Supporting Information (SI) for Art has no gender, only gender bias

This PDF file includes:

Supporting text

Figures S1 to S4

Table S1

SI References

Other supporting materials for this manuscript include the following:

All data pertinent to this study, including R code, raw data, full details of statistical models and a copy of the current PDF can be freely accessed at: [https://doi.org/10.17605/OSF.IO/ZG965].

Citation:

Ruta, N., & Damiano, C. (2025, June 11). Art has no gender, only gender bias. https://doi.org/10.17605/OSF.IO/ZG965

Supporting Information for

Art has no gender, only gender bias.

Stefanie De Winter[1]*†, Nicole Ruta[2]*†, Claudia Damiano[3], Johan Wagemans[2]*

Affiliations
[1] Department of Art History, KU Leuven, Leuven, Belgium
[2] Laboratory of Experimental Psychology, Department of Brain and Cognition, KU Leuven, Leuven, Belgium
[3] Department of Psychology, University of Toronto, Toronto, Canada
* E-mail: stefanie.dewinter@kuleuven.be; nicole.ruta@kuleuven.be; johan.wagemans@kuleuven.be
† These two authors contributed equally to this work.

Keywords: Empirical aesthetics, Gender Bias, Art judgments, Art authorship, Abstract Expressionism

Experiment 1

Authorship sex categorization

The main aim of Experiment 1 was to test if participants could distinguish between artworks made by women and men artists. To test this, we asked participants to assess the sex authorship (male or female) for each of the paintings selected for our study. Moreover, we asked participants to assess the criterion which they relied on when categorizing the painting authorship as male or female. Participants could choose from the following three: 1) colour, 2) line and 3) composition.

Q1. Are people biased towards male authorship for abstract expressionism art?

To answer this question, we first conducted two Wilcoxon Signed-Rank tests to determine if the true median frequencies of choosing male and female chosen sex authorship significantly differed from chance level ($\mu = 80$). The results indicated that the median frequency for chosen male authorship choices (Mdn = 91) was significantly higher than chance level, $V = 16700$, $p < 0.001$. Similarly, the median frequency for chosen female authorship choices (Mdn = 69) was also significantly higher than chance level, $V = 2410$, $p < 0.001$. After applying a Bonferroni correction for multiple comparisons, both p-values remained below the adjusted alpha level ($p < .025$), confirming that these findings were not due to chance.

We then inspected the overall frequencies assigned to male and female authorship, according to the actual gender of the artist. The results of the Chi-square goodness-of-fit test revealed a significant association between the chosen sex for the paintings' authorship and the actual gender of the artist ($\chi^2(1) = 266.69$, $p < .001$). The observed counts in each cell of

the contingency table were compared to the expected equally distributed counts[1], assuming no association between the variables. The chi-square test statistic indicated a small effect size (Cramér's $V = 0.09$) with a p-value of less than .001, suggesting that the observed association was unlikely to occur by chance alone, confirming a bias towards choosing male authorship for abstract expressionism art made by both women (52.6%) and men (61.6%) artists, as illustrated in Fig. **S*1***.

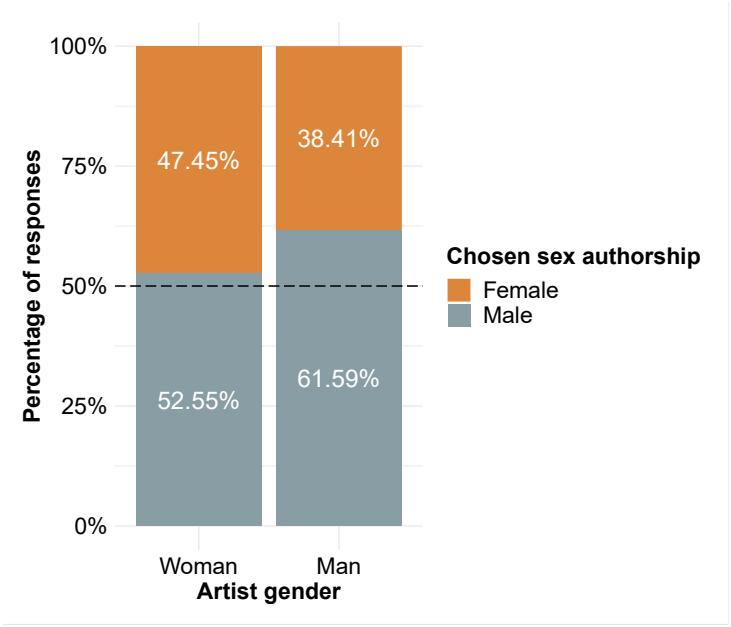

Fig. S1 *Stacked bar plot of chosen sex authorship (percentage of responses) by artist gender (Male and Female).*

Q2. Are people more likely to assign male authorship if relying on a specific stylistic criterion?

We then tested if participants were more likely to assign male authorship when relying on a specific criterion, i.e., colours, line, or composition. The results of the Chi-square goodness-of-fit test revealed a significant association between the chosen sex for the

[1] Expected counts were assumed to be equally distributed, so each cell entry had 8000 counts (32000 total entries / 4 categories).

paintings' authorship and the criterion used ($\chi^2(2) = 447.77$, $p < .001$). As before, observed counts were compared to equally distributed expected counts[2]. The chi-square test statistic indicated a small association (Cramér's $V = 0.12$) with p-value of less than .001, suggesting that the observed association was unlikely to occur by chance alone. The results confirmed a bias towards choosing male authorship when participants relied on composition (60.71%) and line (62.3%), and a bias towards choosing female authorship when relying on colours (51.1%; see Fig. **S2**).

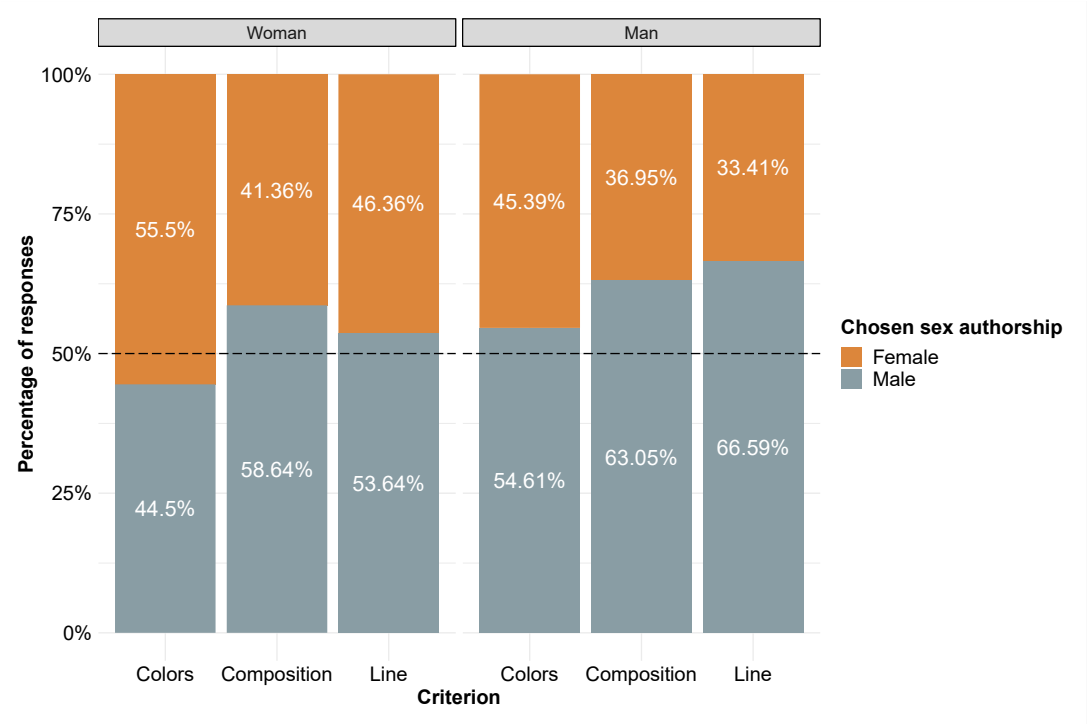

Fig. S2 *Stacked bar plot of chosen sex authorship (percentage of responses) by artist gender (Male and Female) and criteria used by participants (Colors, Composition and Line).*

[2] Expected counts were assumed to be equally distributed, so each cell entry had 5333.33 counts (32000 total entries / 6 categories). Observed counts are reported in Table 1 in the main manuscript.

Q3. A) Can we predict the likelihood of choosing male authorship from the actual gender of the artists? B) And what is the role of art history-relevant stylistic properties as quantified by image statistics?

Data analysis

For each painting we calculated the likelihood of choosing male authorship by dividing the total number of frequencies associated with male authorship (*Criterion Freq* $_{Male}$) to the total number of frequencies, according to each of the three criteria used in Experiment 1 (i.e. colours, line and composition), as described by the formula below:

$$Male\ authorship\ likelihood\ (male\ \Lambda) = \frac{Criterion\ Freq_{Male}}{Criterion\ Freq_{Male} + Criterion\ Freq_{Female}}$$

The likelihood value could vary between 0 and 1, with 0 indicating that, when relying on a certain criterion, no participant categorised a certain painting being made by a male artist; and 1 indicating that all participants chose male authorship instead.

We adopted a stepwise model selection and fitted a series of nested hierarchical linear mixed-effects models to assess the effect of *artist identity*, *criterion*, *response time* and eight relevant Image Features (IF) on the likelihood of choosing male authorship. To quantify colours, we used the following IFs: *percentage of warm hue, percentage of cool hue* and *percentage of black*; to quantify composition: *alloverness, central* and *vertical symmetry*; and to quantify line: *average length* and *average angularity.* The models compared included a baseline model with random intercept within stimuli (160 paintings) and no predictors, and subsequent models that progressively added *artist identity* (Pollock *vs* Krasner *vs* Kline *vs*

Hartigan *vs* Louis *vs* Frankenthaler *vs* Twombly *vs* Mitchell), *criterion* (*Colour* vs *Line* vs *Composition*) and their interaction, *response time* and the eight aforementioned *IFs* as fixed effects. The models were assessed based on a number of parameters, i.e., Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood, deviance, Chi-squared test statistic, degrees of freedom, and p-value using both the *anova* and the *compare_performance* (specifically for R2 conditional and marginal; Lüdecke et al., 2002) functions in R. Retaining the main effects of *response time*, IFs for composition (*alloverness, central* and *vertical symmetry*), *percentage of cool hue,* and the interaction between *artist identity* and *criterion* did not significantly improve the model, therefore these effects were not retained in the final model and will not be further discussed when reporting the results (details of additional stepwise model selection are available in the file [Exp1_model selection.pdf], [Experiment 1] folder, at: https://osf.io/zg965/?view_only=9ae0dca9286a4f029d120aec2672a883).

Results

The final model's total explanatory power is substantial (conditional R^2 = 0.74) and the part related to the fixed effects alone (marginal R^2) is of 0.47 (Table S1). The planned contrasts for *artist identity* were set to compare Pollock against all other artists, as the most famous artist among the selected ones in our study. The planned contrasts for *criterion* compared line *vs* composition and colour *vs* the other two. The model's intercept, corresponding to *artist identity* = Pollock, *percentage warm hue* = 0, *percentage black* = 0, *average length* = 0 and *average angularity* = 0, is at -0.32 (95% CI [-1.10, 0.46], $t(466) = -0.80$, $p = 0.424$).

Within this model, the main effect of *artist identity* when comparing respectively Krasner's, Frankenthaler's and Louis' artworks likelihood of male authorship with Pollock's

was statistically significant and negative (Krasner: beta = -0.65, 95% CI [-1.03, -0.27], $t(466) = -3.39$, $p < .001$; Std. beta = -0.65, 95% CI [-1.02, -0.27]); Frankenthaler: (beta = -0.59, 95% CI [-1.01, -0.17], $t(466) = -2.75$, $p = 0.006$; Std. beta = -0.58, 95% CI [-0.99, -0.17]); Louis: (beta = -1.07, 95% CI [-1.55, -0.58], $t(466) = -4.29$, $p < .001$; Std. beta = -1.11, 95% CI [-1.59, -0.64]), meaning that participants were more likely to say that Pollock's artworks were made by a male compared to the ones by Krasner, Frankenthaler and Louis. When comparing Kline and Pollock, the main effect of *artist identity* was statistically significant and positive (beta = 0.44, 95% CI [0.05, 0.83], $t(466) = 2.23$, $p = 0.026$; Std. beta = 0.45, 95% CI [0.06, 0.83]), meaning that participants were more likely to attribute male authorship to Kline's artworks compared to the ones made by Pollock. Comparisons between Pollock and Hartigan (beta = -0.11, 95% CI [-0.51, 0.29], $t(466) = -0.53$, $p = 0.595$; Std. beta = -0.10, 95% CI [-0.50, 0.29]), Pollock and Twombly (beta = -0.13, 95% CI [-0.52, 0.26], $t(466) = -0.64$, $p = 0.523$; Std. beta = -0.12, 95% CI [-0.51, 0.26]) and Pollock and Mitchell (beta = -0.05, 95% CI [-0.46, 0.35], $t(466) = -0.26$, $p = 0.794$; Std. beta = -0.05, 95% CI [-0.45, 0.35]) were statistically non-significant. Post-hoc comparisons performed with *emmeans* (Lenth, 2023) and with *p* value adjusted with Bonferroni correction revealed that the other art history relevant comparisons between Kline and Hartigan ($t(148) = 2.79$, $p = 0.16$), Louis and Frankenthaler ($t(148) = -2.29$, $p = 0.66$) as well as Twombly and Mitchell ($t(148) = -0.36$, $p = 1.00$) were all statistically non-significant.

The effects of *percentage black* (beta = 0.36, 95% CI [0.25, 0.47], $t(466) = 6.28$, $p < .001$; Std. beta = 0.36, 95% CI [0.25, 0.47]) and *average angularity* (beta = 0.15, 95% CI [0.03, 0.28], $t(466) = 2.37$, $p = 0.018$; Std. beta = 0.16, 95% CI [0.03, 0.28]) are both statistically significant and positive, meaning that participants were more likely to attribute male authorship to darker and more angular paintings. On the other hand, the effects of *percentage warm hue* (beta = -0.05, 95% CI [-0.17, 0.06], $t(466) = -0.88$, $p = 0.380$; Std. beta

= -0.05, 95% CI [-0.17, 0.06]) and *average length* [log] (beta = 0.16, 95% CI [-0.05, 0.37], $t(466) = 1.51$, $p = 0.132$; Std. beta = 0.40, 95% CI [0.04, 0.77]) were both statistically non-significant.

Table S1. *Linear mixed-effects model for likelihood of choosing male authorship in Experiment 1.*

	Model for likelihood of choosing male autorship			
Predictors	*Estimates*	*CI*	*p*	*df*
(Intercept)	-0.32	-1.11 – 0.47	0.425	148.00
artist name [Krasner]	-0.65	-1.03 – -0.27	**0.001**	148.00
artist name [Kline]	0.44	0.05 – 0.83	**0.027**	148.00
artist name [Hartigan]	-0.11	-0.51 – 0.29	0.596	148.00
artist name [Louis]	-1.07	-1.56 – -0.57	**<0.001**	148.00
artist name [Frankenthaler]	-0.59	-1.01 – -0.16	**0.007**	148.00
artist name [Twombly]	-0.13	-0.52 – 0.27	0.523	148.00
artist name [Mitchell]	-0.05	-0.46 – 0.35	0.794	148.00
percentageWarmHue	-0.05	-0.17 – 0.06	0.381	148.00
percentageBlack	0.36	0.24 – 0.47	**<0.001**	148.00
averageLength [log]	0.16	-0.05 – 0.37	0.133	148.00
averageCurvature	0.15	0.03 – 0.28	**0.019**	148.00
Random Effects				
σ^2	0.27			
$\tau_{00\ image\_meaningful}$	0.27			
ICC	0.50			
$N_{\ image\_meaningful}$	160			
Observations	480			
Marginal R^2 / Conditional R^2	0.474 / 0.738			

Experiment 2

Evaluating gendered and normative art criticism derived attributes

Data analysis

We adopted the same stepwise model selection approach as outlined in Experiment 1. For Experiment 2, we fitted four different linear models to assess the effect of *artist identity*, *average likelihood of choosing male authorship* (from Experiment 1), and their interaction on each of the four categories of gendered and normative art criticism derived attributes used to describe works made by 1) women and 2) men art, and 3) good and 4) bad art. Retaining the interaction effect between *artist identity* and *average likelihood of choosing male authorship* did not significantly improve any of the models, while retaining *average likelihood of choosing male authorship* did not significantly improve only the model for attributes used to describe art made by men, therefore these effects were not retained in the final models and will not be further discussed when reporting the results (details of stepwise model selection are provided in the online supplementary materials at: https://doi.org/10.17605/OSF.IO/ZG965, file [ArtHasNoGender_analysis_Aug_2023.R], [Data Analysis] folder).

Results

Attributes used to describe women art

We fitted four separate linear models to assess the main effects of *artist identity* and *average likelihood of choosing male authorship* on respectively the value associated with art criticism derived attributes used to describe works made by 1) women and 2) men art, as well as 3) good and 4) bad art. Average ratings on the four attributes categories are illustrated in Figure S3.

The first model predicting the value associated with attributes used to describe works made by woman artists explains a statistically significant and substantial proportion of variance (R2 = 0.58, F(8, 151) = 26.12, $p < .001$, adj. R2 = 0.56). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at -0.82 (95% CI [-1.12, -0.52], $t(151) = -5.36$, $p < .001$). The full model output is available in the online Supplementary Materials: file [WomenArt_model .pdf], [Experiment 2] folder.

Within this model and in line with previous findings, the effects of Pollock *vs* Kline (beta = 1.27, 95% CI [0.85, 1.69], $t(151) = 5.97$, $p < .001$; Std. beta = 1.27, 95% CI [0.85, 1.69]), Pollock *vs* Louis (beta = 1.62, 95% CI [1.15, 2.09], $t(151) = 6.80$, $p < .001$; Std. beta = 1.62, 95% CI [1.15, 2.09]), Pollock *vs* Frankenthaler (beta = 1.66, 95% CI [1.21, 2.11], $t(151) = 7.29$, $p < .001$; Std. beta = 1.66, 95% CI [1.21, 2.11]), Pollock *vs* Twombly (beta = 1.28, 95% CI [0.86, 1.69], $t(151) = 6.07$, $p < .001$; Std. beta = 1.28, 95% CI [0.86, 1.69]), Pollock *vs* Mitchell (beta = 0.66, 95% CI [0.24, 1.08], $t(151) = 3.09$, $p = 0.002$; Std. beta = 0.66, 95% CI [0.24, 1.08]) were all statistically significant and positive. When comparing Pollock *vs* Krasner (beta = -0.11, 95% CI [-0.54, 0.32], $t(151) = -0.49$, $p = 0.626$; Std. beta = -0.11, 95% CI [-0.54, 0.32]) and Pollock *vs* Hartigan (beta = 0.19, 95% CI [-0.23, 0.61], $t(151) = 0.91$, $p = 0.366$; Std. beta = 0.19, 95% CI [-0.23, 0.61]) the effect is statistically non-significant.

Post-hoc comparisons revealed that among the four art history relevant comparisons, only the one between Kline and Hartigan ($t(151) = -4.96$, $p = 0. 0001$) was statistically significant, with the man artists, Kline (M = 2.83), obtaining significantly higher values for attributes used to describe woman art compared to his woman colleague, Hartigan (M = 2.49). All other art history relevant comparisons were not statistically significant ($p > .05$).

All post-hoc comparisons are available in the online Supplementary Materials: file [WomenArt_post_hoc.pdf], [Experiment 2] folder.

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -0.22, 95% CI [-0.36, -0.08], $t(151) = -3.21$, $p = 0.002$; Std. beta = -0.22, 95% CI [-0.36, -0.08]), meaning that the artworks that were more likely to be associated with male authorship reported significantly lower score for the art-relevant attributes used to describe woman artists.

Attributes used to describe men art

We fitted a linear model to assess the effect of *artist identity* on the value associated with attributes used to describe works made by man artists.

The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.45$, $F(7, 152) = 17.82$, $p < .001$, adj. $R^2 = 0.43$). The model's intercept, corresponding to *artist identity* = Pollock, is at 1.01 (95% CI [0.68, 1.35], $t(152) = 5.98$, $p < .001$). The full model output is available in the online Supplementary Materials: file [MenArt_model .pdf], [Experiment 2] folder.

Within this model, the effects of Pollock *vs* Kline (beta = -1.41, 95% CI [-1.89, -0.94], $t(152) = -5.90$, $p < .001$; Std. beta = -1.41, 95% CI [-1.89, -0.94]), Pollock *vs* Louis (beta = -1.32, 95% CI [-1.79, -0.84], $t(152) = -5.49$, $p < .001$; Std. beta = -1.32, 95% CI [-1.79, -0.84]), Pollock *vs* Frankenthaler (beta = -1.88, 95% CI [-2.36, -1.41], $t(152) = -7.86$, $p < .001$; Std. beta = -1.88, 95% CI [-2.36, -1.41]), Pollock *vs* Twombly (beta = -1.85, 95% CI [-.32, -1.37], $t(152) = -7.70$, $p < .001$; Std. beta = -1.85, 95% CI [-2.32, -1.37]) and Pollock *vs* Mitchell (beta = -0.92, 95% CI [-1.40, -0.45], $t(152) = -3.85$, $p < .001$; Std. beta = -0.92, 95% CI [-1.40, -0.45]) are statistically significant and negative. Instead, the effects Pollock *vs*

Krasner (beta = -0.34, 95% CI [-0.82, 0.13], $t(152) = -1.43$, $p = 0.155$; Std. beta = -0.34, 95% CI [-0.82, 0.13]) and Pollock *vs* Hartigan (beta = -0.38, 95% CI [-0.86, 0.09], $t(152) = -1.60$, $p = 0.111$; Std. beta = -0.38, 95% CI [-0.86, 0.09]) are statistically non-significant and negative.

Post-hoc comparisons revealed that among the four art history relevant comparisons, only the two between Kline and Hartigan ($t(152) = -4.30$, $p = 0.0009$) and Twombly and Mitchell ($t(152) = -3.85$, $p = 0.0049$) were statistically significant, with men artists –Kline (M = 2.80) and Twombly (M = 2.67)– obtaining significantly lower values for attributes used to describe man art compared to their women colleagues – Hartigan (M = 3.11) and Mitchell (M = 2.95). All post-hoc comparisons are available in the online Supplementary Materials: file [MenArt_post_hoc.pdf], [Experiment 2] folder.

Attributes used to describe good art, regardless of artists' gender

The second model predicting the value associated with attributes used to positively describe art explains a statistically significant and substantial proportion of variance (R2 = 0.53, $F(8, 151) = 21.11$, $p < .001$, adj. R2 = 0.50). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at 0.94 (95% CI [0.62, 1.26], $t(151) = 5.81$, $p < .001$). The full model output is available in the online Supplementary Materials: file [GoodArt_model .pdf], [Experiment 2] folder.

Within this model, the effects of Pollock *vs* Kline (beta = -1.40, 95% CI [-1.85, -0.96], $t(151) = -6.21$, $p < .001$; Std. beta = -1.40, 95% CI [-1.85, -0.96]), Pollock *vs* Louis (beta = -1.05, 95% CI [-1.55, -0.55], $t(151) = -4.15$, $p < .001$; Std. beta = -1.05, 95% CI [-1.55, -0.55]), Pollock *vs* Frankenthaler (beta = -1.67, 95% CI [-2.15, -1.19], $t(151) = -6.91$, $p < .001$; Std. beta = -1.67, 95% CI [-2.15, -1.19]), Pollock *vs* Twombly (beta = -1.86, 95% CI [-2.31, -1.42], $t(151) = -8.35$, $p < .001$; Std. beta = -1.86, 95% CI [-2.31, -1.42]) and Pollock

vs Mitchell (beta = -0.93, 95% CI [-1.37, -0.48], $t(151) = -4.12$, $p < .001$; Std. beta = -0.93, 95% CI [-1.37, -0.48]) were all statistically significant and negative. However, when comparing Pollock *vs* Krasner (beta = -0.42, 95% CI [-0.87, 0.04], $t(151) = -1.81$, $p = 0.073$; Std. beta = -0.42, 95% CI [-0.87, 0.04]) and Pollock *vs* Hartigan (beta = -0.21, 95% CI [-0.65, 0.23], $t(151) = -0.94$, $p = 0.348$; Std. beta = -0.21, 95% CI [-0.65, 0.23]) the effect was statistically non-significant.

Post-hoc comparisons revealed that among the four art history relevant comparisons, only the two between Kline and Hartigan ($t(151) = -5.17$, $p < .0001$) Twombly and Mitchell ($t(151) = -4.18$, $p = 0.0014$) were statistically significant, with the men artists, Kline (M=2.63) and Twombly (M=2.53), obtaining significantly higher values for attributes used to describe woman art compared to his women colleagues, Hartigan (M = 3.30) and Mitchell (M = 3.00). All other art history relevant comparisons were not statistically significant ($p > .05$). All post-hoc comparisons are available in the online Supplementary Materials: file [GoodArt_post_hoc.pdf], [Experiment 2] folder.

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -0.38, 95% CI [-0.53, -0.24], $t(151) = -5.23$, $p < .001$; Std. beta = -0.38, 95% CI [-0.53, -0.24]).

Attributes used to describe bad art, regardless of artists' gender

The third model predicting the value associated with attributes used to negatively describe art explains a statistically significant and substantial proportion of variance (R2 = 0.58, F(8, 151) = 25.61, $p < .001$, adj. R2 = 0.55). The model's intercept, corresponding to *artist identity* = Pollock and *likelihood of choosing male authorship* = 0, is at 0.89 (95% CI [0.59, 1.20], $t(151) = 5.81$, $p < .001$). The full model output is available in the online Supplementary Materials: file [BadArt_model .pdf], [Experiment 2] folder.

Within this model, the effect of *artist identity* was significant and negative when comparing Pollock *vs* Kline (beta = -2.00, 95% CI [-2.43, -1.58], $t(151) = -9.37$, $p < .001$; Std. beta = -2.00, 95% CI [-2.43, -1.58]), Pollock *vs* Hartigan (beta = -1.13, 95% CI [-1.55, -0.71], $t(151) = -5.30$, $p < .001$; Std. beta = -1.13, 95% CI [-1.55, -0.71]), Pollock *vs* Louis (beta = -0.95, 95% CI [-1.42, -0.47], $t(151) = -3.95$, $p < .001$; Std. beta = -0.95, 95% CI [-1.42, -0.47]), Pollock *vs* Frankenthaler (beta = -1.68, 95% CI [-2.13, -1.22], $t(151) = -7.32$, $p < .001$; Std. beta = -1.68, 95% CI [-2.13, -1.22]), Pollock *vs* Twombly (beta = -1.09, 95% CI [-1.51, -0.67], $t(151) = -5.14$, $p < .001$; Std. beta = -1.09, 95% CI [-1.51, -0.67]) and Pollock *vs* Mitchell (beta = -0.69, 95% CI [-1.11, -0.27], $t(151) = -3.23$, $p = 0.002$; Std. beta = -0.69, 95% CI [-1.11, -0.27]). When comparing Pollock vs Krasner, the effect of *artist identity* was statistically non-significant (beta = 0.38, 95% CI [-0.05, 0.82], $t(151) = 1.75$, $p = 0.082$; Std. beta = 0.38, 95% CI [-0.05, 0.82]).

Post-hoc comparisons among the four art history relevant comparisons, only the two between Kline and Hartigan ($t(151) = -4.02$, $p = 0.0026$) Louis and Frankenthaler ($t(151) = 3.43$, $p = 0.0219$) were statistically significant. The two pairs showed opposite trends for negative attributes, with Kline (M=2.63) having significantly lower ratings than Hartigan (M = 2.30) while Louis (M = 3.15) had significantly higher ratings than Frankenthaler (M = 2.81). The other art history relevant comparisons between Pollock *vs* Krasner and Twombly *vs* Mitchell were not statistically significant ($p > .05$). All post-hoc comparisons are available in the online Supplementary Materials: file [BadArt_post_hoc.pdf], [Experiment 2] folder.

The effect of *likelihood of choosing male authorship* on the value associated with attributes used to negatively describe art is statistically significant and positive (beta = 0.22, 95% CI [0.08, 0.36], $t(151) = 3.18$, $p = 0.002$; Std. beta = 0.22, 95% CI [0.08, 0.36]).

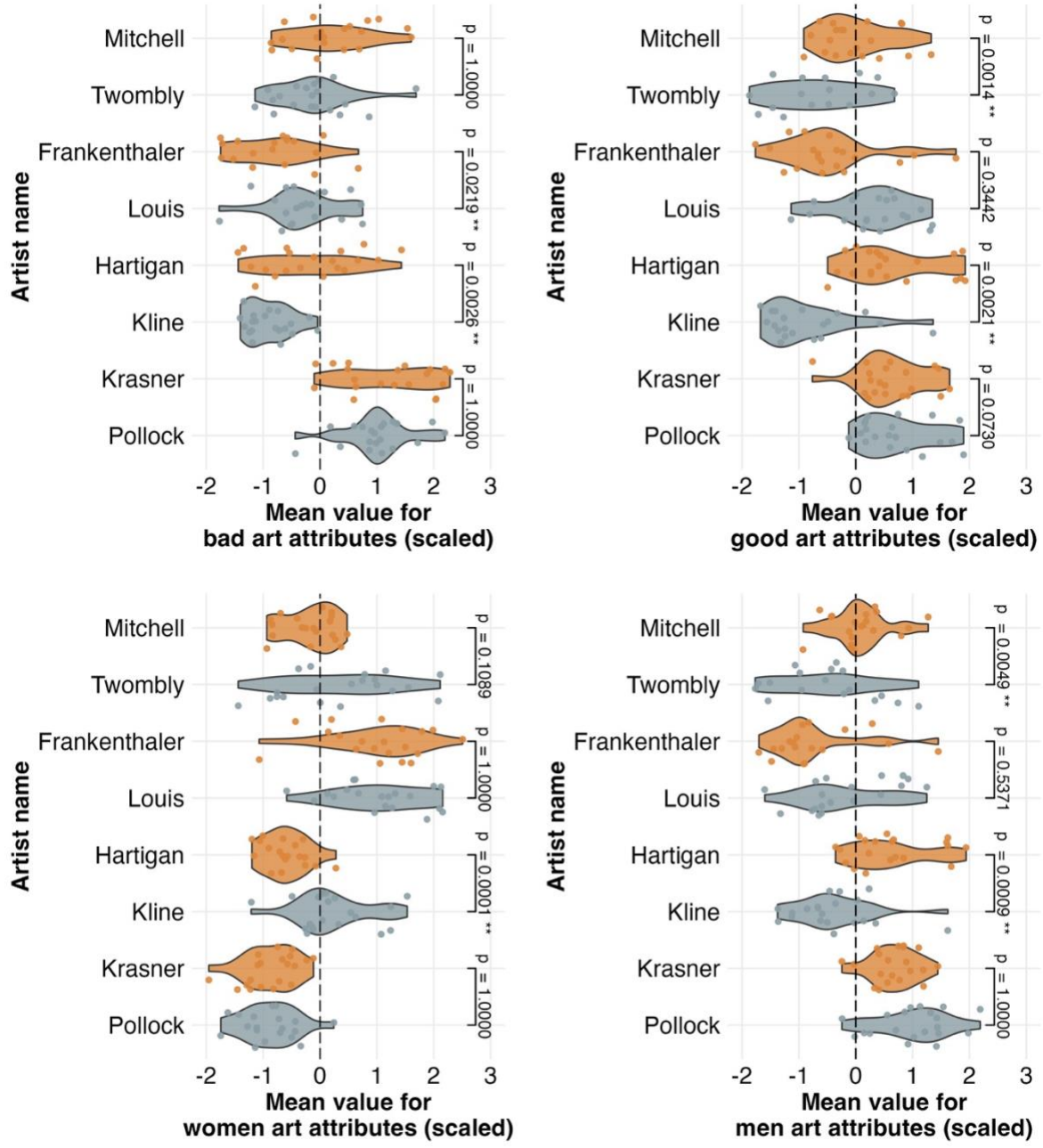

Fig. S3 *On the top row, violin plots of average value for normative attributes –bad art (top left) and good art (top right)–, according to artist identity. On the bottom row, violin plots of average value for gendered attributes used to describe works made by women (bottom left) and man (bottom right) artists, according to artist identity. Women artists are indicated in orange, while men artists are in light blue.*

Experiment 3

Aesthetic Appreciation

Data analysis

Following the same rationale of the previous experiments, we fitted five separate linear models to assess the main effects of *artist identity* and *average likelihood of choosing male authorship* on aesthetic appreciation ratings, respectively: liking, pleasure, interest, order and complexity (Figure S4).

Retaining the interaction effect between *artist identity* and *average likelihood of choosing male authorship* did not significantly improve any of the models, except the model predicting complexity ratings, while retaining *average likelihood of choosing male authorship* did not significantly improve only the model for complexity, therefore these effects were not retained in the final models and will not be further discussed when reporting the results (details of stepwise model selection are provided in the online supplementary materials at: https://doi.org/10.17605/OSF.IO/ZG965, file [ArtHasNoGender_analysis_Aug_2023.R], [Data Analysis] folder).

Results

Liking ratings

We fitted a linear model to predict liking ratings with *artist identity* and *average likelihood of choosing male authorship*. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.55$, $F(8, 151) = 23.14$, $p < .001$, adj. $R^2 = 0.53$). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at 3.65 (95% CI [3.38, 3.92], $t(151) = 26.69$, $p < .001$). The

full model output is available in the online Supplementary Materials: file [Liking_model .pdf], [Experiment 3] folder.

Within this model, the following effects were statistically significant and negative: Pollock *vs* Kline (beta = -0.39, 95% CI [-0.61, -0.17], $t(151) = -3.54$, $p < .001$; Std. beta = -0.78, 95% CI [-1.21, -0.34]) ; Pollock *vs* Frankenthaler (beta = -0.49, 95% CI [-0.72, -0.25], $t(151) = -4.10$, $p < .001$; Std. beta = -0.97, 95% CI [-1.43, -0.50]); Pollock *vs* Twombly (beta = -0.74, 95% CI [-0.96, -0.52], $t(151) = -6.72$, $p < .001$; Std. beta = -1.46, 95% CI [-1.89, -1.03]) and Pollock *vs* Mitchell (beta = -0.23, 95% CI [-0.45, -5.55e-03], $t(151) = -2.03$, $p = 0.045$; Std. beta = -0.45, 95% CI [-0.88, -0.01]). The effects of Pollock *vs* Krasner (beta = -0.15, 95% CI [-0.38, 0.07], $t(151) = -1.32$, $p = 0.189$; Std. beta = -0.30, 95% CI [-0.74, 0.15]), Pollock *vs* Hartigan (beta = -0.14, 95% CI [-0.36, 0.08], $t(151) = -1.26$, $p = 0.209$; Std. beta = -0.28, 95% CI [-0.71, 0.16]) and Pollock *vs* Louis (beta = -0.07, 95% CI [-0.32, 0.17], $t(151) = -0.59$, $p = 0.556$; Std. beta = -0.15, 95% CI [-0.63, 0.34]) were all statistically non-significant and negative.

Post hoc comparisons revealed that the only art history relevant pairs that reported significant differences in complexity ratings were Louis vs Frankenthaler ($t(152) = 3.75$, $p < .0071$) and Twombly vs Mitchell ($t(152) = -4.65$, $p = .0002$), with Louis (man, M = 3.17) and Mitchell (woman, M = 2.71) being liked significantly more compared to Frankenthaler (woman, M = 2.66) and Twombly (man, M = 2.11) respectively. Interestingly, Louis was the artist that had the highest liking ratings as well as being the least likely to be associated with male authorship overall. However, Louis reported significantly higher liking ratings compared only to Twombly ($t(152) = 5.46$, $p < .0001$), which was also the least liked artist. All other post hoc comparisons between Louis and Pollock, Krasner, Hartigan, Kline and Mitchell were non-significant ($p > .05$). On the other hand, Twombly was liked significantly

less compared to most artists, except for Kline and Frankenthaler ($p > .05$). All post-hoc comparisons are available in the online Supplementary Materials: file [Liking_post_hoc.pdf], [Experiment 3] folder.

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -1.21, 95% CI [-1.53, -0.89], $t(151) = -7.46$, $p < .001$; Std. beta = -0.53, 95% CI [-0.67, -0.39]).

Pleasure ratings

We fitted a linear model (estimated using OLS) to predict pleasure ratings with *artist identity* and *average likelihood of choosing male authorship*. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.68$, $F(8, 151) = 39.74$, $p < .001$, adj. $R^2 = 0.66$). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at 3.65 (95% CI [3.40, 3.90], $t(151) = 29.07$, $p < .001$).

Within this model, only the effects of Pollock *vs* Twombly (beta = -0.59, 95% CI [-0.79, -0.39], $t(151) = -5.86$, $p < .001$; Std. beta = -1.08, 95% CI [-1.44, -0.72]) is statistically significant and negative, while the effects of comparing Pollock with the other artists were non statistically significant. The full model output is available in the online Supplementary Materials: file [Pleasure_model .pdf], [Experiment 3] folder.

Post hoc comparisons revealed that among the art history relevant pairs, only the comparisons between Louis vs Frankenthaler ($t(151) = 4.24$, $p = .0011$) and Twombly vs Mitchell ($t(151) = -5.04$, $p = <.0001$) were statistically significant, showing the same pattern reported by liking ratings. All post-hoc comparisons are available in the online Supplementary Materials: file [Pleasure_post_hoc.pdf], [Experiment 3] folder.

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -1.51, 95% CI [-1.81, -1.22], $t(151) = -10.18$, $p < .001$; Std. beta = -0.61, 95% CI [-0.73, -0.50]).

Interest ratings

We fitted a linear model to predict interest ratings with *artist identity* and *average likelihood of choosing male authorship*. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.43$, $F(8, 151) = 14.37$, $p < .001$, adj. $R^2 = 0.40$). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at 3.78 (95% CI [3.48, 4.09], $t(151) = 24.42$, $p < .001$). The full model output is available in the online Supplementary Materials: file [Interest_model.pdf], [Experiment 3] folder.

Withing this model, the effects of Pollock *vs* Kline is statistically significant and negative (beta = -0.63, 95% CI [-0.88, -0.38], $t(151) = -4.97$, $p < .001$; Std. beta = -1.23, 95% CI [-1.72, -0.74]), Pollock *vs* Louis (beta = -0.58, 95% CI [-0.86, -0.30], $t(151) = -4.11$, $p < .001$; Std. beta = -1.14, 95% CI [-1.69, -0.59]), Pollock *vs* Frankenthaler (beta = -0.73, 95% CI [-1.00, -0.46], $t(151) = -5.42$, $p < .001$; Std. beta = -1.43, 95% CI [-1.96, -0.91]), Pollock *vs* Twombly (beta = -0.75, 95% CI [- .99, -0.50], $t(151) = -5.98$, $p < .001$; Std. beta = -1.46, 95% CI [-1.95, -0.98]) and Pollock *vs* Mitchell (beta = -0.33, 95% CI [-0.58, -0.08], $t(151) = -2.62$, $p = 0.010$; Std. beta = -0.65, 95% CI [-1.14, -0.16]) were statistically significant and negative. While the effects of Pollock vs Krasner (beta = -0.17, 95% CI [-0.43, 0.08], $t(151) = -1.33$, $p = 0.186$; Std. beta = -0.34, 95% CI [-0.84, 0.16]) and Pollock *vs* Hartigan is statistically non-significant and negative (beta = -0.11, 95% CI [-0.36, 0.14], $t(151) = -0.90$, $p = 0.372$; Std. beta = -0.22, 95% CI [-0.71, 0.27]) were statistically non-significant and negative.

Similarly, as before for liking and pleasure ratings, post hoc comparisons for interest revealed that Twombly (M= 2.36) obtained significantly lower ratings compared to Mitchell (M= 2.85; $t(151) = -3.32$, $p = 0.0315$), while Kline (M = 2.34) was rated significantly less interesting compared to Hartigan (M = 3.03; $t(151) = -3.99$, $p = .0029$). All post-hoc comparisons are available in the online Supplementary Materials: file [Interest_post_hoc.pdf], [Experiment 3] folder.

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -1.03, 95% CI [-1.39, -0.66], $t(151) = -5.60$, $p < .001$; Std. beta = -0.45, 95% CI [-0.61, -0.29]).

Order ratings

We fitted a linear model to predict order ratings with *artist identity* and *average likelihood of choosing male authorship* and their interaction. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.60$, $F(15, 144) = 14.62$, $p < .001$, adj. $R^2 = 0.56$). The model's intercept, corresponding to *artist identity* = Pollock and *average likelihood of choosing male authorship* = 0, is at 3.80 (95% CI [2.78, 4.81], $t(144) = 7.38$, $p < .001$).

Adding *artist identity* ($F(7, 152) = 27.1310$, $p < 0.000001$) and its interaction with *average likelihood of choosing male authorship* ($F(7, 144) = 3.5931$, $p = 0.0014$) significantly improved the model. None of the post-hoc comparisons between the art history relevant pairs of artists were significant ($p > .05$).

The effect of *average likelihood of choosing male authorship* is statistically significant and negative (beta = -2.19, 95% CI [-3.62, -0.76], $t(144) = -3.03$, $p = 0.003$; Std. beta = -0.65, 95% CI [-1.08, -0.23]).

The full model output is available in the online Supplementary Materials: file [Order_model .pdf], [Experiment 3] folder. All post-hoc comparisons are available in the online Supplementary Materials: file [Order_post_hoc.pdf], [Experiment 3] folder.

Complexity ratings

We fitted a linear model to predict complexity ratings with *artist identity*. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.64$, $F(7, 152) = 38.05$, $p < .001$, adj. $R^2 = 0.62$). The model's intercept, corresponding to *artist identity* = Pollock, is at 3.90 (95% CI [3.67, 4.13], $t(152) = 33.35$, $p < .001$). The full model output is available in the online Supplementary Materials: file [Complexity_model .pdf], [Experiment 3] folder.

Within this model, the effects of Pollock vs Kline (beta = -1.68, 95% CI [-2.01, -1.35], $t(152) = -10.17$, $p < .001$; Std. beta = -1.98, 95% CI [-2.37, -1.60]), Pollock *vs* Hartigan (beta = -0.49, 95% CI [-0.81, -0.16], $t(152) = -2.95$, $p = 0.004$; Std. beta = -0.57, 95% CI [-0.96, -0.19]), Pollock *vs* Louis (beta = -1.78, 95% CI [-2.11, -1.45], $t(152) = -10.77$, $p < .001$; Std. beta = -2.10, 95% CI [-2.49, -1.72]), Pollock *vs* Frankenthaler (beta = -1.62, 95% CI [-1.95, -1.30], $t(152) = -9.81$, $p < .001$; Std. beta = -1.91, 95% CI [-2.30, -1.53]), Pollock *vs* Twombly (beta = -1.32, 95% CI [-1.65, -0.99], $t(152) = -7.99$, $p < .001$; Std. beta = -1.56, 95% CI [-1.94, -1.17]) and Pollock *vs* Mitchell (beta = -0.52, 95% CI [-0.84, -0.19], $t(152) = -3.12$, $p = 0.002$; Std. beta = -0.61, 95% CI [-0.99, -0.22]) are statistically significant and negative. The effect of Pollock *vs* Krasner (beta = -0.23, 95% CI [-0.56, 0.10], $t(152) = -1.39$, $p = 0.168$; Std. beta = -0.27, 95% CI [-0.66, 0.11]), which is statistically non-significant and negative.

Post hoc comparisons revealed that the art history relevant pairs that reported significant differences in complexity ratings were Kline vs Hartigan ($t(152) = -7.21$, $p <$

.0001) and Twombly vs Mitchell ($t(152) = -4.88$, $p = 0.0001$), where both men artists (Kline M = 2.22; Twombly M = 2.58) had significantly lower complexity ratings compared to their women colleagues (Hartigan M = 3.41; Mitchell M = 3.38). All post-hoc comparisons are available in the online Supplementary Materials: file [Complexity_post_hoc.pdf], [Experiment 3] folder.

Finally, it is interesting to report that even when artist identity had a significant interaction with male authorship likelihood, none of the art-historical relevant comparisons between women and men artists resulted to be significant. On the other hand, complexity was not influenced by male authorship likelihood, but only by artist identity. Post hoc comparisons revealed that the art history relevant pairs that reported significant differences in complexity ratings were Kline vs Hartigan ($t(152) = -7.21$, $p < .0001$) and Twombly vs Mitchell ($t(152) = -4.88$, $p = 0.0001$), where both men artists (Kline M = 2.22; Twombly M = 2.58) had significantly lower complexity ratings compared to their women colleagues (Hartigan M = 3.41; Mitchell M = 3.38).

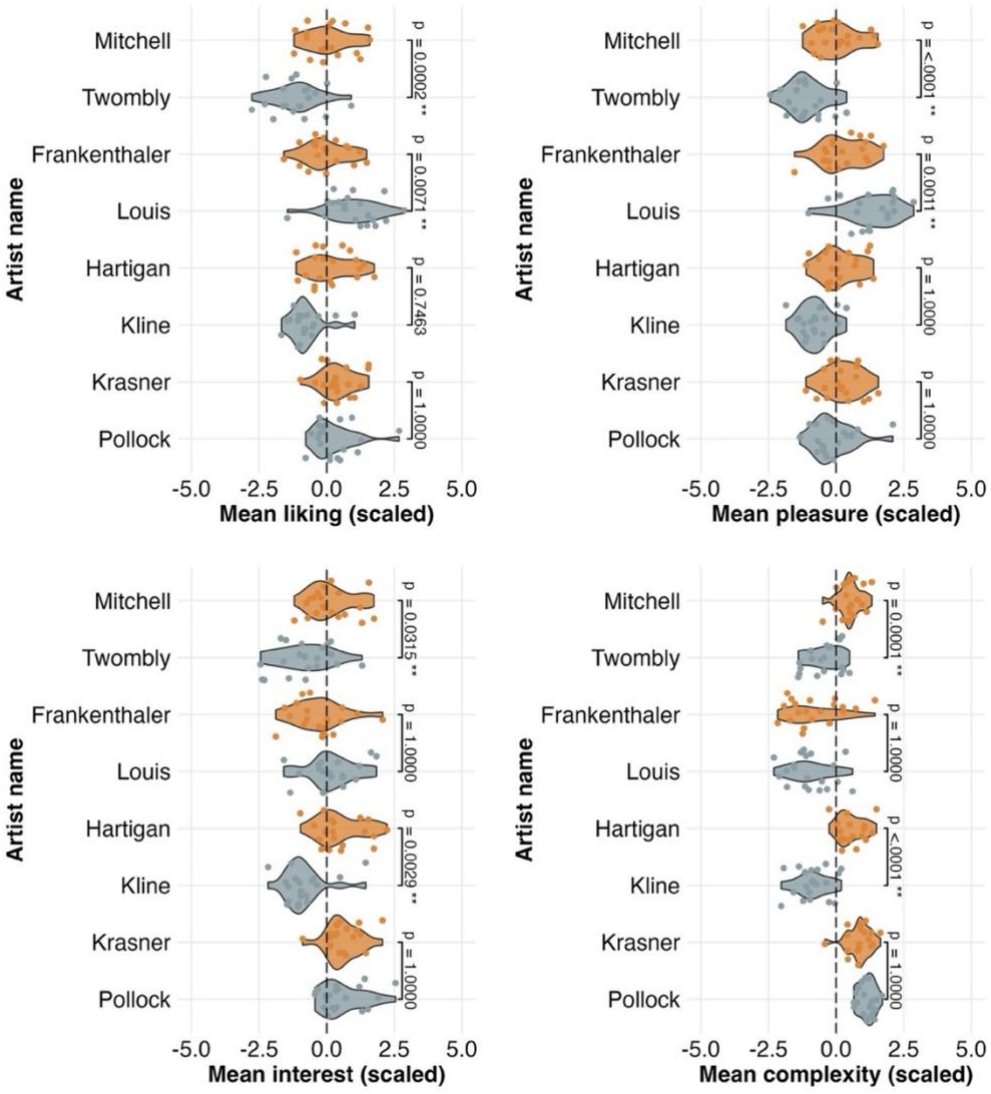

Fig. S4 *Violin plots of average value for four aesthetic appreciation variables (liking, pleasure, interest and complexity) by artist identity.. Women artists are represented in orange, while men artists are in light blue.*

SI References

1. Lenth R (2023). _emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.8.6, https://CRAN.R-project.org/package=emmeans

2. Lüdecke, Patil, Ben-Shachar, Wiernik, & Makowski (2022). easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting. CRAN. Available from https://easystats.github.io/easystats/