologie niet kan inspireren en onthult de mechanische aard van algoritmes."*

**AI Description 4.** *"Despite being produced by advanced technology, this AI-generated piece highlights the absence of genuine human inspiration and artistic intent."*

*"Hoewel dit kunstwerk met behulp van geavanceerde technologie is gemaakt, benadrukt het de afwezigheid van echte menselijke inspiratie en artistieke intentie."*

***Neutral AI Description***

**AI Description 5 and 6. "***This artwork is generated by AI."*

*”Dit kunstwerk is gegenereerd door AI.”*

## Appendix C

**Translations into Dutch**

***Beauty***

“I find this work beautiful” - “Ik vind dit werk mooi”

A score of 0 = “strongly disagree” - “erg mee oneens”

A score of 100 = “strongly agree” - “erg mee eens”

***Intentionality and Creativity***

“In my opinion, the level of intentionality involved in the creation of this work is…” - “Naar mijn mening is het niveau van intentionaliteit dat betrokken is bij het maken van dit werk…”

“In my opinion, the level of creativity involved in the creation of this work is…” - “Naar mijn mening is het niveau van creativiteit dat betrokken is bij het maken van dit werk…”

A score of 0 = “very low” - “heel laag”

A score of 100 = “very high” - “heel hoog”

***Aesthetic Fluency Scale***

“I don’t really know anything about this artist or term” - “Ik weet eigenlijk niets over deze kunstenaar of term”

“I’m familiar with this artist or term” - “Ik ben bekend met deze kunstenaar of term”

“I know a lot about this artist or term” - “Ik weet een hoop over deze kunstenaar of term”

***General Attitudes toward Artificial Intelligence Scale***

“Artificial Intelligence is exciting” - “Kunstmatige Intelligentie is uitdagend”

"I am impressed by what Artificial Intelligence can do" - "Ik ben onder de indruk van wat Kunstmatige Intelligentie kan doen"

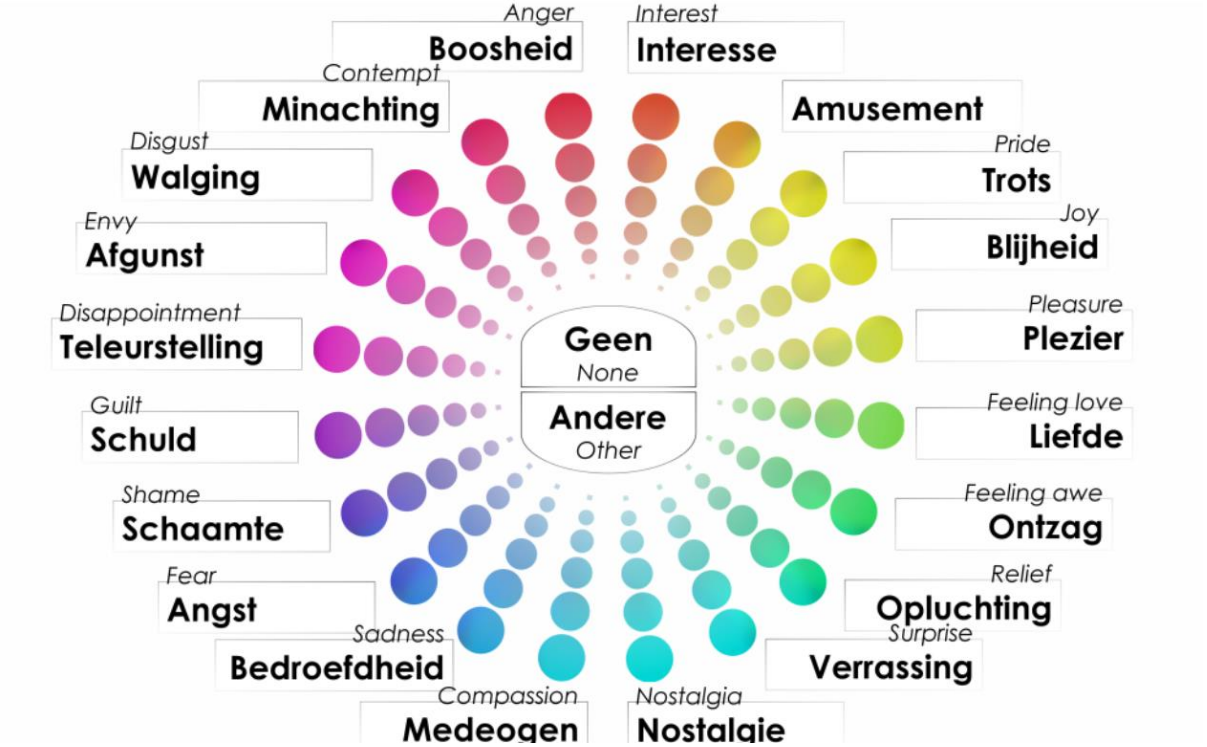
"There are many beneficial applications of Artificial Intelligence" - "Er zijn veel nuttige toepassingen van Kunstmatige Intelligentie"

"I am interested in using artificially intelligent systems in my daily life" - "In mijn dagelijks leven ben ik geïntereseerd in het gebruik van Kunstmatige Intelligente systemen"

"Artificial Intelligence can have positive impacts on people's wellbeing" - "Kunstmatige Intelligentie kan een positieve impact hebben op het welzijn van mensen"

## Appendix D

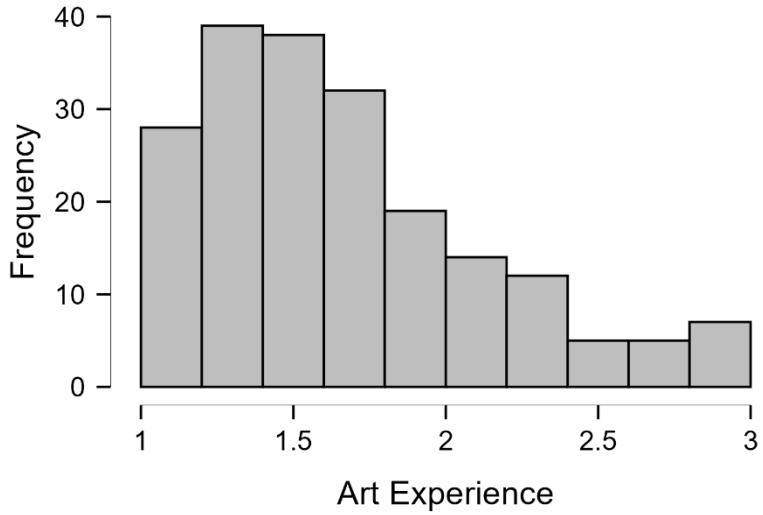
**The Geneva Emotion Wheel (GEW)**



## Appendix E

**Figure E1**

*Histogram of Aesthetic Fluency Scores Among Participants*




*Note.* Response options: (1) *I don't really know anything,* (2) *I'm familiar, and* (3) *I know a lot.*

**Figure E2**

*Q-Q plot of residuals for beauty ratings. Used to assess the normality assumption for the repeated-measures ANOVA (within-subject factor: artwork origin; between-subject factor: art experience).*

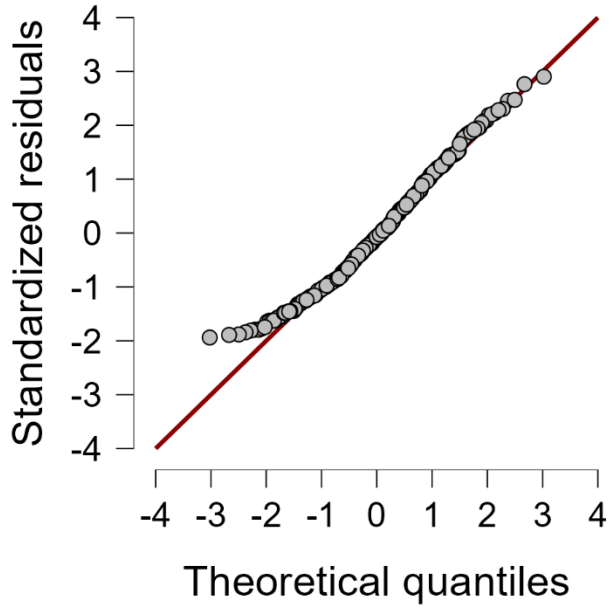


**Figure E3**

*Q-Q plot of residuals for creativity ratings. Used to assess the normality assumption for the repeated-measures ANOVA (within-subject factor: artwork origin; between-subject factor: art experience).*

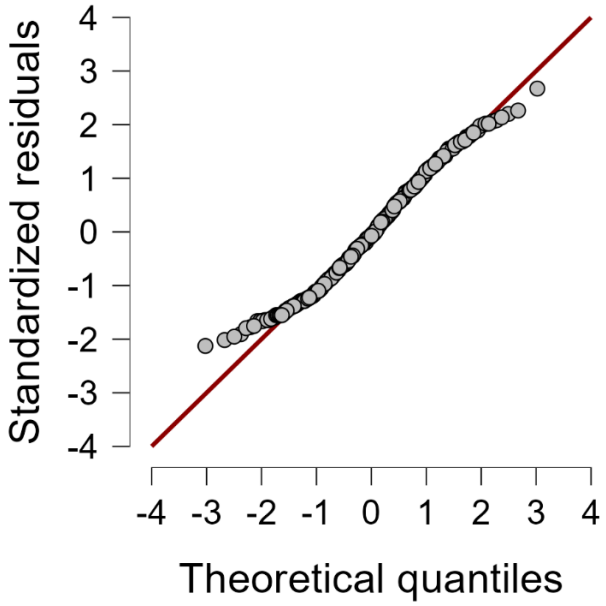



**Figure E4**

*Q-Q plot of residuals for emotional intensity ratings. Used to assess the normality assumption for the repeated-measures ANOVA (within-subject factor: artwork origin; between-subject factor: art experience).*