EXHIBIT 142

Bellaiche *et al.*
Cognitive Research: Principles and Implications (2023) 8:42
https://doi.org/10.1186/s41235-023-00499-6

Cognitive Research: Principles and Implications

ORIGINAL ARTICLE

Open Access

Humans versus AI: whether and why we prefer human-created compared to AI-created artwork

Lucas Bellaiche[1*], Rohin Shahi[1], Martin Harry Turpin[2], Anya Ragnhildstveit[3], Shawn Sprockett[4], Nathaniel Barr[5], Alexander Christensen[6] and Paul Seli[1]

Abstract

With the recent proliferation of advanced artificial intelligence (AI) models capable of mimicking human artworks, AI creations might soon replace products of human creativity, although skeptics argue that this outcome is unlikely. One possible reason this may be unlikely is that, independent of the physical properties of art, we place great value on the imbuement of the human experience in art. An interesting question, then, is whether and why people might prefer human-compared to AI-created artworks. To explore these questions, we manipulated the purported creator of pieces of art by randomly assigning a "Human-created" or "AI-created" label to paintings actually created by AI, and then assessed participants' judgements of the artworks across four rating criteria (Liking, Beauty, Profundity, and Worth). Study 1 found increased positive judgements for human- compared to AI-labelled art across all criteria. Study 2 aimed to replicate and extend Study 1 with additional ratings (Emotion, Story, Meaningful, Effort, and Time to create) intended to elucidate *why* people more-positively appraise Human-labelled artworks. The main findings from Study 1 were replicated, with narrativity (Story) and perceived effort behind artworks (Effort) moderating the label effects ("Human-created" vs. "AI-created"), but only for the sensory-level judgements (Liking, Beauty). Positive personal attitudes toward AI moderated label effects for more-communicative judgements (Profundity, Worth). These studies demonstrate that people tend to be negatively biased against AI-created artworks relative to purportedly human-created artwork, and suggest that knowledge of human engagement in the artistic process contributes positively to appraisals of art.

Keywords Aesthetics, Artificial intelligence, Visual art, Creativity, Judgements

*Correspondence:
Lucas Bellaiche
lucas.bellaiche@duke.edu
[1] Department of Psychology and Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708, USA
[2] Department of Psychology, University of Waterloo, Waterloo, ON, Canada
[3] Department of Psychiatry, University of Cambridge, Cambridge, UK
[4] MDes in Interaction Design Program, California College of the Arts, San Francisco, CA, USA
[5] School of Humanities and Creativity, Sheridan College, Oakville, ON, Canada
[6] Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, TN, USA

Springer Open

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

"We hear much these days about the remarkable new thinking machines. We are told that these machines can be made to take over much of men's thinking…eventually about the only economic value of brains left would be in the creative thinking of which they are capable…"

- J. P. Guilford, "Creativity" (1950), p. 446.

Introduction

Art is widely considered to be a uniquely human phenomenon. It can encapsulate and communicate our emotions, be used to express our individualistic and communal experiences, and serve as a social commentary on these experiences, all of which are elements that are commonly thought to be human-specific (Chatterjee, 2014). Yet, art is also a perceptual product that is engaged through the senses, sometimes independently of the experient's awareness of the inclusion (or lack thereof) of these human elements in artworks they evaluate. Here, an important distinction emerges between art as (1) a *purely physical stimulus* and (2) a *deeper communicative medium* of the human experience. The importance of this distinction lies in its ability to permit investigations of the separate and joint influences that these two conceptions of art may have on people's appraisals of art, which in turn contributes to our understanding of the factors and processes involved in aesthetic appraisals. And, with the recent development of highly advanced artificial intelligence (AI) models that can produce, as *purely physical stimuli*, high-quality artworks that are indiscernible from human-created artworks (Gangadharbatla, 2022; Johnson, 1997; Mazzone & Elgammal, 2019), the Computer Age in which we live has afforded considerable experimental control over these two conceptions of art. Here, across two studies, we explored the question of whether and why humans might tend to prefer ostensibly human-created artworks over AI-created artworks.

Although the process of appraising artworks undoubtedly involves a subjective element, consistent preferential patterns have nevertheless emerged across peoples' aesthetic judgements of such artworks (e.g., blue is the most preferred color, whereas yellow is most disliked [Bornstein, 1975; Komar & Melamid, 1999; Palmer et al., 2013]; and representational art is enjoyed over abstract art [Heinrichs & Cupchik, 1985; Kettlewell et al., 1990; Knapp & Wulff, 1963; Mastandrea et al., 2011, 2021]). However, the ways in which we judge art are often sensitive to factors in the environment, not just intrinsic preferences. One common finding across literatures of judgement and decision-making is that labels play an important role in people's general evaluations of things. For instance, people prefer wine they are falsely told is more expensive than other wine (Plassmann et al., 2008). The influence that a mere label has on our judgements is also made evident by research showing that, when participants are provided Coca-Cola either with or without its label, they report greater enjoyment of the labelled beverage than the non-labelled beverage, despite their otherwise identical compositions (McClure et al., 2004). As demonstrated by these studies, our judgements rely heavily on contextual information. And, perhaps unsurprisingly, the role that context plays in our appraisals extends to judgements of art as well. For example, when Newman and Bloom (2012) provided participants artworks labelled either as originals or as identical forgeries, participants preferred the originals over the forgeries: a finding that was taken to indicate that humans are sensitive to an authentic process of creative production as determined simply by a label. In other words, people consider non-sensory aspects of art, like context and background, in their judgements and evaluations of art (Blank et al., 1984; Chatterjee & Vartanian, 2016; Winner, 1982).

The influential role that labels play in our judgements is also made clear in contemporary lines of research focused on AI. The general finding emerging from this research is that, in many instances, the label of "AI" is taken to be a pejorative. For instance, Liu et al. (2022) conducted an experiment wherein participants were asked to read a number of emails and were told that some of the emails (but not others) were drafted with the help of an AI-based language model. Results indicated that, when participants were told that AI played a role in drafting the email, their reports of trust in the email writer decreased. This anti-AI bias seems especially apparent in domains that people assume to be human-specific (e.g., those concerning affect or creativity; Wilson, 2011). Research has only just begun to connect the spheres of aesthetics and AI, yet, the discussion of AI-produced creativity arose many decades ago. Indeed, in his 1950 Presidential Address to the American Psychological Association (APA), JP Guilford foreshadowed the development of AI in modern society, noting that, if advances in "thinking machines" (i.e., AI) reached their forecasted height, then "the only economic value of brains left would be the creative thinking of which they are capable" (Guilford, 1950). While Guildford's position is certainly provocative, he may have overestimated the extent to which the apparent 'last bastion' of human utility— "creativity"—is uniquely human; instead, it could be that, with recent advances in highly advanced "thinking machines," even our creative abilities could lose their economic value when competing with the abilities of AI.

Initial forays into the AI-art interaction seem to suggest that, not only is AI art sufficiently sensorily similar to human-created art that people fail to accurately discern its true creator (human or AI), but also that people tend

to derogate AI-created art as compared to human-created art (Chamberlain et al., 2018; Gangadharbatla, 2022; Mazzone & Elgammal, 2019). For example, Chamberlain et al. (2018) found that participants rated human-created artworks as higher in aesthetic value than AI-created artworks. Interestingly, however, this bias was malleable: indeed, when participants were shown videos of robots that were physically producing art, participants' anti-AI art judgements were reduced, suggesting that people may consider the level of artistic effort exerted by the creator as a marker of the quality of the produced artwork. Similar findings emerged in a more-recent study in which participants were shown, and asked to provide judgements of, two abstract paintings (Chiarella et al., 2022). While both paintings were in fact created by humans, by randomly assigning a label of "human-created" or "AI-created" to each painting, the researchers deceived participants into believing that one of the paintings was created by AI. Results of this study indicated that participants tended to prefer the painting that was labelled "human-created" relative to the painting labelled "AI-created." Notably, this preference for human over AI art is not specific to visual art. In fact, several recent studies have reported similar anti-AI findings in music (Shank et al., 2022), creative writing (Raj et al., 2023), dance (Darda & Cross, 2023), poetry (Köbis & Mossink, 2020), and non-art texts (Darda et al., 2023). This bias in aesthetic judgement is further intensified when judging more "human" aspects of art, like evoked emotion, suggesting the need for multiple judgement criteria in AI aesthetics research (Raj et al., 2023).

Critically, however, other studies have yielded results that are at odds with those from the aforementioned studies, finding little-to-no differences in evaluations of human- and AI-created artworks (Hong & Curran, 2019; Israfilzade, 2020; Xu et al., 2020). Additionally, though not an explicit investigation of aesthetic-judgements, per se, one study found that participants were equally likely to consider hypothetical products created by human and AI artists as "art" (Mikalonytė & Kneer, 2022). Put differently, people did not tend to perceive AI artists' hypothetical products as being lesser than human artists' hypothetical products. Given these mixed results, it is challenging to ascertain whether humans do in fact prefer human-created artworks over AI-created artworks, and, if so, why this is so. Thus, here, we examined this question with the intention of shedding more light on their answers.

The current studies

Across two studies, we sought to extend the extant—albeit sparse—literature on attitudes toward artworks created by humans as compared to artworks created by AI platforms. Our studies had three primary aims. First, while a handful of studies have examined the question of whether people differ with respect to their judgements of human-created versus AI-created artworks, at noted above, mixed results have been yielded, leaving the answer to this question unclear. Thus, by employing two large-sample ($Ns = 150$ and 151), highly powered, within-subjects studies including relatively large stimulus sets, we sought to provide greater clarity on the answer to this question. We hypothesized that humans will indeed show a preference for human-labelled art relative to AI-labelled art.

Second, we sought to identify some of the more-complex criteria that people might rely upon when making aesthetic judgements. Many aesthetic studies have used simple probes assessing general "liking" and "aesthetic beauty" to index art judgements (for a review, see Chatterjee & Cardillo, 2022). However, most models of aesthetics consider aesthetic judgements to be a multifaceted and/or hierarchical cascade of evaluations, implicating more-complex and elaborative processes in judgements of art that seek to understand meaning and communication through the piece (Berlyne, 1960, 1971; Chatterjee & Vartanian, 2016; Cupchik & Berlyn, 1979; Graf & Landwehr, 2015; Leder et al., 2004; Silvia, 2005). This is supported, in part, by the growing discussion of art not only as a bottom-up appraisal of only visual features that induces liking or judgements of beauty, but a device for communication and socio-epistemic value (see Sherman & Morrissey, 2017). This distinction considers human expression and, as such, is essential to investigate in aesthetic judgements of art created by AI. Thus, in Study 1, in addition to simply indexing Liking and Beauty, which we consider to be more-passive, surface-level appraisals in line with fluency models of aesthetics (see Leder et al., 2004; Graf & Landwehr, 2015), we also obtained assessments of how profound participants found each painting to be, and how much money they would (hypothetically) spend on each painting. These latter criteria require higher cognitive elaboration to determine communicative properties of art, as discussed in Graf and Landwehr (2015) and Sherman and Morrissey (2017). In Study 2, we further developed our probing procedure by, in addition to asking the four questions from Study 1 (i.e., Liking, Beauty, Profundity, Worth), asking participants to rate each painting on the following additional communicative attributes of the art to understand potential interactions between surface-level and communicative engagement processes: emotionality (how much the artwork evoked an emotion in the viewer; Emotion), narrativity (the degree to which the artwork portrayed an imagined narrative in the viewer; Story), meaningfulness (how personally meaningful the artwork was to the

viewer; Meaningful), perceived effort (how much effort the viewer thought went into the creation of the artwork; Effort), and perceived time (how much time the viewer thought went into the creation of the artwork; Time). We predicted that the anticipated preference for human-labelled art over AI-labelled art can be, at least partially, explained by a perceived lack of integration of the human experience in AI-created artworks as measured by these wide-ranging aesthetic criteria.

Third, given that aesthetic appraisals are personal and subjective (Chatterjee, 2014; Roseman & Evdokas, 2004), also imperative in their investigation is consideration of individual differences that might reliably predict these judgements. In this vein, past work has considered the personality trait of openness to experience and its interactions with artwork and other creative outputs (Kaufman, 2013; McCrae, 2007; McCrae & Greenberg, 2014), with Silvia et al. (2015) considering it "an essentially aesthetic trait" (p. 376). In addition, openness has been shown to influence judgements of art depending on painting type, such that people who are more open to experience tend to have an increased appreciation for abstract art compared to less-open participants (Feist & Brady, 2004). Other personality traits explored in aesthetics and creativity research include empathy and embodied cognitions (Freedberg & Gallese, 2007; Rusu, 2017), and beliefs of creative mindsets (Hass et al., 2016; Karwowski, 2014). To provide greater clarity on the possible influence of individual-differences measures on appraisals of artwork, in Study 1, we assessed age and scores on the cognitive reflection test (CRT), and in Study 2, age, CRT scores, openness to experience, personal attitudes toward AI, empathy, growth mindsets, and fixed mindsets (below, we outline our rationale for inclusion of these individual-differences measures).

Study 1

Method

Study 1 was approved by the University of Waterloo Research Ethics Board (31067). All data and stimuli can be found at https://osf.io/cgw8v/.

Participants

One-hundred and fifty participants, each with at least 100 approved human intelligence tasks and an approval rating of above 90%, were recruited through Amazon's Mechanical Turk. Power analyses were not performed a priori. Instead, we reviewed the most closely related studies we could find, finding that, in Study 1 of Chamberlain et al. (2018), 65 participants completed the study. To ensure our study was well-powered, we decided to nearly triple this sample size. Participants were told at sign-up that they would be evaluating pieces of art and filling out questionnaires meant to probe the way they think. Participants were directed to Qualtrics to complete the study, and then were debriefed following completion of the study and were compensated for their time. One participant was excluded for bad data (i.e., they used the same response throughout the entire study), resulting in 149 participants (M_{age} = 42.35, SD = 11.59; female = 65).

Materials

Thirty AI-created paintings, considered by the authors of this article to be of high quality, were taken from ArtBreeder, which is a machine learning website that produces art (see Fig. 1 for sample images). The images were open-source and



Fig. 1 Sample images (representational and abstract, respectively) taken from Artbreeder. See https://osf.io/cgw8v/ for all stimuli

pre-existing and were not created by the authors. While it might seem most reasonable to obtain artworks from both AI and humans (and not deceive participants), one possible problem with this approach is that there could be error introduced into the selection process. For instance, although one could do their best to ensure that any selected human and AI artworks are comparable, without first conducting a norming study first (across all of our measures), this outcome could not be verified. For this reason, we opted to present a stimulus set that was exclusively AI-created.

Of the 30 AI-created paintings, half (15) were representational (i.e., reflected an easily recognizable figure or object), and half (15) were abstract (i.e., they included partially or completely unrecognizable referents), as rated by the first author. This distinction was made in the present study because past results have repeatedly revealed that people tend to prefer representational over abstract paintings (Heinrichs & Cupchik, 1985; Kettlewell et al., 1990; Knapp & Wulff, 1963; Mastandrea et al., 2011, 2021). Here, we not only wanted to attempt to replicate this finding, but also to determine whether evaluations of different painting types (representational and abstract) differ as a function of whether the purported creator was a human or AI. Research on AI-art perceptions has primarily used abstract paintings as stimuli, which could limit our understanding of the true role of painting type in aesthetic judgements (Chiarella et al., 2022; Israfilzade, 2020). However, Chamberlain et al. (2018) presented varied painting types but found no interaction between painting creator and painting type on general aesthetic value ratings; with our extension of rating criteria, potentially different processes of judgements that may rely both on painting type and creator could be illuminated. Relatedly, Gangadharbatla (2022) found an increased willingness for people to believe that abstract art tends to be created by AI (compared to humans), whereas Chamberlain et al. (2018) found an increased willingness for people to believe that representational art tends to be created by humans. Thus, it seems that painting type does bear some relationship to the creator of the painting; to date, however, this is unclear.

Importantly, each image was presented in random order and had individual-level randomization of a label of "human-created" or "AI-created." In other words, different labels were assigned to different images across participants. Thus, on average, participants were presented 15 images with an AI label, and 15 images with a human label, even though all images were in fact AI-created.[1]

Participants were asked to rate each image on the following criteria: "How much do you like this image?" (Liking), "How beautiful/aesthetically pleasing is this image?" (Beauty), "How profound or meaningful is this image?" (Profundity), and "How much money would this work be worth?" (Worth). All criteria were answered on a 1–5 Likert scale ["Not at all"… "Very much"], except for Worth, which was answered on a 1–5 Likert scale with possible responses ["None at all"… "Worth quite a lot"]. In addition, for each image, participants were asked "Based on the label above, was this image created by a human or an artificial intelligence computer program?" (Label-Check). This question was used as an attention check. Incorrect responses to the Label-Check were excluded from data analysis, resulting in the removal of 113 trials (out of 4470 total trials across the 149 participants). No participants themselves were removed on the basis of this attention check, only trials.

Before rating the images, participants were asked to complete the 7-item cognitive reflection test (CRT; Frederick, 2005; Toplak et al., 2014). This extended CRT, which has been shown to have high internal consistency ($\alpha = 0.74$), consists of seven numeracy questions and is administered to assess an individual's cognitive ability to override instinctual responses for a correct answer to the problem (Campitelli & Gerrans, 2014). Thus, the CRT considers individual differences in quantitative skills and bias-overriding. Our rationale for including the CRT was to determine whether CRT scores predict different ratings across human- and AI-created artwork: If we do observe an anti-AI bias—with more-positive judgements for human artworks than AI artworks—then one possibility is that individuals who have a greater ability to override their intuitions may show less of a bias, given that this bias may be the result of an automatic (intuitive) response to devalue artworks created by AI. Notably, both CRT and age were included as exploratory variables.

Procedure

Participants first provided informed consent following a description of the nature of the study. Next, they were given two bot checks (zero participants failed these checks, and all of them therefore moved forward with the study). Participants were then directed to complete demographic information, followed by the CRT. When finished with these tasks, participants received instructions for the remaining portion of the survey, which asked them to rate the 30 AI-created paintings on Liking, Beauty, Profundity, and Worth, with randomized labels of AI- or human-created. Following ratings, participants were debriefed on the purpose of the study, once again provided consent for their data to be used for analyses, and were redirected for compensation.

[1] Here, we use the term "anti-AI bias." Of note, however, is that one could reasonably refer to the same bias as a "pro-human bias." For clarity, and to allow consistency with similar terminology used in previous research on the topic (Chamberlain et al., 2018; Chiarella et al., 2022), we continue to use the term "anti-AI bias." We thank a reviewer for drawing our attention to this issue of terminology.

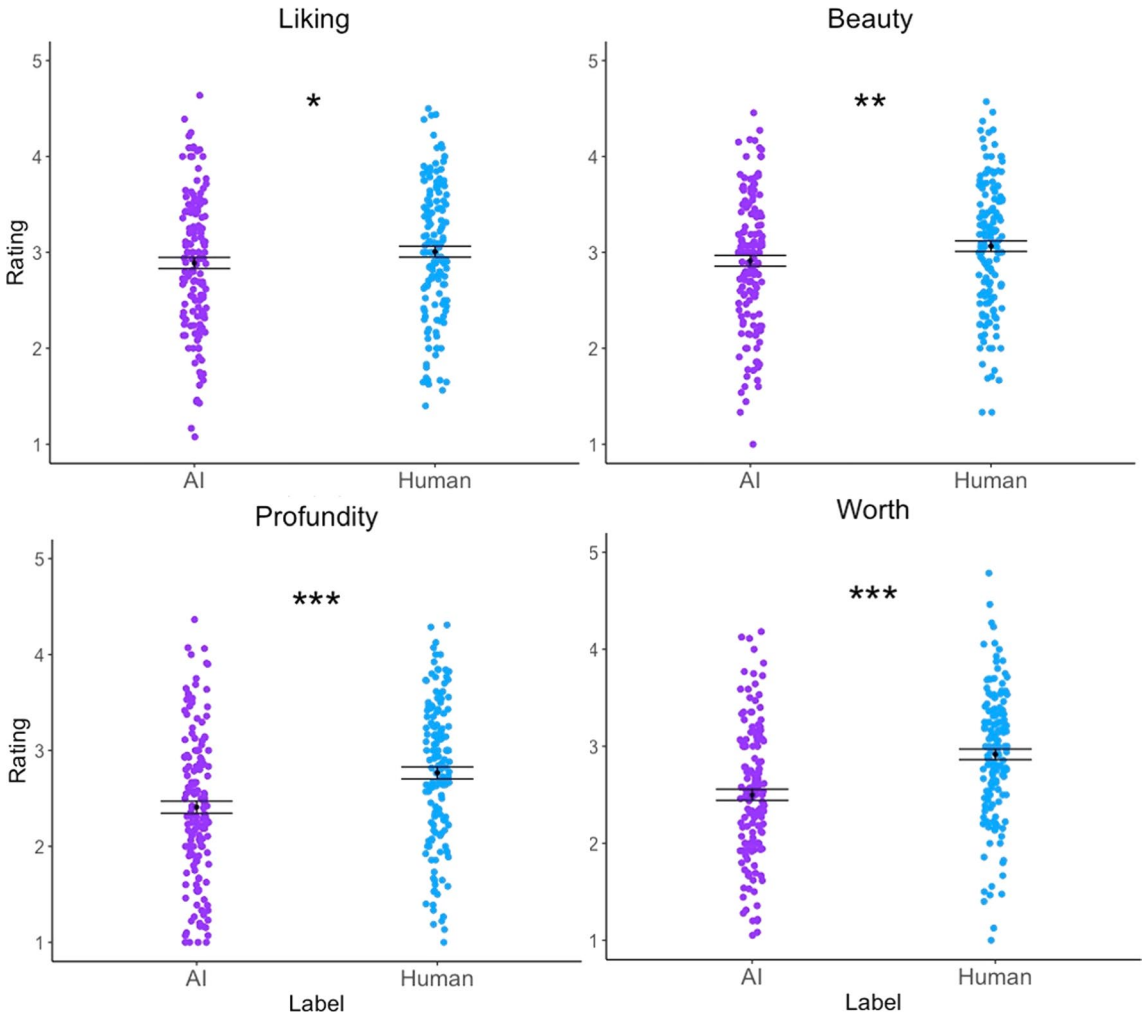


Fig. 2 Distribution of values on artworks associated with AI and Human for each criterion (Liking, Beauty, Profundity, Worth). Error bars represent standard error of the mean. Following multiple-comparison corrections, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Study 1 results

Paired *t*-tests with Bonferroni corrections show increased ratings for human-labelled over AI-labelled art for all four criteria: Liking ($t(148) = 2.644$, $p = 0.036$, $d = 0.17$), Beauty ($t(148) = 3.499$, $p = 0.002$, $d = 0.22$), Profundity ($t(148) = 7.725$, $p < 0.001$, $d = 0.47$), and Worth ($t(148) = 10.042$, $p < 0.001$, $d = 0.61$) (Fig. 2). In addition, for each criterion, a difference score between average human-labelled art rating and average AI-labelled art rating was calculated for each participant and was plotted against the participant's CRT score. Only a relationship between CRT and Beauty difference scores emerged significant, with lower CRT scores associated with higher Beauty difference scores (that is, higher average Beauty judgements were given for human-labelled than AI-labelled art; $r = -0.17$, $p = 0.042$).[2] No such significant relationships emerged elsewhere (Liking: $p = 0.07$, Profundity: $p = 0.95$, Worth: $p = 0.37$). The same difference scores were also plotted against age; again, no significant relationships emerged ($p = 0.29$, $p = 0.20$, $p = 0.26$, $p = 0.59$). A significant relationship emerged between age and CRT performance ($r = 0.23$, $p = 0.004$), such that older participants performed better on the CRT.

Lastly, exploratory linear mixed-effects models using maximum likelihood estimation were performed for each of the four main criteria as the outcome variables: Liking, Beauty, Profundity, and Worth. Across models,

[2] ANCOVAs controlling for CRT produced the same results, such that images with the "human" label elicited increased ratings than those with an "AI" label. Only for Beauty (as in our difference-score analyses) did CRT moderate this main effect marginally ($F(1, 147) = 4.200$, $\eta^2 = .004$, $p = .042$).

participants and paintings were random effects, resulting in crossed random effects models. For all models, our manipulated variables of the art's Label (AI or Human) and Painting Type (abstract or representational) were included, and we tested a potential interaction between Label and Painting Type. Across all criteria, a significant main effect of Painting Type emerged, such that participants preferred representational over abstract paintings regardless of Label (p's < 0.001). For Liking, a non-significant main effect of Label ($p = 0.057$) was qualified by a significant interaction between Painting Type and Label ($p = 0.041$). Additionally, for Profundity, a main effect of Label ($p < 0.001$) was qualified by a significant interaction between Painting Type and Label ($p = 0.036$). When a painting is representational, art is liked more and found more profound when it is created by a human as opposed to AI. No significant interaction between Painting Type and Label emerged in Beauty ($p = 0.21$) or Worth ($p = 0.18$).

Study 1 discussion

Across all four rating criteria (Liking, Beauty, Profundity, and Worth), we found an anti-AI bias, with participants showing a preference for art labelled as "human-created" compared to "AI-created." This finding is particularly important given the within-subjects design of our study, its relatively large sample size, range of rating criteria, range of painting types, and randomization of labels across a large set of paintings.

Study 1 primarily served as a test of whether an anti-AI-art bias exists across a range of aesthetic judgement criteria. Given our relatively large sample, we were confident in the findings of this study, which, on the whole, mirrored findings of Chamberlain et al. (2018). However, extending these results, here we showed that this pattern of judgements for AI-art biases is reflected across appraisal processes outside of general aesthetic preferences. For instance, not only is AI-art liked less, but it is also viewed as less worthy and less profound, which may have interesting implications for the ways in which people will consume AI-art in the future.

In addition, though some studies also used a randomization of labels (as in the present study; e.g., Chiarella et al., 2022), our study was unique in its use of only AI-created images, across many more stimuli, in a within-subjects design. Statistically, this ensured higher power and increased confidence in our findings. In other words, participants were given the opportunity to judge both human- and AI-labelled artworks, and this permitted a sounder comparison of participants' judgements across the two sources. Perhaps most importantly, though, was that all artworks were from AI in actuality. This underscores the fact that participants are generally incapable of noticing what source created what paintings—as supported by Gangadharbatla (2022) and Chamberlain et al. (2018)—which ultimately reflects the noteworthy quality of AI-art today. Though we did not explicitly ask participants for their confidence in the accuracy of our labels, our decision to exclusively select AI-created stimuli enabled us to dissociate top-down from bottom-up processing in these aesthetic judgements. That is, actual human versus actual AI images may differ in certain visual, bottom-up qualities.[3] Thus, by exclusively selecting AI artwork and randomizing labels on a random subset of paintings, we ensured that any observed effects were necessarily top-down and isolated to source manipulation, as opposed to bottom-up, or driven by potential inherent differences across human- and AI-created artworks.

Somewhat surprising was the lack of statistically significant associations between our individual-differences measures (CRT and age) with the difference scores of human-AI ratings for each participant, which might be expected to track an AI bias. Although lower cognitive reflection scores were associated with higher Beauty preferences of human than AI art, no other significant relationship emerged between our individual-differences measures and the human-AI difference scores. Past work with the CRT has found negative relationships between CRT scores and perceived profundity of randomly generated statements (Pennycook et al., 2015), implying a relationship between cognitive reflection traits and subjective judgements of profundity that could extend to aesthetics. However, in Study 1, we found no support for this relationship. Given that the CRT assesses intuition-overriding in mathematical domains, we can conclude that (1) anti-AI biases largely do not rely on quantitative skills, (2) the anti-AI bias is not an intuitive response as probed by the mathematical problems in the CRT, and/or (3) we need more-sophisticated statistical analyses and/or power to uncover the true underlying relationships (if any) among these measures. In addition, the lack of statistically significant relationships between age and judgements of art was somewhat surprising. Though exploratory, this is at odds with some previous aesthetic studies that have reported age-dependent differences in judgements of art. Specifically, Mockros (1993) reported that general aesthetic judgements were rated higher by novice professional adults than novice undergraduates. More research is warranted on this topic given that we found no such age effect, which perhaps implies that younger participants—with their lives being more heavily dominated by AI than older participants—do not have

[3] The only way to ensure they are not would be to run a norming study; to our knowledge, no such study has yet been run. Future work is therefore merited here.

different views on the creative products of AI than those of older adults.

Study 2

The results from Study 1 provide initial insight into participant judgements of art that is believed to be generated either by AI algorithms or by humans. While not reliably predictable by individual differences such as age or CRT scores, a bias against AI art emerged. This bias, however, clearly depended on the criterion used to assess the artwork. Certainly, individuals do not just consider measures of Liking, Beauty, Profundity, and Worth when assessing artwork, but a host of other engagement processes.

This nuance to a general "anti-AI bias" matches with contemporary aesthetics models that argue that aesthetic judgements are rather complex. Indeed, largely pioneered by David Berlyne's new experimental aesthetics (1960, 1971), art appreciation has been argued to be a consequence of elaborative appraisals of criteria like novelty, ambiguity, and complexity (which he deemed "collative" properties; Cupchik & Berlyn, 1979). Other emerging models have since provided more nuance to Berlyne's traditional behaviorist approach (see Berlyne, 1975; Silvia, 2005), including more cognitive principles, and an increased understanding of the processes people utilize to engage with art. These models all agree in their proposal that humans engage in a multi-process evaluation of art, considering both sensory and non-sensory aspects (e.g., the Aesthetic Triad, Chatterjee & Vartanian, 2016; the pleasure-interest model, Graf & Landwehr, 2015; the information-processing model, Leder et al., 2004). Accordingly, many aesthetic properties—some more literal and surface-level, some non-sensory and more elaborative—can be probed during interactions with art. For instance, findings from Chamberlain et al. (2018) show that participants specifically cite that brush-strokes and other surface-level properties of the artistic process influence their evaluations of the final product. Moreover, when viewing anthropomorphized videos of robots painting, participants show greater appreciation for computer-generated art than when no anthropomorphized video of the robot is provided. This mirrors findings by Hong et al. (2022), who found that anthropomorphization of an AI-music algorithm led to higher acceptance of the algorithm as a true "musician," which in turn led to increased aesthetic appreciation. Collectively, these studies suggest that there may be an implicit role of effortfulness or embodiment behind peoples' evaluations of art. More specifically, it may be that peoples' preferences for art increase as their beliefs about the amount of effort that went into creating a piece of art increase.

As informed by multi-level models as above, including additional judgement criteria could help us to shed further light on the reasons as to why people are making the aesthetic judgements they're making. Thus, in addition to Liking, Beauty, Profundity, and Worth, in Study 2 we also included the following judgement criteria: emotionality (Emotion), perceived narrativity (Story), personal meaning (Meaningful), perceived effort (Effort), and estimated time to create (Time) as additional criteria for judgement. Importantly, while these now nine criteria may or may not behave in similar ways to one another, some may moderate ratings of others. In this way, we could gain insight on how levels of processing act interpedently or influence one another in this new sphere of aesthetic judgements with art created by non-humans. We specifically included these criteria given emotion's prominent role in aesthetics and creativity (Silvia, 2005), the use of narratives as engines of consumption in other forms of art (a growing literature in music cognition investigates narratives in response to music, e.g., Margulis et al., 2019; McAuley et al., 2021), the act of deriving meaning from artwork that is often posited in models of aesthetics (e.g., Leder et al., 2004; Pelowski et al., 2017), and that effort and time involved in creation of a product is often a heuristic for quality (Kruger et al., 2004).

In addition, we extended the battery of individual-differences measures given our mostly null results from the individual-differences measures in Study 1. While we choose to keep both age and CRT scores as potential individual differences, we also included empathy skills (to determine if participants differ in judgement based on the ability to empathize with other agents, including AI, perhaps explaining AI-art judgements), openness to experience (given its role in aesthetic encounters; Kaufman, 2013; McCrae, 2007; McCrae & Greenberg, 2014; Silvia et al., 2015), personal attitudes toward AI (which we hypothesized could predict judgements of artwork made by AI), and facets of the creative mindset scale (CMS; Karwowski, 2014), which include growth and fixed mindsets. The CMS specifically asks about views of who can produce products of creativity, and we thus deemed it important to include in this study to determine if growth mindsets lead to higher AI-art ratings as we hypothesized. In addition, to detect more sensitive relationships between both individual-differences measures and judgement criteria, we aimed to use more-sophisticated statistical modeling with a pre-registered design. In sum, through implementing linear mixed models, we sought to determine whether we would replicate and explain the anti-AI effect with wider criteria of judgements—to parallel common multi-level processing models of aesthetics—and an extension of individual-differences measures.

Method

Study 2's design and analysis plan were pre-registered. Study 2 was approved by the Duke University Campus Institutional Review Board (2022-0534). All data and stimuli, and the pre-registration, can be found at https://osf.io/cgw8v/.

Participants

One-hundred and fifty-one participants were recruited through Prolific.[4] Only native English speakers with at least a 90% approval rate were permitted to enroll in the study. Like Study 1, they were told at sign-up that they would be evaluating pieces of art and filling out questionnaires meant to probe the way they think. Participants were debriefed following completion of the study and were compensated for their time. As per our pre-registration, data from two participants were excluded for failing an attention check (age and birth-year needed to correspond appropriately), and one participant was excluded for bad data, resulting in 148 participants (M_{age} = 36.66, SD = 12.73; female = 74, non-binary = 2).

Materials

The same 30 images from Study 1 were presented in random order in Study 2, again with a randomly assigned label of "human-created" or "AI-created." Ratings were done on the same criteria: Liking, Beauty, Profundity, Worth; note that the phrasing for the Profundity question in this study changed from "How profound or meaningful is this image?" to "How profound is this artwork?" for simplicity. Participants rated the images on five additional criteria as well: "To what extent does this artwork elicit an emotional response in you?" (Emotion), "To what extent can you imagine a story communicated through this artwork?" (Story), "To what extent do you find this artwork personally meaningful?" (Meaningful), "How much effort do you believe was involved in making this artwork?" (Effort), and "How much time do you believe it took to create this piece?" (Time). All questions were on a 1–5 Likert ["None/not at all"… "Very much"] except Time, which was a free-response question that allowed for numeric entry in a "Days," "Hours," and "Minutes" field. Participants were told to simply enter "0" if a certain field did not apply. Lastly, for each image, participants were asked the Label-Check question from Study 1 for trial-based exclusions: "Based on the label above this image, was this image created by a human or an artificial intelligence computer program?" Those who did not provide the correct answer had that trial eliminated, resulting in 114 trials (of 4440 total trials across the 148 participants) excluded from analyses. As in Study 1, no participants were removed on the basis of this attention check, only trials. This specific exclusion criterion was not specified in our original pre-registration plan, though, for purposes of consistency, it follows the procedure from Study 1.

Before the ratings, participants filled out various questionnaires. Participants completed the same CRT as in Study 1. In addition, to probe trait-level empathy skills, participants completed the Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009). The TEQ has been demonstrated to be a highly internally consistent (α = 0.87) and reliable (r = 0.81) self-report empathy measure. The 16-item questionnaire was completed on a 1–5 Likert ["Never"… "Always"], and responses were reverse-scored when appropriate and summed for a final empathy score.

Participants also completed the General Attitudes toward Artificial Intelligence Scale (GAAIS) to assess positive and negative sentiments of AI in general (Schepman & Rodway, 2020). The 20-item GAAIS has been fully validated and found to report a consistent positive subscale (α = 0.88) and negative subscale (α = 0.82), with strong convergent validity with other AI-related measures (Schepman & Rodway, 2022). Scores for each subscore were aggregated through summation, producing a "Positive Attitude" score and a "Negative Attitude" score.

To probe participant beliefs about creativity, we administered the Creative Mindset Scale (CMS: Karwowski, 2014). The CMS assesses how much an individual believes creativity can be developed and trained through time (i.e., "growth" mindset) and how much they believe creativity is static and innate (i.e., "fixed" mindset). Participants rated the 10 items of the CMS on a 1–5 Likert Scale ["Definitely not"… "Definitely yes"]. Scores for each mindset were aggregated through summation, producing a "Growth" value and a "Fixed" value.

Participants also completed the 12-item "Openness to Experience" subset of the NEO-Five Factor Personality Inventory. Though the full NEO-FFI can also assess Consciousness, Extraversion, Agreeableness, and Neuroticism, we chose to implement only the Openness subset given its previous specific involvement in aesthetic encounters (Kaufman, 2013; McCrae, 2007; McCrae & Greenberg, 2014; Silvia et al., 2015), and to reduce study length. Items were answered on a 1–5 Likert scale ["Strongly disagree"… "Strongly agree"] and were reverse-scored when appropriate and aggregated through summing.

Lastly, in an effort to guide future research, we had participants answer a series of exploratory, free-response and single-item questions. These data are listed in our

[4] We aimed to recruit one-hundred and fifty participants, as in Study 1 and according to our pre-registration. However, Prolific mistakenly labelled one participant's task as "unfinished," thereby leading to the recruitment of an additional participant.

pre-registration as "exploratory," and, for the sake of clarity and brevity, they are not reported here.

Procedure

Upon direction to the Qualtrics survey, participants gave informed consent on the nature of the study. Then, they were given two bot checks; no participant failed, and all moved forward with the study. Participants were then directed to complete demographic information, followed by the CRT, TEQ, openness from the NEO-FFI, CMS, GAAIS, and additional single-item questions, in that order. Importantly, participants were specifically asked their age in the demographic questionnaire, and at the end of the study, were asked their year of birth; this served as an attention check. As per our pre-registration, data from two participants with incongruent birth years and ages were excluded from our analyses. Following questionnaires, participants completed the ratings of the images, after which they were debriefed and redirected to Prolific.

Statistical analyses

Linear mixed-effects models using maximum likelihood estimation were estimated in *R* (version 4.2.0; *R* Core Team, 2022) using the {lme4} package (version 1.1.33; Bates et al., 2015). A mixed effect model was estimated for each of the four main criteria from Study 1 as the outcome variables: Liking, Beauty, Profundity, and Worth. The linear mixed-effects models diverged from our pre-registration with the use of frequentist rather than Bayesian models.[5] Power analyses were not performed a priori; instead, the sample size was determined using the same procedure reported in our Study 1.

For all models, our manipulated variables of the art's Label (AI or Human) and Painting Type (abstract or representational) were included. Both Label and Painting Type were fixed-effect variables. We did not include the interaction between Label and Painting Type in our reports because this term was not significant in any of the models. Across models, participants and paintings were random effects, resulting in crossed random effects models. For participants, we included the random intercept and Label slope to model people's differences in outcome related to Label. For paintings, we only included the random intercept term to account for differences in the level of the outcome for each painting.

[5] Because of the ordinal nature of the outcome data, we did not presume continuous assumptions (e.g., normal residuals) would be met during pre-registration. And, because ordinal data with crossed random effects demand Bayesian models due to the availability of R packages, we thus pre-registered Bayesian models. However, once the data were collected and revealed normal residuals, we could and did run a (frequentist) crossed random effects model. Thus, no Bayesian models were ever run.

For Model 1, we focused on individual-differences measures to identify characteristics of who were related to our outcome variables. All individual-differences measures were continuous (i.e., sum scores) fixed effect variables. Each individual-difference measure was mean-centered to increase the interpretability of the model (e.g., each intercept represented the mean of each outcome). Interactions between Label and all individual-differences measures were specified in this model.

For Model 2, we added the five additional criteria (Emotion, Story, Meaningful, Effort, and Time), as well as their respective interactions with Label, as predictors to the model. This model represented aspects of why outcomes might be rated differently. Time was log-transformed due to its logarithmic distribution. All additional criteria were level-1 or within-painting variables. Since Model 2 represents the full model, the equation is presented here:

$$
\begin{aligned}
\boldsymbol{L}1 : Y_{ij} = {} & \beta_{0j} + \beta_{1j} PaintingType_{ij} + \beta_{2j} Label_{ij} \\
& + \beta_{3j} Openness_{ij} + \beta_{4j} PositiveAI_{ij} \\
& + \beta_{5j} NegativeAI_{ij} + \beta_{6j} GrowthMindset_{ij} \\
& + \beta_{7j} FixedMindset_{ij} + \beta_{8j} CorrectCRT_{ij} \\
& + \beta_{9j} Age_{ij} + \beta_{10j} Emotion_{ij} + \beta_{11j} Story_{ij} \\
& + \beta_{12j} Meaningful_{ij} + \beta_{13j} Effort_{ij} \\
& + \beta_{14j} \log Time_{ij} + \beta_{15j} Label_{ij} \times Openness_{ij} \\
& + \beta_{16j} Label_{ij} \times PositiveAI_{ij} + \beta_{17j} Label_{ij} \\
& \times NegativeAI_{ij} + \beta_{18j} Label_{ij} \\
& \times GrowthMindset_{ij} + \beta_{19j} Label_{ij} \\
& \times FixedMindset_{ij} + \beta_{20j} Label_{ij} \\
& \times CorrectCRT_{ij} + \beta_{21j} Label_{ij} \times Age_{ij} \\
& + \beta_{22j} Label_{ij} \times Emotion_{ij} + \beta_{23j} Label_{ij} \\
& \times Story_{ij} + \beta_{24j} Label_{ij} \times Meaningful_{ij} \\
& + \beta_{25j} Label_{ij} \times Effort_{ij} + \beta_{26j} Label_{ij} \\
& \times \log Time_{ij} + Participant_{oj\,i} \\
& + Painting_{1j} + \epsilon_{ij}
\end{aligned}
$$

$$
\boldsymbol{L}2 : Participant_{0j} = \gamma_{00} + \gamma_{01} Label_{oj} + u_{0j}
$$

$$
\boldsymbol{L}2 : Painting_{1j} = \gamma_{10} + u_{1j}
$$

For Model 3, we included only the significant predictors from Model 2. For each model, we report model fit using Akaike's information criterion (AIC) and variance explained using marginal R^2 (fixed effects only), intraclass correlation (ICC; random effects only), and conditional R^2 (all effects combined). For brevity, we focus our report on Model 3 but provide all models' results in Tables 1–4 to be comprehensive.

Study 2 results

Liking

Results for the Liking models are reported in Table 1. There were two interactions involving Label: Story ($b = -0.09$, $p < 0.001$) and Effort ($b = 0.12$, $p < 0.001$). For Story, Liking was greater when Label was "AI" than when Label was "human" and this difference increased as Story increased (Fig. 3). Based on Fig. 3 and the difference between the conditional means, for low-to-moderate values of Effort (ratings = 1–4; Fig. 3), Liking was greater when Label was "AI" than when Label was "human." For high values of Effort (rating = 5), Liking was greater when Label was "human" than when Label was "AI." There were main effects for Emotion ($b = 0.32$, $p < 0.001$), Meaningful ($b = 0.24$, $p < 0.001$), and Painting Type ($b = 0.28$, $p < 0.001$). For Emotion and Meaningful, Liking increased as these criteria increased. For Painting Type, Liking was greater for representational paintings than abstract paintings. The fixed effects of the model explained 57.9% of the variance, the random effects explained 32.0% of the variance, and all effects explained 71.2% of the variance.

Beauty

Results for the Beauty models are reported in Table 2. Similar to Liking, there were interactions involving Label: Story ($b = -0.08$, $p < 0.001$) and Effort ($b = 0.10$, $p < 0.001$). The effects similarly parallel Liking in that Beauty was greater when Label was "AI" than when Label was "human" and this difference increased as Story increased and for low-to-moderate values of Effort (ratings = 1–4; Fig. 3), Beauty was greater when Label was "AI" than when Label was "human" whereas the reverse was found for high values of Effort (rating = 5, Fig. 3). There was an additional interaction between Label and Openness to Experience ($b = -0.01$, $p = 0.003$) such that well below average Openness to Experience (−20) had higher Beauty for the human Label than the AI Label but this difference reversed with below average (−10) Openness to Experience and higher such that Beauty was higher for the AI Label than the human Label. Similar to Liking, there were main effects for Emotion ($b = 0.23$, $p < 0.001$), Meaningful ($b = 0.23$, $p < 0.001$), and Painting Type ($b = 0.32$, $p < 0.001$) such that Beauty increased as Emotion and Meaningful increased and Beauty was greater for representational paintings than abstract paintings. There was an additional main effect of Age ($b = -0.01$, $p = 0.007$) such that people younger than average (-20) rated Beauty higher than people older than average (40). The fixed effects of the model explained 53.1% of the variance, the random effects explained 34.0% of the variance, and all effects explained 69.0% of the variance.

Table 1 Estimates and associated CIs predicting "Liking" judgements

Liking	Model 1	Model 2	Model 3
Fixed effects [95% CI]			
(Intercept)	***2.38 [2.17, 2.60]***	***0.55 [0.38, 0.71]***	***0.57 [0.41, 0.73]***
Label (Human = 1)	***0.24 [0.15, 0.32]***	*− 0.26 [− 0.45, − 0.07]*	***− 0.28 [− 0.43, − 0.13]***
Painting Type (Representational = 1)	***0.75 [0.48, 1.01]***	***0.28 [0.13, 0.44]***	***0.28 [0.13, 0.44]***
Openness	0.01 [− 0.01, 0.03]	0.01 [− 0.00, 0.03]	–
Positive AI	*0.03 [0.01, 0.04]*	0.00 [− 0.01, 0.01]	–
Negative AI	0.01 [− 0.01, 0.03]	0.01 [− 0.00, 0.02]	–
Growth	*0.06 [0.01, 0.11]*	0.01 [− 0.03, 0.04]	–
Fixed	*0.06 [0.02, 0.09]*	0.02 [− 0.01, 0.04]	–
Empathy	0.00 [− 0.01, 0.02]	− 0.00 [− 0.01, 0.01]	–
CRT	− 0.04 [− 0.09, 0.01]	0.01 [− 0.03, 0.04]	–
Age	− 0.00 [− 0.01, 0.01]	− 0.01 [− 0.01, 0.00]	–
Emotion	–	***0.34 [0.29, 0.39]***	***0.32 [0.29, 0.36]***
Story	–	***0.23 [0.19, 0.27]***	***0.23 [0.19, 0.27]***
Meaningful	–	***0.20 [0.14, 0.25]***	***0.24 [0.20, 0.27]***
Effort	–	***0.11 [0.06, 0.15]***	***0.11 [0.07, 0.15]***
Time (log)	–	0.01 [− 0.00, 0.03]	–
Label × Openness	0.00 [− 0.01, 0.02]	− 0.01 [− 0.02, 0.00]	–
Label × Positive AI	*− 0.02 [− 0.03, − 0.00]*	− 0.01 [− 0.01, 0.00]	–
Label × Negative AI	− 0.01 [− 0.02, 0.01]	0.00 [− 0.01, 0.01]	–
Label × Growth	0.03 [− 0.01, 0.07]	0.01 [− 0.02, 0.04]	–
Label × Fixed	0.01 [− 0.02, 0.04]	0.01 [− 0.01, 0.03]	–
Label × Empathy	0.00 [− 0.01, 0.01]	0.00 [− 0.01, 0.01]	–
Label × CRT	− 0.03 [− 0.06, 0.01]	0.00 [− 0.03, 0.03]	–
Label × Age	− 0.00 [− 0.01, 0.01]	− 0.00 [− 0.01, 0.00]	–
Label × Emotion	–	− 0.04 [− 0.10, 0.03]	–
Label × Story	–	***− 0.10 [− 0.15, − 0.04]***	***− 0.09 [− 0.14, − 0.05]***
Label × Meaningful	–	0.06 [− 0.00, 0.13]	–

Table 1 (continued)

Liking	Model 1	Model 2	Model 3
Label × Effort	–	***0.12 [0.06, 0.18]***	***0.12 [0.07, 0.17]***
Label × Time (log)	–	−0.01 [−0.04, 0.01]	–
Random effects			
Participant (Intercept)	0.44	0.19	0.21
Label (Slope)	0.15	0.05	0.06
Painting (Intercept)	0.13	0.04	0.04
Residual	0.95	0.53	0.53
Model			
Marginal	0.16	0.59	0.58
ICC	0.37	0.30	0.32
Conditional	0.47	0.72	0.71
AIC	12,819.13	10,300.48	10,126.86

Note. CI = confidence interval, Marginal = variance explained by fixed effects, ICC = intraclass correlation or variance explained by random effects, Conditional = variance explained by fixed and random effects, AIC = Akaike Information Criterion. Italicized text = $p < 0.05$, bolded text = $p < 0.001$

Profundity

Results for the Profundity models are reported in Table 3. There was one interaction involving Label and Positive AI attitudes ($b = -0.01$, $p = 0.016$) such that, based on Fig. 4 and the difference between the conditional means, there was slightly higher Profundity when Label was "AI" and Positive AI attitudes were well below average (−20; Fig. 4). This difference grew larger as Positive AI attitudes increased to well above average (20). There were main effects for all five additional criteria (Emotion: $b = 0.22$, $p < 0.001$; Story: $b = 0.19$, $p < 0.001$; Meaningful: $b = 0.26$, $p < 0.001$; Effort: $b = 0.19$, $p < 0.001$; Time (log): $b = 0.02$, $p = 0.011$) such that Profundity increased as each criterion increased. For both growth ($b = 0.04$, $p = 0.005$) and fixed ($b = 0.02$, $p = 0.017$) mindsets, Profundity increased as these characteristics went from well below average (−10 and −8, respectively) to well above average (6 and 10, respectively). The fixed effects of the model explained 63.5% of the variance, the random effects explained

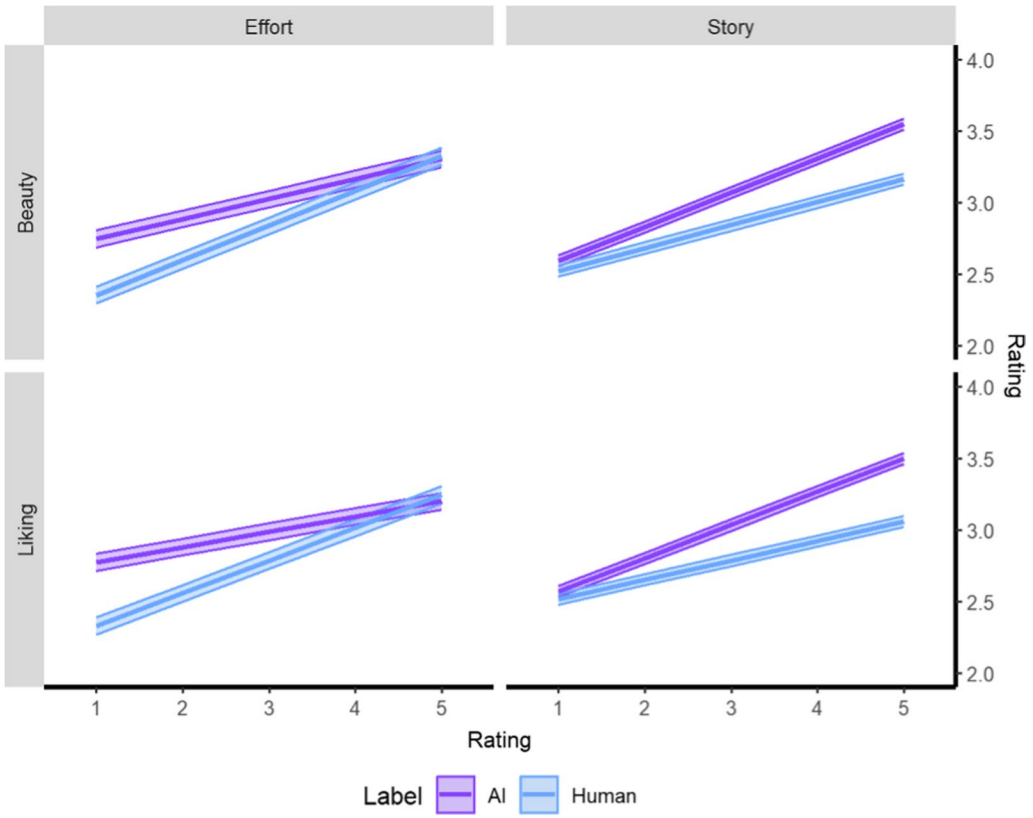


Fig. 3 Interaction plots showing moderations by Effort and Story on Liking and Beauty, two sensory-level criteria. Confidence bands represent the 95% confidence interval

Case 5:24-cv-03811-EKL Document 842-146 Filed 09/22/26 Page 14 of 23 PageID #: 99830

Table 2 Estimates and associated CIs predicting "Beauty" judgements

Beauty	Model 1	Model 2	Model 3
Fixed effects [95% CI]			
(Intercept)	***2.43 [2.20, 2.67]***	***0.75 [0.56, 0.95]***	***0.74 [0.55, 0.93]***
Label (Human = 1)	***0.23 [0.15, 0.31]***	***− 0.37 [− 0.57, − 0.18]***	***− 0.28 [− 0.43, − 0.13]***
Painting Type (Representational = 1)	***0.75 [0.46, 1.05]***	*0.32 [0.11, 0.52]*	*0.32 [0.11, 0.52]*
Openness	0.01 [− 0.01, 0.03]	*0.02 [0.00, 0.03]*	0.01 [− 0.00, 0.02]
Positive AI	*0.02 [0.01, 0.03]*	0.00 [− 0.01, 0.01]	–
Negative AI	0.01 [− 0.01, 0.03]	0.01 [− 0.00, 0.03]	–
Growth	0.04 [− 0.01, 0.09]	− 0.01 [− 0.05, 0.02]	–
Fixed	*0.05 [0.01, 0.08]*	0.01 [− 0.02, 0.04]	–
Empathy	− 0.00 [− 0.02, 0.01]	− 0.01 [− 0.02, 0.00]	–
CRT	− 0.04 [− 0.06, 0.04]	0.03 [− 0.00, 0.07]	–
Age	− 0.01 [− 0.02, 0.01]	*− 0.01 [− 0.02, − 0.00]*	*− 0.01 [− 0.01, − 0.00]*
Emotion	–	***0.24 [0.19, 0.29]***	***0.23 [0.19, 0.26]***
Story	–	***0.24 [0.20, 0.28]***	***0.24 [0.20, 0.28]***
Meaningful	–	***0.21 [0.15, 0.26]***	***0.23 [0.20, 0.27]***
Effort	–	***0.14 [0.10, 0.19]***	***0.14 [0.10, 0.18]***
Time (log)	–	− 0.00 [− 0.02, 0.02]	–
Label × Openness	− 0.01 [− 0.02, 0.01]	*− 0.01 [− 0.02, 0.00]*	*− 0.01 [− 0.02, − 0.00]*
Label × Positive AI	*− 0.01 [− 0.03, − 0.00]*	− 0.00 [− 0.01, 0.00]	–
Label × Negative AI	− 0.01 [− 0.02, 0.01]	− 0.00 [− 0.01, 0.01]	–
Label × Growth	0.02 [− 0.02, 0.05]	0.00 [− 0.03, 0.03]	–
Label × Fixed	− 0.01 [− 0.03, 0.02]	− 0.01 [− 0.03, 0.01]	–
Label × Empathy	0.00 [− 0.01, 0.01]	− 0.00 [− 0.01, 0.01]	–
Label × CRT	− 0.03 [− 0.06, 0.01]	0.00 [− 0.03, 0.03]	–
Label × Age	− 0.00 [− 0.01, 0.01]	0.00 [− 0.00, 0.01]	–
Label × Emotion	–	− 0.02 [− 0.09, 0.04]	–
Label × Story	–	*− 0.19 [− 0.14, − 0.03]*	***− 0.08 [− 0.12, − 0.04]***
Label × Meaningful	–	0.04 [− 0.02, 0.11]	–
Label × Effort	–	*0.09 [0.03, 0.15]*	***0.10 [0.05, 0.16]***
Label × Time (log)	–	0.01 [− 0.01, 0.04]	–
Random effects			
Participant (Intercept)	0.44	0.21	0.24
Label (Slope)	0.13	0.05	0.05
Painting (Intercept)	0.16	0.08	0.08
Residual	0.90	0.55	0.55
Model			
Marginal	0.13	0.53	0.53
ICC	0.40	0.33	0.34
Conditional	0.48	0.69	0.69
AIC	12,606.95	10,449.15	10,300.65

Note. CI = confidence interval, Marginal = variance explained by fixed effects, ICC = intraclass correlation or variance explained by random effects, Conditional = variance explained by fixed and random effects, AIC = Akaike Information Criterion. Italicized text = $p < 0.05$, bolded text = $p < 0.001$

26.0% of the variance, and all effects explained 73.0% of the variance.

Worth

Results for the Worth models are reported in Table 4. Similar to Profundity, there was one interaction involving Label and Positive AI attitudes (b = -0.01, p = 0.014) but it differed such that there was slightly higher Profundity when Label was "human" and Positive AI attitudes were well below average (− 20; Fig. 4). This difference reversed as Positive AI attitudes went from well below average (− 20) to below average (− 10) and continued to diverge to well above average (20). Also similar to Profundity, there were main effects for all five additional criteria (Emotion: b = 0.18; Story: b = 0.12; Meaningful: b = 0.15; Effort: b = 0.27; Time (log): b = 0.05; all p's < 0.001) such that Profundity increased as each criterion increased. The fixed effects of the model explained 57.4% of the variance, the random effects explained 32.0% of the variance, and all effects explained 71.3% of the variance.

Additional analyses

We found a significant main effect of Label on the additional five criteria (p's < 0.001), Emotion, Story, Meaningful, Effort, and Time, such that these criteria were rated higher when the label was "human-created" than "AI-created." These full models (constructed as the previous Model 1's with individual-difference measures) are reported in Additional file 1: Table A.

Case 5:24-cv-03811-EKL Document 842-146 Filed 09/22/26 Page 15 of 23 PageID #: 99831

Table 3 Estimates and associated CIs predicting "Profundity" judgements

Profundity	Model 1	Model 2	Model 3
Fixed effects [95% CI]			
(Intercept)	***2.20 [2.02, 2.38]***	***0.42 [0.31, 0.53]***	***0.40 [0.31 0.50]***
Label (Human = 1)	***0.37 [0.29, 0.58]***	−0.15 [−0.31, 0.01]	***−0.14 [−0.20, −0.07]***
Painting Type (Representational = 1)	***0.40 [0.21, 0.58]***	−0.04 [−0.11, 0.02]	–
Openness	−0.01 [−0.03, 0.01]	−0.00 [−0.01, 0.01]	–
Positive AI	*0.02 [0.01, 0.04]*	0.01 [−0.00, 0.01]	0.01 [−0.00, 0.01]
Negative AI	0.00 [−0.02, 0.02]	0.00 [−0.01, 0.01]	–
Growth	***0.11 [0.05, 0.16]***	*0.04 [0.02, 0.07]*	*0.04 [0.01, 0.06]*
Fixed	*0.07 [0.03, 0.11]*	*0.02 [0.00, 0.04]*	*0.02 [0.00, 0.04]*
Empathy	0.00 [−0.01, 0.02]	−0.00 [−0.01, 0.01]	–
CRT	*−0.08 [−0.13, −0.02]*	−0.03 [−0.06, 0.00]	–
Age	0.01 [−0.00, 0.02]	0.00 [−0.00, 0.01]	–
Emotion	–	***0.22 [0.18, 0.26]***	***0.22 [0.19, 0.25]***
Story	–	***0.18 [0.15, 0.21]***	***0.19 [0.17, 0.22]***
Meaningful	–	***0.26 [0.22, 0.31]***	***0.26 [0.22, 0.29]***
Effort	–	***0.20 [0.16, 0.24]***	***0.19 [0.16, 0.22]***
Time (log)	–	*0.02 [0.00, 0.03]*	*0.02 [0.00, 0.03]*
Label × Openness	0.01 [−0.01, 0.02]	−0.01 [−0.01, 0.00]	–
Label × Positive AI	*−0.01 [−0.03, −0.01]*	*−0.01 [−0.01, −0.00]*	*−0.01 [−0.01, −0.00]*
Label × Negative AI	−0.01 [−0.02, 0.01]	0.00 [−0.00, 0.01]	–
Label × Growth	−0.00 [−0.04, 0.03]	−0.01 [−0.03, 0.01]	–
Label × Fixed	−0.01 [−0.03, 0.02]	−0.00 [−0.02, 0.01]	–
Label × Empathy	0.01 [−0.00, 0.02]	0.00 [−0.00, 0.01]	–
Label × CRT	−0.02 [−0.06, 0.01]	0.00 [−0.02, 0.02]	–
Label × Age	−0.00 [−0.01, 0.00]	−0.00 [−0.01, 0.00]	–
Label × Emotion	–	−0.00 [−0.06, 0.06]	–
Label × Story	–	0.03 [−0.01, 0.08]	–
Label × Meaningful	–	−0.01 [−0.07, 0.05]	–
Label × Effort	–	−0.02 [−0.07, 0.04]	–
Label × Time (log)	–	−0.00 [−0.02, 0.02]	–
Random effects			
Participant (Intercept)	0.53	0.13	0.13
Label (Slope)	0.12	0.02	0.02
Painting (Intercept)	0.06	0.00	0.01
Residual	0.76	0.41	0.41
Model			
Marginal	0.15	0.64	0.64
ICC	0.44	0.25	0.26
Conditional	0.52	0.73	0.73
AIC	11,905.68	9098.20	8940.49

Note. CI = confidence interval, Marginal = variance explained by fixed effects, ICC = intraclass correlation or variance explained by random effects, Conditional = variance explained by fixed and random effects, AIC = Akaike Information Criterion. Italicized text = $p < 0.05$, bolded text = $p < 0.001$

Study 2 discussion

Study 2 replicated and extended the general anti-AI art sentiments from Study 1. Using more-sophisticated modeling, we again found that participants tended to prefer artworks that were labelled human-created over those labelled AI-created art. Importantly, this pattern of results was observed across all four original criteria, as well as the additional five criteria added in Study 2.

We designed our analyses such that three models predicted scores for each criterion. In Model 1, individual-difference traits were targeted, as they were the only predictor variables inputted besides the Label and Painting Type. In Model 2, we added predictor variables that could vary with each trial—rather than remain static, as with the individual-difference metrics—which were the 5 additional criteria: Emotion, Story, Meaningful, Effort, and Time. Model 2 aimed to identify the contributions that more-communicative criteria could predict over our original four criteria. For instance, how does the ability of an art piece to evoke a story relate to the general liking of that piece, and does this relationship depend on the label? Due to the abundance of predictor variables, several of which were not significant in Model 2, we created a third and final model consisting only of significant main effects and interactions from Model 2, which allowed for easier interpretation and discussion. This tripartite model approach was not specified in the pre-registration, though the basic model type and structure (linear mixed models) were.

As determined by our significant-only model (Model 3), when participants reported Liking some paintings

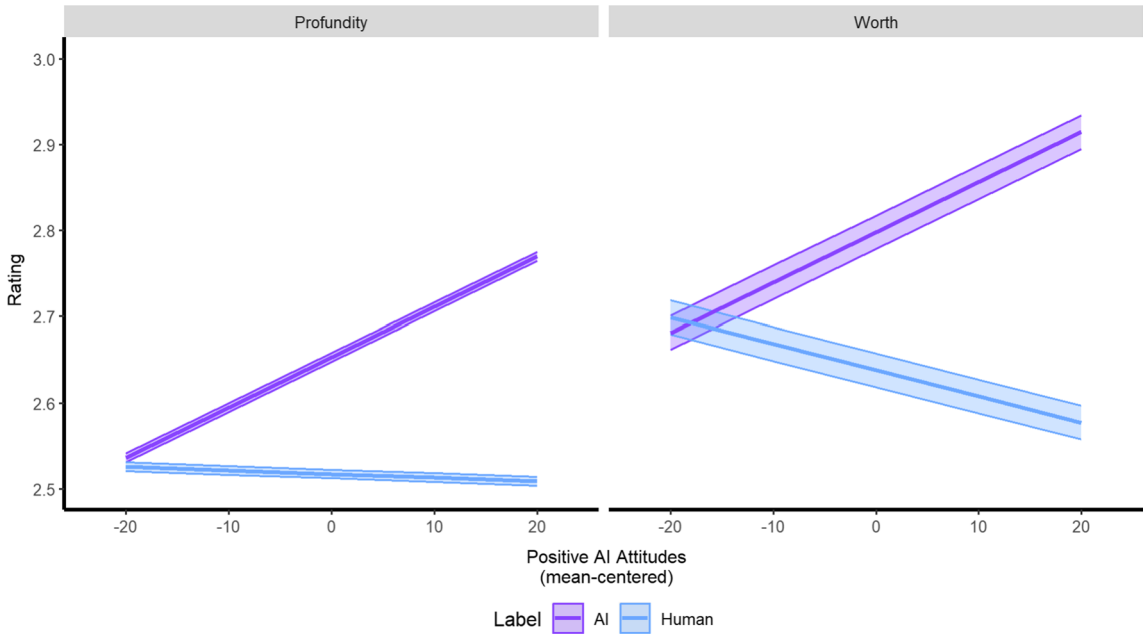


Fig. 4 Interaction plot of Positive AI attitudes on communicative judgements of Profundity and Worth. Confidence bands represent the 95% confidence interval

more than others, regardless of Label, we also found an increased ability to extract emotion out of the painting (Emotion), to create a story alongside the painting (Story), to find personal meaning from the painting (Meaningful), and to believe the painting required more effort to produce (Effort). Interestingly, interactions between Label and two criteria (Story, Effort) emerged, but in opposite directions. Specifically, with a greater ability to extract a story from a painting (i.e., higher narrativity), participants liked the painting more if that painting was created by AI as compared to if it was created by humans. Narratives, then, may serve as a catalyst for appreciating products on surface levels. However, if participants believed the painting involved more effort in its creation than other paintings, participants liked paintings made by humans more than those made by AI. This may be due to a lack of understanding, on the participants' end, of how AI produce pieces, and a need to actually witness the artistic production as investigated in Chamberlain et al. (2018). Additionally, it may reflect an effort heuristic. As suggested by Kruger et al. (2004), people use effort as a metric to inform their judgements of the quality of a product (e.g., poems, paintings, suits of armor; Kruger et al., 2004). These findings suggest this effort heuristic may only exist for human products and not AI products. No individual-difference traits emerged in Liking's Model 3, but its Model 1 originally found that those with both higher growth-mindset scores and higher fixed-mindset scores liked paintings more in general. That fixed and growth mindsets were both positively associated with aesthetic judgements was surprising, yet interesting. This finding partly mirrors results from Karwowski et al. (2019), who found some people can hold both fixed and growth attitudes toward creativity, and further nuances perceptions of creativity that were formerly considered along a continuum between fixed or growth poles (Karwowski, 2014). The Model 1 also reported a main effect of positive attitudes toward AI, which was qualified with an interaction with the Label, such that with higher appreciation for AI, participants liked AI-labelled art more than human-labelled art.

The final model of Beauty revealed similar main effects and interactions as the final Liking model. There was also an additional main effect of age, such that younger participants found art, regardless of its purported creator, to be more beautiful; this may be in line with findings from Tröndle et al. (2014), who found that younger participants are more willing to ascribe contemporary artistic creations as "art." An interaction between Label and Openness emerged, such that with higher openness to experience, participants found AI-labelled paintings more beautiful as compared to human-labelled paintings, nuancing claims of openness and aesthetic appreciation by McCrae (2007). No other individual-difference measures emerged in this final model, but in Model 1, we originally had main effects of positive AI attitudes (qualified with an interaction with the Label, similar to the Liking Model 1), and a main effect of fixed mindset. Again, that

Table 4. Estimates and associated CIs predicting "Worth" judgements

Worth	Model 1	Model 2	Model 3
Fixed effects [95% CI]			
(Intercept)	***2.29 [2.12, 2.45]***	***0.65 [0.53, 0.77]***	***0.66 [0.55, 0.77]***
Label (Human = 1)	***0.52 [0.43, 0.61]***	*−0.19 [−0.37, −0.02]*	*−0.21 [−0.35, −0.07]*
Painting Type (Representational = 1)	***0.35 [0.18, 0.52]***	−0.01 [−0.09, 0.06]	–
Openness	−0.01 [−0.03, 0.01]	0.01 [−0.01, 0.02]	–
Positive AI	*0.02 [0.00, 0.03]*	0.00 [−0.01, 0.01]	0.01 [−0.00, 0.01]
Negative AI	0.01 [−0.01, 0.03]	0.01 [−0.00, 0.02]	–
Growth	*0.07 [0.02, 0.12]*	0.01 [−0.02, 0.04]	–
Fixed	*0.05 [0.01, 0.09]*	0.01 [−0.01, 0.03]	–
Empathy	0.00 [−0.01, 0.02]	−0.00 [−0.01, 0.01]	–
CRT	*−0.08 [−0.13, −0.03]*	−0.03 [−0.06, 0.00]	–
Age	0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	–
Emotion	–	***0.17 [0.13, 0.21]***	***0.18 [0.15, 0.20]***
Story	–	***0.14 [0.11, 0.17]***	***0.12 [0.10, 0.14]***
Meaningful	–	***0.18 [0.14 0.23]***	***0.15 [0.12, 0.18]***
Effort	–	***0.25 [0.21, 0.29]***	***0.27 [0.23, 0.31]***
Time (log)	–	***0.04 [0.02, 0.06]***	***0.05 [0.03, 0.06]***
Label × Openness	0.01 [−0.01, 0.03]	−0.00 [−0.01, 0.01]	–
Label × Positive AI	*−0.02 [−0.03, −0.00]*	*−0.01 [−0.02, −0.00]*	*−0.01 [−0.02, −0.00]*
Label × Negative AI	−0.01 [−0.03, 0.01]	−0.00 [−0.01, 0.01]	–
Label × Growth	0.00 [−0.04, 0.05]	0.00 [−0.03, 0.02]	–
Label × Fixed	−0.01 [−0.04, 0.02]	0.00 [−0.02, 0.02]	–
Label × Empathy	0.00 [−0.01, 0.02]	0.00 [−0.01, 0.01]	–
Label × CRT	−0.02 [−0.06, 0.03]	−0.00 [−0.03, 0.02]	–
Label × Age	−0.01 [−0.01, 0.00]	−0.00 [−0.01, 0.00]	–
Label × Emotion	–	0.01 [−0.04, 0.07]	–
Label × Story	–	−0.04 [−0.09, 0.00]	–
Label × Meaningful	–	−0.06 [−0.11, 0.00]	–
Label × Effort	–	*0.06 [0.00, 0.11]*	0.02 [−0.02, 0.06]
Label × Time (log)	–	0.01 [−0.01, 0.04]	–
Random effects			
Participant (Intercept)	0.47	0.16	0.17
Label (Slope)	0.24	0.10	0.10
Painting (Intercept)	0.05	0.01	0.01
Residual	0.61	0.36	0.36
Model			
Marginal	0.16	0.59	0.57
ICC	0.45	0.31	0.32
Conditional	0.54	0.72	0.71
AIC	11,040.92	8711.31	8557.09

Note. CI = confidence interval, Marginal = variance explained by fixed effects, ICC = intraclass correlation or variance explained by random effects, Conditional = variance explained by fixed and random effects, AIC = Akaike Information Criterion. Italicized text = $p < 0.05$, bolded text = $p < 0.001$

fixed mindsets emerge is an interesting point that merits future investigation.

We also implemented the same models for our communicative criteria. The final models for Profundity and Worth showed a main effect of Emotion, Story, Meaningful, Effort, and (nonsignificant in Liking and Beauty) Time. For both Profundity and Worth, there was also a significant interaction between the label and positive attitudes toward AI, such that more positive attitudes toward AI led to higher perceived Profundity and Worth when the art was created by AI as compared to by a human. The Profundity final model, but not Worth's, also had significant effects of growth and fixed mindsets. However, in both Model 1's of Profundity and Worth, main effects of positive AI attitudes (qualified by an interaction with Label), growth and fixed mindsets, and CRT scores emerged. The main effect specifically suggests that lower CRT scores (i.e., poorer ability to override intuitive responses) led to higher ratings of Profundity and Worth regardless of label. This mirrors work by Pennycook et al. (2015) that investigated profundity and judgements of randomly generated statements, suggesting an overall increased ability to find Profundity and Worth in the environment when having a lowered filter for regulation. Lastly, supplemental analyses (see Additional file 1) also showed that the additional five communicative criteria (Emotion, Story, Meaningful, Effort, Time) behaved similarly in response to an "AI" or "human" label (such that the latter has consistently higher ratings). In other words, a label of "human" acts as a vessel for elaborative and communicative engagement.

Overall, results from Study 2 supported a distinction in judgement patterns between Liking/Beauty and Profundity/Worth across models, replicating and extending these patterns from Study 1. Interestingly, it appears that when raters judge sensory-level criteria, they may rely on some properties that require greater engagement (e.g., story-telling and effort) depending on the label provided. In this regard, it is possible that the sensory-level judgements can be both a consequence of more-passive appraisals based on solely visual features *and* an interaction with other judgement mechanisms that require active engagement with the piece, in line with dual-process models like that of Graf and Landwehr (2015). Meanwhile, when judging communicative criteria like Profundity and Worth, people tend to rate along the lines of other judgements (e.g., time spent on painting), only moderated by personal positive AI attitudes depending on the label. Individual differences, then, may only affect how viewers differently judge AI or human art when judging their communicative properties.

General discussion

In line with the extant literature on the role of labels in subjective evaluations—both in aesthetics and other domains of preferences (e.g., Kirk et al., 2009; McClure et al., 2004; Newman & Bloom, 2012; Plassmann et al., 2008; Turpin et al., 2019)—our findings demonstrate that labels significantly influence people's judgements of visual artworks. Specifically, we investigated how a label of "AI-created" on a painting affects the evaluation of the artwork, as compared to a label of "human-created." In a time of unprecedented growth in AI-art algorithms, websites, and exhibits, it is evermore timely to consider how people perceive, judge, and interact with artwork produced by AI. We found a bias against AI-labelled art across all aesthetic judgement criteria, but more so for appraisals related to art as a deeper communicative medium (Profundity, Worth). In fact, differences in surface-level judgements between human- and AI-labelled art, though influenced by some communicative processes (e.g., Effort and Story), were nearly non-significant. These findings have crucial implications for art, AI, and creativity at large.

Art created by AI has been met with both success and controversy (see Chatterjee, 2022); and, very recently, these creative products have permeated into mainstream culture like museums and artistic industries (Diaz, 2022; Kaleagasi, 2017; Kinsella, 2018; Zulić, 2019). Though contention ensues over whether artificial programs are capable of producing "creative" products (in line with the computer science principle known as the Lovelace Objection; Natale & Henrickson, 2022), it is nonetheless critical to understand how audiences perceive such art given that contemporary psychological literature considers creativity to be not an absolute construct, but a consequence of changing interactions between subjective criteria and socio-cultural factors like technology (i.e., "in the eye of the beholder"; Amabile, 1982; Cseh & Jeffries, 2019; Hennessey et al., 2011; Kaufman & Sternberg, 2010). Probing these attitudes toward AI-art, our studies demonstrate clear preferences for artwork labelled as human-created as compared to AI-created, despite all art being AI-created (which controlled for potential bottom-up discrepancies between actual human- versus AI-created paintings, while also leveraging findings that show AI art is indiscernible from human art; Chamberlain et al., 2018; Gangadharbatla, 2022). However, the degree of preference depended on the type of judgement, as our results show a marked distinction between judgements of Liking/Beauty and Profundity/Worth. In Study 1, for instance, we found increased effect sizes for differences between human and AI labels for the more-communicative criteria of Profundity ($d = 0.47$) and Worth ($d = 0.61$) than Liking ($d = 0.17$) and Beauty ($d = 0.22$). This distinction is further emphasized by the different interaction mechanisms discovered in Study 2. Ultimately, these results align well with hierarchical and multi-processing models of aesthetic encounters that view sensory versus communicative engagements as different pathways, including Graf and Landwehr's (2015) dual-process model of aesthetic liking and Chatterjee and Vartanian's (2016) Aesthetic Triad, which delineates a distinction between sensory- and meaning-level systems of aesthetics. Corroborating these models, our results suggest that people may feel cognitively obstructed when engaging deeply with and deriving meaning from artworks that are labelled as created by AI (or have any other label that is pejorative). Equivalently, a "human" label encourages elaborative engagement (e.g., deriving emotion, effort, narratives). However, on quick, surface-level evaluations rather than elaborative appraisals, AI art may be better appreciated.

Several interpretations can be drawn from these results. Firstly, participants may hold a nuanced view of creativity: while humans can achieve creativity across cognition and production, AI can produce only surface-level reflections of creative thinking according to raters. Artworks that are created by humans may reflect a profound human experience—and thus be deemed more monetarily worthy—that AI cannot produce. Indeed, absent the human experience, AI can produce only sensorily similar—still comparatively beautiful and liked—pieces of visual art. This provides a fascinating perspective of future creativity and aesthetics research: one that does not consider creativity as an all-or-nothing trait, but a spectrum of abilities that AI may be able to penetrate.

Even though these two lines of aesthetic engagement emerged, Study 2 showed that sensory-level judgements actually seemed to be influenced in interesting ways by additional judgements and processes. Such moderation effects support sensory versus communicative pathways of engagement with art, but also reflect how they may intertwine. Indeed, this approach of aesthetics reconciles how aesthetic liking has been found to be both a consequence of passive viewing based on solely bottom-up processing of visual features and product of deeper and more-elaborative engagement with the piece (see Graf & Landwehr, 2015). We found support for this dual-process model in Study 2. Specifically, for Liking and Beauty (two sensory judgements), we isolated interactions between the narrativity of a painting and the label, where higher narrativity led to higher Liking and Beauty when the label of the painting was "AI-created" more than when the label was "human-created." This seems to imply that narrativity is a key element in surface-level evaluations of artwork, especially with non-human artists. This finding also has strong implications for marketing and product branding, as it is reminiscent of the mass production advertising practices in the early twentieth century where consumers were overall hesitant to purchase goods from multinational conglomerates until narratives were paired with the product (Freeman, 2014; Pulizzi, 2012). In other words, imagined backstories helped to overcome biases against machine-made products. This vessel for deeper engagement (and consumerism) provides a way for AI art to, potentially, have higher evaluations outside of simple bottom-up appraisals.

Narrativity has also played key roles in other modes of art. Music has specifically been investigated as a vessel for stories and resulting aesthetic engagements. Such work (e.g., Margulis et al., 2019; McAuley et al., 2021) finds interesting cross-cultural differences, implicating the role of media in our consumption of art and underlying associations (see Wu et al., 2020 for a cross-cultural study that finds higher appreciation for AI products of art from Chinese participants than American participants). The fact that narrativity scores interact with the label of another modality in our study (paintings) strengthens the idea that stories serve as an engine for general art judgement (that is sensitive to context). Future work could further investigate this possibility by experimentally manipulating the level of narrativity of artworks, which would in turn allow for the assessment of the influence of such a manipulation on people's aesthetic encounters and appraisals. In such a case, we would anticipate that increasing the narrativity associated with a given piece of art would produce more-positive judgement of that art relative to artworks with lower levels of narrativity. Narratives could also be indirectly stimulated; that is, certain titles and terms may automatically evoke stories which may affect ratings, similar to how pseudo-profound randomly generated titles increase perceived profundity for both computer-generated and human-created abstract paintings (Turpin et al., 2019).

Another interaction that emerged in Study 2 was between the perceived effort of paintings and the label, though in the opposite direction as the narrativity interaction: higher perceived effort led to higher Liking and Beauty when the label was "human-created" more than when it was "AI-created." Kruger et al.'s (2004) "effort heuristic," then, may only be immediately available for human-made products, and people seek markers of effort in such products. For instance, Chamberlain et al. (2018) found that people observe the meticulous brush strokes (i.e., an/isotropy) in artworks and use these brush strokes as a heuristic for aesthetic judgements. In other words, people may show an appreciation for an artist's effortfulness (and/or meticulousness) as the labor infuses the product with some temporal or monetary value. That this effort heuristic doesn't emerge for AI-labelled art could also reflect a general lack of knowledge about how much effort goes into AI-created art (and what "effort" even reflects in the context of AI algorithms). Interestingly, however, Chamberlain et al. (2018) were able to increase appreciation for AI-created art by showing participants videos of anthropomorphized robots creating paintings, which could have reasonably led participants to believe that the AI engaged in greater levels of effort in its artistic process (thereby mitigating the anti-AI bias). Thus, effort heuristics may be probed even in instances of AI-art in future work.

In regard to relationships between judgements and with painting type (abstract versus representational), we only found modest significant interactions between the label and painting type for Liking and Profundity in Study 1, and did not include these interactions in Study 2 as none emerged significantly in initial analyses. This ultimately supports findings from Chamberlain et al. (2018), who found no significant interactions between the label and painting type of general aesthetic preferences. In addition, these findings give nuance to the attribution effects investigated by Gangadharbatla (2022) and Chamberlain et al. (2018), who found an increased willingness to attribute an AI creator to abstract paintings, and a human creator to representational paintings, respectively. It should be emphasized, however, that in the present study, we defined abstract paintings as having partially or completely unrecognizable referents, and representational paintings as having completely recognizable referents. Notably, however, more-strict definitions and criteria to dissociate abstract and representational art could be used in future research. Importantly, by employing stricter

criteria for separating abstract from representational art, future research may, unlike in our study, find significant differences across these two painting types. Future work could also explore intermediary categories like partially representational art in such investigations.

Surprisingly, we also found very few significant relationships between our individual-differences measures and our various indices of art judgement. Only in Study 2 did we see robust effects of an individual trait on judgement patterns. That is, personal positive attitudes toward AI interacted with the label such that higher appreciation for AI led to higher appreciation for art labelled as created by AI as compared to human-labelled art; perhaps as we accept and utilize machine learning in a growing technological world, these traits will naturally proliferate among the layperson, leading to growing acceptance of AI art. Overall, though, we were surprised by the lack of other clear predicting individual-difference measures in judgements of art. Though some small effects for age and openness emerged in models of Beauty in Study 2, ultimately, more research is needed to determine what, if any, personal traits predict why one may judge art by AI higher than art by humans across criteria.

Limitations

Notably, there are some limitations to our studies that should be considered, to appropriately qualify our findings, that will inform improved designs in future research on the topic. Firstly, we cannot be certain of how well we deceived participants with some of our labels (i.e., those that were false in cases wherein the labels indicated that the artworks were created by a human), although we can nevertheless minimize these concerns (at least to some extent) by considering the past literature showing participants' inability to discriminate between AI- and human-created paintings (Chamberlain et al., 2018; Gangadharbatla, 2022). In any case, moving forward, this concern can be eliminated by explicitly ensuring that participants are not able to accurately determine that all of our stimuli were in fact created by AI. Alternatively, a future study could employ both (actual) AI- and human-created artwork that has been normed on the criteria of note. Secondly, though our judgement criteria provided more nuance to simple "aesthetic pleasure" by asking for an artwork's Profundity or Worth, for instance, it is possible the very act of asking participants to (perhaps atypically) consider these attributes of an artwork alters their natural evaluation. Third, another consideration is that the label of "AI" is quite broad and therefore loose in its meaning, and this may have differentially influenced aesthetic judgements across participants depending on their understanding of the term; future work could specifically ask participants what "AI" means to them, and have independent raters (who are knowledgeable about AI) rate participant responses as an individual-difference measure of AI knowledge that may moderate and/or mediate results. Relatedly, some participants may know more about the specifics of machine learning than others (including, for example, the amount of effort that goes into producing an AI algorithm, and the amount of time it takes an individual to create an artwork via such an algorithm), which could influence their responses to our various measures. Also, some participants may be inclined to consume potentially biased information about AI through news sources, which could influence their responses. A simple teaching intervention on AI or assessment of past experiences with AI could assuage these concerns.

Lastly, though previous studies have investigated the role of art expertise in judgements of AI versus human aesthetics (and aesthetics in general), here, we did not index rater expertise. However, findings on the effects of expertise have been somewhat mixed. Chamberlain et al. (2018) found no moderating effect of expertise on judgement biases on art created by AI or humans, and Moffat and Kelly (2006) found that non-musicians were surprisingly able to discern AI- versus human-created compositions better than musicians. That said, both Moffat and Kelly (2006) and Darda and Cross (2023) found that the anti-AI bias is actually stronger in the experts, which matches general trends that experts rate harsher (e.g., Lundy & Smith, 2017). However, for experts to have a harsher anti-AI bias than non-experts also goes against claims that "novices are… more non-aesthetically biased in their aesthetic judgments compared with experts," which suggests that experts are more resistant to label effects (Lundy & Smith, 2017, p. 140). Regardless, research is merited to delineate the true role that background experience has on judgements of aesthetics (including AI aesthetics), though given that experts may have a stronger anti-AI bias as some studies have supported, our results may, if anything, reflect more-conservative effects given our sample was not limited to experts.

Conclusions

Across two studies, we found evidence of a multi-processing approach to aesthetics, which has strong implications for creative thinking and art industries. Firstly, people preferred (purportedly) human-created art over AI-created art. This preference was particularly evident for criteria that communicated deeper meanings of the art (e.g., Profundity, Worth). On more-sensory levels, the difference between human- and AI-labelled art was much more modest, though significant differences were nonetheless observed. As such, AI-labelled art

can still be greatly appreciated (almost as much as human-labelled art) when people utilize varying levels of engagement processes. Interestingly, the sensory- versus communicative-level judgement processes were further distinguished by different interaction processes: the former were moderated by rates of story-telling and perceived effort (though in opposite directions), while the latter were moderated by personal positive attitudes toward AI. These interactions shed light on when and why individuals may appreciate art made by different creators, as posed in our initial research question, and also reflect how different stages of multi-processing models of aesthetics may interact with one another.

In conclusion, people tend to perceive art as reflecting a human-specific experience, though creator labels seem to mediate the ability to derive deeper evaluations from art. Thus, creative products like art may be achieved—according to human raters—by non-human AI models, but only to a limited extent that still protects a valued anthropocentrism.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41235-023-00499-6.

Additional file 1: Additional file 1 includes the Model 1 output for each of the Study 2 additional criteria.

Acknowledgements
The authors thank Leonard Faul for assistance in statistical analyses, Kibby McMahon for assistance with Study 1 data collection, and Bob Zavon for feedback and assistance throughout.

Significance statement
Creativity and creative products are engines of human communication and societal progress. With the development of artificial intelligence models that can generate visual art that is, in many cases, indistinguishable from human-created art, the necessity of a human creator in the creative process is questioned. However, judgements of artworks created by either humans or AI may vary depending simply on source information about the purported creator of the artwork (AI vs. human), which can result in different levels of acceptance and appreciation of such creative products. We found that human-labelled art is more positively judged across both sensory and communicative properties than AI-labelled art, even though all art was, in actuality, AI-created. This demonstrates that, while there is a seemingly human-specific role that people believe is necessary for the production of high-quality art—particularly for non-sensory properties of art like Profundity and Worth—AI-art is still sufficiently impressive to be mistaken for human art, and that the creator label, alone, has important implications for how people appraise artworks. In sum, people value art through multiple, co-occurring processes, with creator labels influencing people's evaluations of art.

Author contributions
LB served as lead for developing the study plan with PS, analyzing data with AC, organizing and running Study 2, writing—original draft, and project administration. RS and MT organized Study 1 and collected Study 1 data. RS also helped to ideate Study 1 and analyze Study 1 data. AR provided edits and assistance with study design. AC served as lead for data analysis and data visualization. NB and SS provided conceptualization and edits. PS provided conceptualization and major edits. All authors read and approved the final manuscript.

Funding
This work was supported by a National Science Foundation Graduate Fellowship (Lucas Bellaiche) and, for participant compensation, a Charles Lafitte Foundation Graduate Grant Award (Lucas Bellaiche).

Availability of data and materials
The datasets supporting the conclusions of this article, and the scripts used to analyze the data, are available on Open Science Framework, and can be found at: https://osf.io/cgw8v/.

Declarations

Ethics approval and consent to participate
Study 1 was approved by the University of Waterloo Research Ethics Board (31067), and Study 2 was approved by the Duke University Campus Institutional Review Board (2022-0534). All participants provided informed consent to participate in the studies, with knowledge that their anonymized data may be openly shared with others.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Open practices statement
All data, analysis scripts, stimuli, and the Study 2 pre-registration can be found at https://osf.io/cgw8v/.

Received: 9 February 2023 Accepted: 22 June 2023
Published online: 04 July 2023

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*, 997–1013. https://doi.org/10.1037/0022-3514.43.5.997
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01
Berlyne, D. E. (1975). Behaviorism? Cognitive theory? Humanistic psychology? To Hull with them all. *Canadian Psychological Review, 16*, 69–80.
Berlyne, D. E. (1960). Conflict, arousal, and curiosity. *McGraw-Hill Book Company*. https://doi.org/10.1037/11164-000
Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.
Blank, P., Massey, C., Gardner, H., & Winner, E. (1984). Perceiving What Paintings Express. In: W. R. Crozier & A. J. Chapman (Eds.), *Advances in Psychology* (Vol. 19, pp. 127–143). North-Holland. https://doi.org/10.1016/S0166-4115(08)62349-3.
Bornstein, M. H. (1975). Qualities of color vision in infancy. *Journal of Experimental Child Psychology, 19*(3), 401–419. https://doi.org/10.1016/0022-0965(75)90070-3
Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*, 434–447. https://doi.org/10.3758/s13421-013-0367-9
Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts, 12*(2), 177–192. https://doi.org/10.1037/aca0000136
Chatterjee, A. (2014). *The aesthetic brain: How we evolved to desire beauty and enjoy art*. Oxford University Press.
Chatterjee, A. (2022). Art in an age of artificial intelligence. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.1024449
Chatterjee, A., & Cardillo, E. R. (2022). *Brain, beauty, and art: Essays bringing neuroaesthetics into focus*. Oxford University Press.

Chatterjee, A., & Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New York Academy of Sciences, 1369*(1), 172–194. https://doi.org/10.1111/nyas.13035

Chiarella, S., Torromino, G., Gagliardi, D., Rossi, D., Babiloni, F., & Cartocci, G. (2022). Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. *Computers in Human Behavior, 137*, 107406. https://doi.org/10.1016/j.chb.2022.107406

Cupchik, G. C., & Berlyn, D. E. (1979). The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology, 20*, 93–104. https://doi.org/10.1111/j.1467-9450.1979.tb00688.x

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 159–166. https://doi.org/10.1037/aca0000220

Darda, K. M., & Cross, E. S. (2023). The computer, A choreographer? Aesthetic responses to randomly-generated dance choreography by a computer. *Heliyon, 9*(1), e12750. https://doi.org/10.1016/j.heliyon.2022.e12750

Darda, K., Carre, M., & Cross, E. (2023). Value attributed to text-based archives generated by artificial intelligence. *Royal Society Open Science, 10*(2), 220915. https://doi.org/10.1098/rsos.220915

Diaz, J. (2022, November 18). *MoMA's newest artist is an AI trained on 180,000 works, from Warhol to Pac-Man*. Fast Company. https://www.fastcompany.com/90811669/momas-newest-artist-is-an-ai-trained-on-180000-works-from-warhol-to-pac-man

Feist, G. J., & Brady, T. R. (2004). Openness to experience, non-conformity, and the preference for abstract art. *Empirical Studies of the Arts, 22*(1), 77–89. https://doi.org/10.2190/Y7CA-TBY6-V7LR-76GK

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences, 11*(5), 197–203. https://doi.org/10.1016/j.tics.2007.02.003

Freeman, M. (2014). Transmedia critical| advertising the yellow brick road: historicizing the industrial emergence of transmedia storytelling. *International Journal of Communication, 8*, 19.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gangadharbatla, H. (2022). The role of AI attribution knowledge in the evaluation of artwork. *Empirical Studies of the Arts, 40*(2), 125–142. https://doi.org/10.1177/0276237421994697

Graf, L. K. M., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review., 19*(4), 395–410. https://doi.org/10.1177/1088868315574978

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444–454. https://doi.org/10.1037/h0063487

Hass, R. W., Katz-Buonincontro, J., & Reiter-Palmon, R. (2016). Disentangling creative mindsets from creative self-efficacy and creative identity: Do people hold fixed and growth theories of creativity? *Psychology of Aesthetics, Creativity, and the Arts, 10*, 436–446. https://doi.org/10.1037/aca0000081

Heinrichs, R. W., & Cupchik, G. C. (1985). Individual differences as predictors of preference in visual art. *Journal of Personality, 53*(3), 502–515. https://doi.org/10.1111/j.1467-6494.1985.tb00379.x

Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual Assessment. In *Encyclopedia of Creativity* (pp. 253–260). Elsevier. https://doi.org/10.1016/B978-0-12-375038-9.00046-7.

Hong, J.-W., & Curran, N. M. (2019). Artificial Intelligence, artists, and art: Attitudes toward artwork produced by humans versus Artificial Intelligence. *ACM Transactions on Multimedia Computing, Communications, and Applications, 15*(2s), 58:1-58:16. https://doi.org/10.1145/3326337

Hong, J.-W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the AI-composed music evaluation. *Computers in Human Behavior, 131*, 107239. https://doi.org/10.1016/j.chb.2022.107239

Israfilzade, K. (2020). What's in a name? Experiment on the aesthetic judgments of art procured by Artificial Intelligence. *Journal of Arts, 3*, 143–158. https://doi.org/10.31566/arts.3.011

Johnson, G. (1997, November 11). Undiscovered Bach? No, a Computer Wrote It. *The New York Times*. https://www.nytimes.com/1997/11/11/science/undiscovered-bach-no-a-computer-wrote-it.html

Kaleagasi, B. (2017, March 9). *A New AI Can Write Music as Well as a Human Composer*. Futurism. Retrieved December 14, 2022, from https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer

Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 62–70. https://doi.org/10.1037/a0034898

Karwowski, M., Royston, R. P., & Reiter-Palmon, R. (2019). Exploring creative mindsets: Variable and person-centered approaches. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 36–48. https://doi.org/10.1037/aca0000170

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2010). *The Cambridge handbook of creativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511763205

Kaufman, S. B. (2013). Opening up openness to experience: A four-factor model and relations to creative achievement in the arts and sciences. *The Journal of Creative Behavior, 47*(4), 233–255. https://doi.org/10.1002/jocb.33

Kettlewell, N., Lipscomb, S., Evans, L., & Rosston, K. (1990). The effect of subject matter and degree of realism on aesthetic preferences for paintings. *Empirical Studies of the Arts, 8*(1), 85–93. https://doi.org/10.2190/Y8JA-3GUR-M08C-BPEC

Kinsella, E. (2018, October 25). *The First AI-Generated Portrait Ever Sold at Auction Shatters Expectations, Fetching $432,500—43 Times its Estimate*. Artnet News. https://news.artnet.com/market/first-ever-artificial-intelligence-portrait-painting-sells-at-christies-1379902

Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *NeuroImage, 44*(3), 1125–1132. https://doi.org/10.1016/j.neuroimage.2008.10.009

Knapp, R. H., & Wulff, A. (1963). Preferences for abstract and representational art. *The Journal of Social Psychology, 60*(2), 255–262. https://doi.org/10.1080/00224545.1963.9922196

Köbis, N., & Mossink, L. (2020). *Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry* (arXiv:2005.09980). https://doi.org/10.48550/arXiv.2005.09980

Komar, V., & Melamid, A. (1999). *Painting by numbers: Komar and Melamid's scientific guide to art*. University of California Press.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology, 40*, 91–98. https://doi.org/10.1016/S0022-1031(03)00065-9

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95*(4), 489–508. https://doi.org/10.1348/0007126042369811

Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. In *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517731.

Lundy, D. E., & Smith, J. L. (2017). It's Tough To Be a Critic: Professional Versus Nonprofessional Music Judgment. *Empirical Studies of the Arts, 35*(2), 139–168. https://doi.org/10.1177/0276237416661989

Margulis, E. H., Wong, P. C. M., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications*. https://doi.org/10.1057/s41599-019-0363-1

Mastandrea, S., Bartoli, G., & Carrus, G. (2011). The automatic aesthetic evaluation of different art and architectural styles. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 126–134. https://doi.org/10.1037/a0021126

Mastandrea, S., Wagoner, J. A., & Hogg, M. A. (2021). Liking for abstract and representational art: National identity as an art appreciation heuristic. *Psychology of Aesthetics, Creativity, and the Arts, 15*, 241–249. https://doi.org/10.1037/aca0000272

Mazzone, M., & Elgammal, A. (2019). Art, creativity, and the potential of Artificial Intelligence. *Arts*. https://doi.org/10.3390/arts8010026

McAuley, J. D., Wong, P. C. M., Bellaiche, L., & Margulis, E. H. (2021). What drives narrative engagement with music? *Music Perception, 38*(5), 509–521. https://doi.org/10.1525/mp.2021.38.5.509

McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*(2), 379–387. https://doi.org/10.1016/j.neuron.2004.09.019

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion, 31*(1), 5–11. https://doi.org/10.1007/s11031-007-9053-1

McCrae, R. R., & Greenberg, D. M. (2014). Openness to Experience. In *The Wiley Handbook of Genius* (pp. 222–243). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118367377.ch12
Mikalonytė, E. S., & Kneer, M. (2022). Can Artificial Intelligence make art?: Folk intuitions as to whether AI-driven robots can be viewed as artists and produce art. *ACM Transactions on Human-Robot Interaction, 11*(4), 43:1-43:19. https://doi.org/10.1145/3530875
Mockros, C. A. (1993). The development of aesthetic experience and judgement. *Poetics, 21*(5), 411–427. https://doi.org/10.1016/0304-422X(93)90003-Y
Moffat, D., & Kelly, M. G. (2006). An investigation into people's bias against computational creativity in music composition. *Assessment, 13*(11), 1–8.
Natale, S., & Henrickson, L. (2022). The Lovelace effect: Perceptions of creativity in machines. *New Media & Society*. https://doi.org/10.1177/14614448221077278
Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology. General, 141*(3), 558–569. https://doi.org/10.1037/a0026035
Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology, 64*(1), 77–107. https://doi.org/10.1146/annurev-psych-120710-100504
Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). 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, 21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003
Pennycook, G., Cheyne, J. A., Barr, N., Fugelsang, J. A., & Koehler, D. J. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making, 10*(6), 15.
Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 1050–1054. https://doi.org/10.1073/pnas.0706929105
Pulizzi, J. (2012). The rise of storytelling as the new marketing. *Publishing Research Quarterly, 28*(2), 116–123. https://doi.org/10.1007/s12109-012-9264-5
Raj, M., Berg, J. M., & Seamans, R. (2023). *Art-ificial Intelligence: The Effect of AI Disclosure on Evaluations of Creative Content* (SSRN Scholarly Paper No. 4369818). https://doi.org/10.2139/ssrn.4369818
Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition and Emotion, 18*, 1–28. https://doi.org/10.1080/02699930244000390
Rusu, M. (2017). Empathy and communication through art. *Review of Artistic Education, 14*(1), 139–146. https://doi.org/10.1515/rae-2017-0018
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
Schepman, A., & Rodway, P. (2022). the general attitudes towards Artificial Intelligence Scale (GAAIS): confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2022.2085400
Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2022). AI composer bias: listeners like music less when they think it was composed by an AI. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000447
Sherman, A., & Morrissey, C. (2017). What is art good for? The socio-epistemic value of art. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2017.00411
Silvia, P. J. (2005). Cognitive appraisals and interest in visual art: exploring an appraisal theory of aesthetic emotions. *Empirical Studies of the Arts, 23*(2), 119–133. https://doi.org/10.2190/12AV-AH2P-MCEH-289E
Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts, 9*, 376–384. https://doi.org/10.1037/aca0000028
Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto empathy questionnaire. *Journal of Personality Assessment, 91*(1), 62–71. https://doi.org/10.1080/00223890802484381
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729
Tröndle, M., Kirchberg, V., & Tschacher, W. (2014). Is this art? An experimental study on visitors' judgement of contemporary art. *Cultural Sociology, 8*(3), 310–332. https://doi.org/10.1177/1749975513507243
Turpin, M., Walker, A., Kara-Yakoubian, M., Gabert, N., Fugelsang, J., & Stolz, J. (2019). Bullshit makes the art grow profounder. *Judgment and Decision Making, 14*, 658–670.
Wilson, E. A. (2011). *Affect and artificial intelligence*. University of Washington Press.
Winner, E. (1982). *Invented worlds: The psychology of the arts*. Harvard University Press.
Wu, Y., Mou, Y., Li, Z., & Xu, K. (2020). Investigating American and Chinese subjects' explicit and implicit perceptions of AI-generated artistic work. *Computers in Human Behavior, 104*, 106186. https://doi.org/10.1016/j.chb.2019.106186
Xu, K., Liu, F., Mou, Y., Wu, Y., Zeng, J., & Schäfer, M. S. (2020). Using machine learning to learn machines: A cross-cultural study of users' responses to machine-generated artworks. *Journal of Broadcasting & Electronic Media, 64*(4), 566–591. https://doi.org/10.1080/08838151.2020.1835136
Zulić, H. (2019). How AI can change/improve/influence music composition, performance, and education: Three case studies. *INSAM Journal of Contemporary Music, Art and Technology, 2*(1), 100–114.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com