Cognition 257 (2025) 106063

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

AI contextual information shapes moral and aesthetic judgments of AI-generated visual art☆

Ionela Bara [a,*], Richard Ramsey [a,b], Emily S. Cross [a,*]

[a] Social Brain Sciences Group, Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland
[b] Neural Control of Movement Group, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
AI-generated art
Moral judgments
Aesthetic judgments
Contextual information
Implicit moral associations

ABSTRACT

Throughout history, art creation has been regarded as a uniquely human means to express original ideas, emotions, and experiences. However, as Generative Artificial Intelligence reshapes visual, aesthetic, legal, and economic culture, critical questions arise about the moral and aesthetic implications of AI-generated art. Despite the growing use of AI tools in art, the moral impact of AI involvement in the art creation process remains underexplored. Understanding moral judgments of AI-generated art is essential for assessing AI's impact on art and its alignment with ethical norms. Across three pre-registered experiments combining explicit and implicit paradigms with Bayesian modelling, we examined how information about AI systems influences moral and aesthetic judgments and whether human art is implicitly associated with positive attributes compared to AI-generated art. Experiment 1 revealed that factual information about AI backend processes reduced moral acceptability and aesthetic appeal in certain contexts, such as gaining financial incentives and art status. Experiment 2 showed that additional information about AI art's success had no clear impact on moral judgments. Experiment 3 demonstrated that an implicit association task did not reliably link human art with positive attributes and AI art with negative ones. These findings show that factual information about AI systems shapes judgments, while different information doses about AI art's success have limited moral impact. Additionally, implicit associations between human-made and AI-generated art are similar. This work enhances understanding of moral and aesthetic perceptions of AI-generated art, emphasizing the importance of examining human—AI interactions in an arts context, and their current and evolving societal implications.

1. General introduction

Throughout history, art creation has been thought to be a uniquely human activity and viewed as an expression of creativity and ingenuity (Barasch, 1990; Dissanayake, 1995; Goldman, 2001; Ramachandran & Hirstein, 1999). Masterpieces, such as Leonardo da Vinci's *Mona Lisa* and Vincent van Gogh's *The Starry Night* are celebrated globally for their technical and artistic mastery, as well as their ability to encapsulate our shared cultural heritage and identity as species. However, as Generative Artificial Intelligence (GAI) gains prominence in shaping our visual, aesthetic, legal, and economic culture, important questions arise, such as whether it is morally acceptable to use AI tools to produce, exhibit and commercialize AI art. Despite AI's increasing prevalence in the arts, empirical research exploring the ethical implications of using AI in art creation remains limited. Understanding people's moral perceptions of AI-generated art is essential for untangling its broader societal impact. The current project investigates timely questions concerning the moral implications of using AI-systems to produce art and gain financial and social or cultural benefits. By doing so, we aim to provide novel insight into the ethical dynamics between AI systems and art creation, authorship, and aesthetic appreciation.

GAI is a type of artificial intelligence that uses generative models to produce text, create images, and compose music or videos based on different prompts (Anantrasirichai & Bull, 2022; Watson, 2019). AI generative models typically learn patterns and structures from data and generate new data resembling previously learned patterns (Watson, 2019). Regarding the arts domain, some examples refer to text-to-image AI generation systems, such as Stable Diffusion, Midjourney and DALL-E and text-to-video AI generators, such as Lumen5, Midjourney or OpenAI (Midjourney, 2023; OpenAI, 2024; Rombach, Blattmann, Lorenz, Esser,

☆ Special Issue in Honour of Jacques Mehler, Cognition's founding editor.

* Corresponding authors.
E-mail addresses: iobara@ethz.ch (I. Bara), ecross@ethz.ch (E.S. Cross).

https://doi.org/10.1016/j.cognition.2025.106063
Received 1 August 2024; Received in revised form 17 December 2024; Accepted 6 January 2025
Available online 16 January 2025

& Ommer, 2021; Roumeliotis & Tselikas, 2023; Shahriar, 2022; Thoppilan et al., 2022). The use of AI for artistic purposes has provoked wide-ranging public debate over its moral and artistic implications. Concerning the art market, the 2018 auction of *Portrait of Edmond Belamy* for $432,500 at Christie's highlighted AI's potential to reshape the financial landscape of art market by providing new opportunities for economic growth (Hitti, 2018). Initially, media outlets, such as Reuters and NDTV celebrated this breakthrough (Goldberg, 2018), but over time, public reactions have been mixed. A recent example is Jason M. Allen's use of AI to win a digital art prize in 2022, which not only raised accusations of cheating, but also concerns about moral fairness in artistic competitions (Roose, 2022). Critics have argued that awarding a prize to an art generated by AI undermines the efforts of human artists who rely on skill, creativity, and years of practice. The use of AI in this context can be seen as a shortcut, bypassing the time and dedication that traditional artists invest in their skills. Together, these ideas highlight a critical moral debate regarding AI's impact on artistic integrity and fairness. Investigating the moral acceptability of AI-generated art is essential to addressing these tensions. Such inquiries can help clarify the role of AI in artistic expression, ensure ethical standards are maintained, and guide future integration of technology in the creative industries.

People's perceptions of AI-generated art have yielded mixed findings in empirical research. Studies examining the ability of novice observers to distinguish between human-made art and AI-generated art have consistently shown low accuracy rates in detecting AI-generated art (Chamberlain, Mullin, Scheerlinck, & Wagemans, 2018; Darda, Carre, & Cross, 2023; Darda & Cross, 2022; Gangadharbatla, 2021; Samo & Highhouse, 2023). Moreover, across various artistic domains (e.g., visual art, dance, and music), research has demonstrated a clear aesthetic bias against AI-generated art when people are aware of its AI origin (Chamberlain et al., 2018; Darda et al., 2023; Darda & Cross, 2022; Di Dio et al., 2023; Hong & Curran, 2019; Hong, Peng, & Williams, 2021; Moffat & Kelly, 2006; Samo & Highhouse, 2023). These findings suggest that even if people cannot easily detect when an artwork has AI versus human origins, any knowledge of an artwork's artificial origins has a marked negative impact on its emotional value and aesthetic appeal.

This negative bias against AI-generated art can be linked to what is perceived as violations of traditional art definitions. Art has been regarded as inherently human, involving intentional creativity, and cultural conventions (e.g., techniques, styles, mediums of presentation) to express original ideas, emotions, and experiences (Chatterjee, 2022; Davies, 2012; Dissanayake, 1995; Gaut & McIver Lopes, 2001; Shao, Zhang, Zhou, Gu, & Yuan, 2019; Young, 2001). AI-generated art challenges this view, unless, of course, one considers that generative AI models are trained on countless exemplars of human-made art. Furthermore, the negative attitudes people display toward AI-generated art may originate from fears of AI surpassing and replacing human artists in producing art faster and with superior quality and ingenuity (Cha et al., 2020; Hong & Curran, 2019; Tubadji, Huang, & Webber, 2021). For some, AI-generated art threatens beliefs in human exceptionalism and human supremacy in favouring human creativity over machines (Gunkel, 2017; Millet, Buehler, Du, & Kokkoris, 2023; Sawyer, 2012; Schmitt, 2020). Moreover, the lack of knowledge about AI systems appears to modulate the negative bias against AI-generated art. Research by Latikka, Bergdahl, Savela, and Oksanen (2023) has identified unfamiliarity with AI tools, concerns about authenticity, warmth, and safety as the main explanatory variables for negative perceptions of AI-generated art. This indicates that the interplay between art creation and AI-systems is complex and reflects deep-seated beliefs about the nature of art and attitudes toward technology's impact on human innovation and creativity.

Despite the growing use of AI tools to produce art, little research has examined their moral consequences. Investigating moral judgments of AI-generated art is essential for understanding people's perceptions of AI's role in art production, as well as its aesthetic, economic, cultural and ethical implications. Moral judgments and moral responsibility represent cornerstones of moral psychology (Ellemers, van der Toorn, Paunov, & van Leeuwen, 2019; Ladak, Loughnan, & Wilks, 2023; Malle, 2021). Moral evaluations involve assessing an individual's intentions, actions, and outcomes, which can vary based on context or individual differences in reasoning and empathy. These evaluations can be further shaped by perceived harm, social norms, and the level of intentionality shown by a responsible person or agent (Bonnefon, Rahwan, & Shariff, 2024; Malle, Guglielmo, & Monroe, 2014; Malle & Knobe, 1997). The moral responsibility of AI-systems is a multifaced construct that has been primarily linked to whether AI systems or AI-imbued agents are perceived to have agency and motivation, their own mind, and the ability to act intentionally, so that they can be held morally accountable for an event or be assigned blame or punishment (Bonnefon et al., 2024; Gray, Gray, & Wegner, 2007; Gray & Wegner, 2012; Guglielmo, 2015; Ladak et al., 2023; Malle et al., 2014).

To date, only very limited research has explored issues related to moral judgments of AI-systems and visual art. A study by Epstein, Levine, Rand, and Rahwan (2020) used vignettes to describe AI as agentic/anthropomorphised or tool-like/non-anthropomorphised. Framing AI as an agent rather than a tool shifted participants' views on responsibility for AI-generated art. When seen as an agent, participants attributed less responsibility to the artist and more to the programmer and AI. Conversely, viewing AI as a tool led to greater responsibility assigned to the technologist using the AI system. This suggests that changing the perceived agency of AI-systems plays a central role in the responsibility attribution of AI-generated art. Furthermore, Lima, Zhunis, Manovich, and Cha (2021) examined whether evaluating AI's artistic agency before or after viewing AI-generated art images impacts the moral standing of AI systems. Evaluating AI's artistic agency before viewing AI art images resulted in greater artistic agency ratings of AI systems rather than after viewing AI-generated art. This highlights that people's perceptions of AI's artistic agency and moral standing are shaped by how we frame and contextualize information about the roles and creative contributions of AI-systems.

Despite these examples, the moral acceptability of using AI tools to generate art and their implications for AI-generated art's status, artistic acclaim, financial incentives, and aesthetic appraisals remain largely unexplored. Given that moral psychology research has focused on how actions or intentions are judged as morally justified (Malle et al., 2014; Tepe & Byrne, 2022), applying moral acceptability judgments to explore the use of AI in artistic production is timely and valuable for several reasons. First, it allows us to evaluate the ethical implications of using AI in a domain traditionally associated solely with human creativity (Collingwood, 1958; Dissanayake, 1995; Graham, 2005). Second, it addresses issues including authorship, art status, and cultural appropriation (Carroll, 2000; Coombe, 1998; Danto Arthur, 2013; Gaut & McIver Lopes, 2001; Young, 2010), potentially informing the development of ethical guidelines for AI-generated art creation and consumption. In addition, understanding the moral implications of AI systems gaining recognition and fame holds importance given that historically, artistic acclaim has been linked to human creativity, laborious expert skill and practice, originality and authenticity (Barasch, 1990; Gaut & McIver Lopes, 2001; Newman & Bloom, 2012). Moreover, understanding the moral considerations of gaining financial incentives from commercializing AI-generated art is essential in assessing issues related to intellectual property rights and authorship (Coombe, 1998). This can shed light on the development of regulatory frameworks that promote fairness and transparency in the commercialization of AI-generated art.

Furthermore, information processing models emphasize the importance of contextual factors, such as relevant information in shaping moral judgments as they provide essential detail for people to evaluate the ethical implications of actions or decisions (Guglielmo, 2015; Malle et al., 2014). Contextual information also facilitates recontextualization and reinterpretation, allowing individuals to integrate contextual details with pre-existing beliefs and perceptions (Mann & Ferguson, 2015). This dynamic process highlights the context-dependent nature of moral

judgments, where new information can bias the perceived moral acceptability of an action or behaviour (Cone, Mann, & Ferguson, 2017). This is also consistent with theoretical and empirical research from social psychology on social influence and belief change, which shows that the framing of information can shift beliefs in different directions. Negative biases are more likely to emerge when information emphasizes risks or undesirable outcomes, while positive biases frequently occur when information highlights benefits or successes (Kahneman & Tversky, 1984; Nelson, Oxley, & Clawson, 1997).

In terms of how to investigate such perceptions, the study of moral judgments has traditionally used both explicit and implicit measures. Explicit moral judgments are considered deliberate, conscious, and reflecting higher-order cognitive processes (De Houwer, 2006; Yoder, Harenski, Kiehl, & Decety, 2015). These judgments are often shaped by societal norms, cultural values, and moral frameworks transmitted through social institutions (Haidt, 2001). Explicit measures, such as Likert scales or forced-choice responses, provide a window into how people intentionally align with or challenge prevailing moral norms, highlighting the role of morality as a social construct. However, these judgments are often susceptible to biases such as social desirability and cultural conformity. In contrast, implicit measures assess automatic and often unconscious responses to moral situations, which are believed to reveal covert associations shaped by life experiences (Bornstein & Pittman, 1992; Cameron, Payne, Sinnott-Armstrong, Scheffer, & Inzlicht, 2017; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). Tools such as the Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2005) or the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001) have been successfully used in moral or empirical aesthetics research (Bertamini, Palumbo, Gheorghes, & Galatsidas, 2016; Palumbo, Ruta, & Bertamini, 2015). In the context of AI-generated art, explicit measures would facilitate an intentional and direct examination of the moral considerations involved. In contrast, implicit measures would reveal the subtle biases people may have for or against AI-generated art, providing insights into underlying, less overt dynamics of moral judgment.

Given the prominent role of AI systems in generating visual art images using conventional artwork databases and their moral, societal, cultural and economic implications, the current study aims to investigate people's moral judgments of AI-generated art images and their relationship with aesthetic preference. Across three pre-registered experiments, using explicit and implicit paradigms, we examine the extent to which:

1) providing factual information about how AI-generated images are created impacts moral judgments concerning the use of AI algorithms, financial gain, prestige, art status, aesthetic appreciation (Experiment 1);
2) varying information doses (factual, low, medium, and high) regarding the success of AI-generated art impact moral judgments related to financial gain, prestige, and art status (Experiment 2); and
3) human art is implicitly associated with "good" word attributes while AI-generated art is associated with "bad" word attributes (Experiment 3).

Across these three experiments, this study provides insights into current moral and aesthetic attitudes toward AI-generated art, acknowledging that as AI-generated creative outputs become ever more ubiquitous these attitudes are likely to transform in the coming years.

2. Experiment 1

2.1. Introduction

Here we investigated whether different moral and aesthetic judgments of AI-generated art are impacted by factual information on how AI images are produced. We predicted that participants would rate the use of AI algorithms in art creation, and the financial and prestige gains from AI-generated art as less morally acceptable, and that participants would assign lower ratings to the status and aesthetic appeal of AI-generated art as a function of:

a) providing factual information on how the AI images were produced. Specifically, information on how the AI system generates images would lead to less moral acceptability and reduced aesthetic ratings than when providing no information; and
b) the type of AI-generated images. Particularly, AI-generated depicting people would lead to reduced ratings of moral acceptability and reduced aesthetic judgments than AI-generated images depicting landscapes.

The first hypothesis stems from the idea that moral judgments are flexibly updated by known information or the context at hand (Andrejević, Feuerriegel, Turner, Laham, & Bode, 2020; Bartels, Bauman, Cushman, Pizarro, & McGraw, 2015; Slothuus, 2008). Indeed, research has demonstrated that moral decision-making is impacted by different contextual and wording framing effects (Petrinovich & O'Neill, 1996). Furthermore, the first hypothesis also builds upon previous research demonstrating that contextual information actively reshapes moral judgments by enabling reinterpretation. This process impacts how individuals integrate new information into their pre-existing beliefs, often introducing bias into moral decisions (Cone et al., 2017; Mann & Ferguson, 2015). In that sense, Kuklinski, Quirk, Schwieder, and Rich (1998) showed that factual information influences decision-making by reducing uncertainty associated with a lack of knowledge, clarifying ambiguous topics, and expanding upon limited existing knowledge. Applying this to our hypothesis, providing factual information about the backend processes of AI-generated art may lead to a negative bias toward its moral acceptability compared to situations where no such information is provided. This effect may occur possibly because factual information can help recontextualize and potentially accentuate pre-existing skepticism about AI's abilities in artistic and creative domains, or because it expands or enhances previously limited understanding.

The second hypothesis is informed by general social cognition accounts by which human faces and bodies are important signs of social communication, conveying meaning about unique human identities (Burton, Kramer, Ritchie, & Jenkins, 2016; Frith & Frith, 2023; Rossion, 2022). However, advances in AI technologies able to replicate and appropriate people's identity (e.g., facial, bodily or vocal aspects) without their consent, pose new challenges to human dignity and raise moral concerns about identity violation, including identity exploitation, discrimination, theft and fraud (Dunn, 2020). For example, AI-generated art depicting human figures may tap into deeply ingrained mechanisms of person perception, where viewers automatically attribute intentions, agency, and moral accountability to these representations. In contrast, AI-generated depictions of landscapes or abstract concepts may raise fewer moral concerns, as they lack a direct connection to individual identity and its associated rights. While this specific hypothesis is exploratory (as it has not yet been empirically tested), we argue that it is indirectly supported by social cognition research, which highlights the importance of authentic identity signals in building trust and ensuring moral accountability. Furthermore, the fast-growing body of research on AI-generated art and its moral implications makes it timely to situate our hypothesis within this evolving context, where AI technologies and person perception carry important societal and moral ramifications. Therefore, we anticipated that AI-generated artworks that featured humans would be associated with amplified moral unacceptability and compromised aesthetic quality.

2.2. Method

2.2.1. Pre-registration and Open Science statement

Prior to data collection, the research questions, hypotheses, planned analyses, sample sizes and exclusion criteria were pre-registered (https://osf.io/835zf). In addition, consistent with recent metascience recommendations (Munafò et al., 2017), all raw data, stimuli, and analysis code for each experiment are openly available on the open science framework (https://osf.io/d4zme/).

2.2.2. Participants

All participants in Experiment 1 were recruited online from Prolific (https://www.prolific.com/) in exchange for payment (£9/h – recommended Prolific rates). All participants provided informed consent, were pre-screened for English fluency (self-reported native fluency), normal or corrected-to-normal vision, 100 % approval rate on Prolific's system, and had completed at least 100 prior tasks on Prolific. All the experimental procedures for Experiment 1 were granted ethical approval by ETH Zurich Ethics Commission. As pre-registered, the sample size was determined by the largest participant number we could recruit given the resources available for multiple connected experiments. This approach is consistent with (Lakens, 2022) who emphasised that sample sizes in research are inherently constrained by available resources. Given that, we aimed to test 50 participants per group (info vs. no info about AI-systems), which represents 100 participants in total after exclusions. As exclusion criteria, we pre-registered that we will exclude participants who fail two mandatory attention checks.

The attention checks involved clear and straightforward instructions and followed Prolific's recommendations for fair and transparent practices in evaluating participants' engagement across research tasks. For example, participants were prompted to select either "not at all" or "very" to answer all the questions on one question screen. Participants were randomly recruited to one of the two experimental groups. One hundred fifty-two participants (97 females, 55 males, $Mean_{age} = 38.64$, $SD_{age} = 11.06$) were recruited for the AI art images information group. Ninety-two participants (49 females, 43 males, $Mean_{age} = 35.65$, $SD_{age} = 9.87$) were recruited for no information group. After exclusions, the final sample consisted of 100 participants in total (50 participants for the AI images information group – 28 females, 22 males. $Mean_{age} = 37.84$, $SD_{age} = 11.18$, and 50 participants for no information group (25 females, 25 males. $Mean_{age} = 34.76$, $SD_{age} = 9.23$). The exclusion rate, particularly in the information group where participants were required to read longer texts, was notably high. This likely highlights on going challenges associated with online data collection environment. However, empirical research on online data quality has emphasised the critical importance of routinely incorporating attention checks, such as those implemented in this experiment. These checks are essential for identifying careless responses (Douglas, Ewell, & Brauer, 2023; Muszyński, 2023).

2.2.3. Stimuli, design, tasks and procedure

Stimuli. Experiment 1's stimuli included 40 AI art images (20 landscapes and 20 depicting people) generated using DALL-E3 by OpenAI (openai.com). All AI art images were created using textual prompts based on Impressionist artworks by Spanish artist Joaquín Sorolla. For more information, please see Supplementary Material (Exp.1, Section A). An example of AI-generated images used in Experiment 1 can be seen in Fig. 1. All the AI-generated images and their corresponding textual prompts are available on our open science framework page (https://osf.io/d4zme/).

Design. Experiment 1 used a 2 (image type: people, landscapes) x 2 (group: no info, info) mixed within- and between-participant design. The image type was a within-participant factor, meaning that all

Fig. 1. Example of AI-generated images.

participants completed moral judgments (AI algorithms use, financial gain, prestige, art status) and aesthetic appreciation judgments of AI-generated images depicting people and landscapes. In contrast, the group was a between-participant factor, meaning that participants were randomly assigned to one of two information groups. While the no information group viewed and made moral and aesthetic judgments of AI art images accompanied by no information, the information group was provided with a brief text about how the AI-images were produced (please see the full text in Fig. 2).

Tasks and Procedure. All the tasks in Experiment 1 were produced in Qualtrics (https://www.qualtrics.com/). The moral and aesthetic judgment task involved participants from info and no info group rating all 40 AI-generated art images on five dependent variables (AI algorithms use, financial gain, prestige, art status, aesthetic appreciation). All ratings were assessed on a 5-point Likert scale (1–5; not at all – extremely). The AI art images remained on the screen until participants made a rating response. The order of the AI images and rating questions was fully randomized across participants. Regardless of the group, all participants completed 200 AI image ratings in total. An example of an experimental trial can be seen in Fig. 2.

Following the rating task, all participants completed a series of questionnaires. We report these exploratory results in the Supplementary material (Exp.1, Section C, Figs. S5-S11).

2.2.4. Data analyses

We preregistered a multilevel Bayesian estimation similar to previous work (Bara, Cross, & Ramsey, 2023; Bara, Darda, Kurz, & Ramsey, 2021; Bara, Ramsey, & Cross, 2024). Primarily, we reported and discussed the posterior distribution of our key parameters within the full model, which had the maximum number of varying parameters that the design allowed. In the full model, we discussed the posterior distributions of key parameters, identifying the point of highest density

This image was generated by an AI algorithm that produces images from textual descriptors. To accomplish that, several steps are required: First, the AI algorithm is trained by learning a large dataset of art images and their corresponding text descriptors, such as the artist's name. Then, the AI algorithm is able to generate new images based on different textual prompts (e.g., artist's name, artistic style, whether it depicts a seascape, landscape, or people).

	Not at all	Slightly	Moderately	Very	Extremely
Is it morally acceptable to **use AI algorithms** to generate this art image?	○	○	○	○	○
Is it morally acceptable to **derive financial gain** from selling this AI-generated art image in art auction houses?	○	○	○	○	○
Is it morally acceptable **to label** this AI-generated art image **as conventional artwork**?	○	○	○	○	○
How much do you **aesthetically appreciate** this AI-generated art image?	○	○	○	○	○
Is it morally acceptable to **gain prestige** by exhibiting this AI-generated art image in art museums?	○	○	○	○	○

Fig. 2. Example of experiment trial in the info group.

(median) and the lower and upper bounds of the 95 % quantile intervals. Therefore, our interpretations are grounded in the posterior (density) distributions and the 95 % and 66 % quantile intervals, which represent the range of most plausible values for each parameter.

In practical terms, we implemented a translation of McElreath's (2020) general modelling principles (Kurz, 2020) using the Bayesian modelling package 'brms' to build multi-level models (Bürkner, 2017, 2018). Similar to previous research work (Bara et al., 2021; Bara et al., 2023; Bara et al., 2024), we adopted a weakly informative approach for setting priors (Gelman, 2006) as detailed in the Supplementary material (Exp1, Section A, TableS1). Moreover, for data wrangling we used the 'tidyverse' principles (Wickham & Grolemund, 2016) and we generated plots using the associated data plotting package 'ggplot2', as well as the 'tidybayes' package (Kay, 2020). All analyses were conducted in the R programming language (Version 4.4.0; RCore Team, 2024). Given that the dependent variables are an ordered category (a 1–5 rating scale), we used an ordinal regression model. The formula is provided below:

```
brms formula = bf(mvbind(AI-algorithms use, financial gain, prestige, art
status, aesthetic appreciation) | thres(4, gr=item)
        ~ 1 + info type * image type +
        (1 |p| item) +
        (1 + image type |a| participant))
```

Note: info type = info vs. no info; image type = landscapes vs. people; item = stimuli. We acknowledge a small deviation from our preregistration, due to an error in its formulation. That is, the model above omits the random effects for "info type" by "participant".

2.3. Results

Rating summary data for all five dependent variables (AI-algorithms use, financial gain, prestige, art status, aesthetic appreciation) across factual information and no information groups and for AI art images depicting people and landscapes are shown below (Fig. 3).

The posterior estimates across all five dependent variables are shown in Fig. 4 and Supplementary Tables (Exp.1, Section B, Tables S2). While we visualise the full model, we only discuss the main pre-registered parameters of interest that address our key hypotheses, specifically the main effects of group and image type. For the average effect of group (info>none), only for financial gain did the 95 % quantile intervals of the posterior distribution exclude zero, indicating a clear and consistent response. The posterior distribution's shift toward the left, with no overlap at zero, indicates that participants in the no information group perceived gaining financial incentives from AI-generated art as more morally acceptable compared to those in the information group. This also indicates the converse: participants in the information group perceived financial gain as less morally acceptable compared to those in the no-information group. Furthermore, for art status, AI use, and prestige, the effects were weaker, as their 95 % quantile intervals included zero. However, the 66 % quantile intervals for these DVs did not overlap zero, indicating a less robust but a trend in the same direction (Info > None). This suggests that AI-generated art was perceived as less morally acceptable regarding its financial gain, its prestige and its art status in the info group rather than no info group. In addition, the posterior distribution for aesthetic appreciation showed no clear levels of aesthetic appeal across both groups.

Regarding the effect of image type (people>landscape), across all DVs, the posterior distribution showed no clear response, as both the 95 % and 66 % quantile intervals overlapped zero. This suggests that participants rated AI-generated landscapes and people images similarly on moral acceptability and aesthetic appeal. Taken together, the results suggest that AI-generated art is deemed less morally acceptable when factual information about AI-system operations is provided, as opposed to when no such information is given, across financial gain, prestige, and art status.

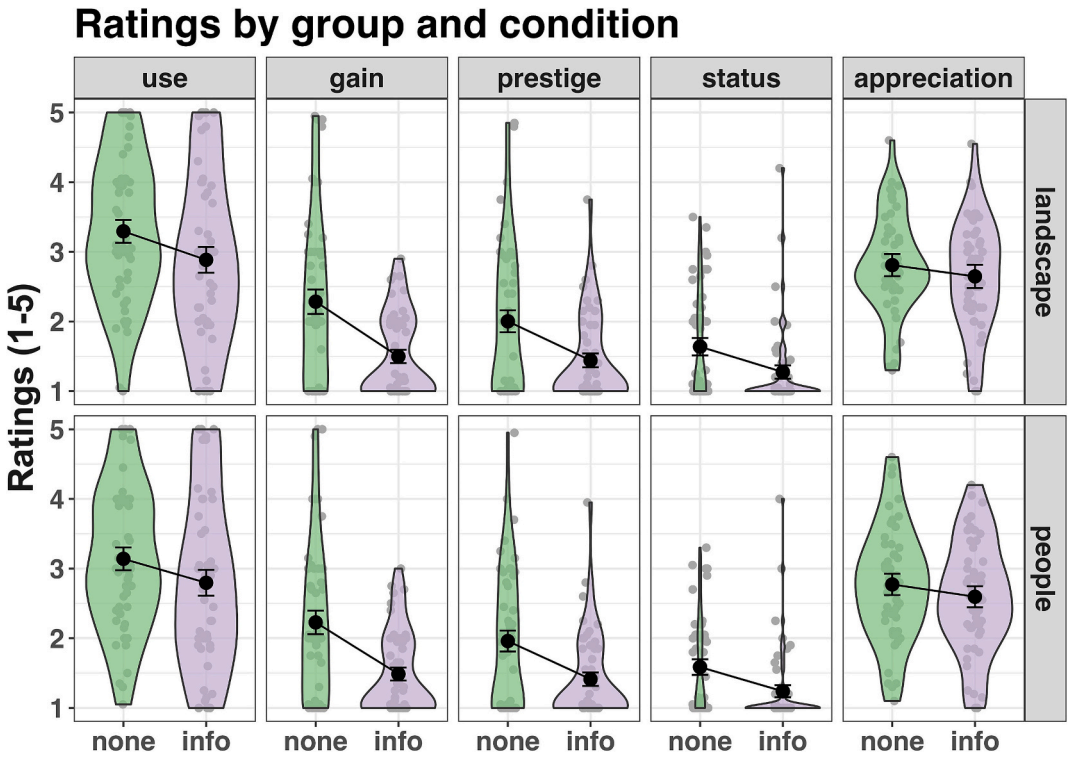

Fig. 3. Ratings across group (no info vs. info) and image type (landscapes vs. people) for all five DVs. The rows show the ratings across landscape and people. The five columns illustrate our main DVs: use = AI-algorithms use, gain = financial gain, prestige, status = art status, appreciation = aesthetic appreciation. The ratings are reported on a 5-point Likert scale (1 = not at all morally acceptable to 5 = extremely morally acceptable). Error bars represent 95 % confidence intervals. The black markers (circles) and interval estimates represent the group mean average, whereas the grey markers represent the individual participants.

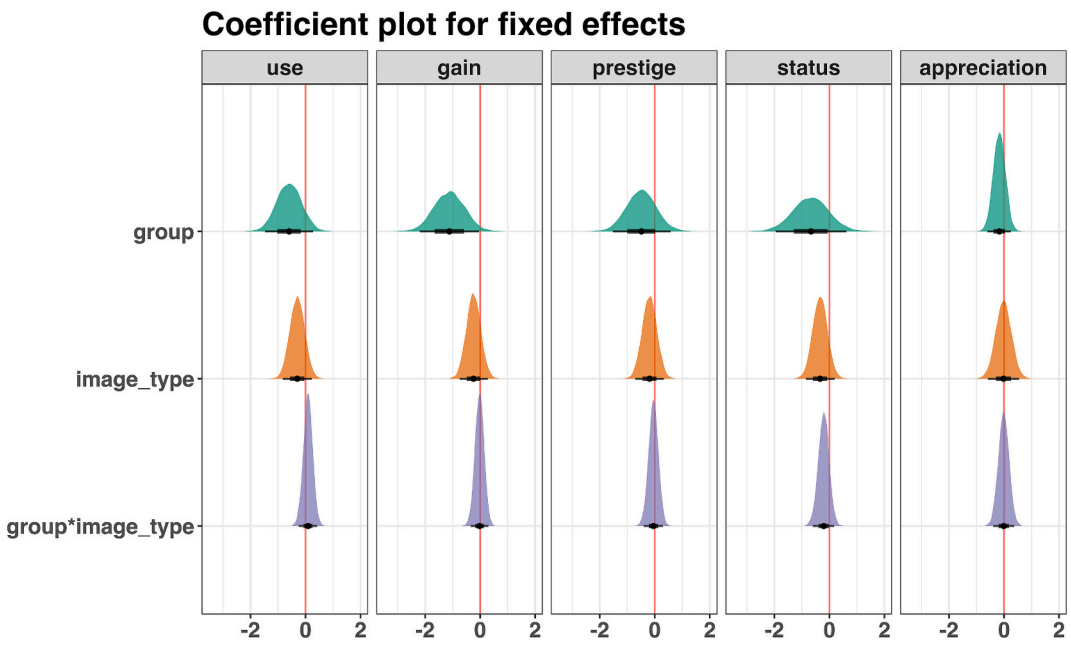

Fig. 4. Multivariate coefficient plot for the full model. The main parameters of interest are: average effect of group (in green) and image type (in orange) across all five DVs (use = AI-algorithm use, gain = financial gain, prestige, status = art status, appreciation = aesthetic appreciation). The coloured half-eye plots (green and orange) indicate the posterior (density) distribution each response category. The width and height of the plots show the probability density of the responses at each level. The point estimate (black dots) represents the median of the posterior distribution for each category. Error bars represent 66 % quantile intervals (thick black lines) and 95 % quantile intervals (thin black lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Discussion

Experiment 1 demonstrated that providing factual information about how AI algorithms generate art results in lower moral acceptability ratings across all five dependent variables compared to providing no information. This supports our hypothesis that providing information about AI's image generative process diminishes its moral acceptability and reduces aesthetic ratings compared to when no information is provided. Although the direction of results was consistent across all DVs, the strength of the effects varied. Financial gain showed the strongest effect, with participants in the information group perceiving AI-generated art as less morally acceptable compared to the no information group. Prestige and art status demonstrated moderate effects, while AI use and aesthetic appreciation showed progressively weaker effects. These findings suggest that moral acceptability judgments of AI-generated art are most strongly shaped by factual information about AI systems in financial contexts, emphasizing the critical role of moral concerns regarding the financial incentives of AI-generated art. Furthermore, contrary to our hypothesis, the content of the AI-generated images (landscape vs. people as subjects) did not differentially impact moral and aesthetic judgments, suggesting that moral and aesthetic judgments were impervious to the depicted content chosen for this experiment.

3. Experiment 2

3.1. Introduction

In Experiment 2, we examined the extent to which different information doses regarding the financial and artistic reception of AI-generated art impact moral judgments. Specifically, we compared three doses of fictional information about the financial and artistic success of AI art (low, medium, and high impact) with factual information about the AI system's backend processes (same information that was presented in Experiment 1). Our aim was to determine how varying information doses affect moral judgments and to identify whether a particular dose is associated with ratings of maximal moral acceptability of AI-generated art. We hypothesized that:

1) participants will assign higher moral acceptability ratings of financial gain, prestige, and art status to AI-generated images when they are associated with information about the artworks' success compared to only factual information on how the AI system generates art images. We therefore expected moral ratings to be higher in one, two or all information levels (low, medium, high) compared to factual information only (reference category); and
2) participants will also assign moral acceptability ratings of financial gain, prestige, and art status to AI-generated images as a function of the strength of the information provided about the artworks' success. That is, we hypothesized a dose-response function, such that artworks that are associated with more success will result in higher moral acceptability ratings.

Our hypotheses are motivated by prior research showing that judgments of moral fairness and acceptability are shaped by accompanying contextual information. For example, Andrejević et al. (2020) demonstrated that participants adjusted their fairness assessments after receiving contextual details about the deservingness of the action recipient. This suggests that moral judgments are dynamically changing as people refine their understanding of a situation given the contextual information. Our hypotheses are further grounded in information processing theories, which highlight the influence of contextual information on moral judgments (Guglielmo, 2015; Malle et al., 2014). Contextual information allows individuals to integrate new knowledge with existing beliefs and perceptions (Mann & Ferguson, 2015), potentially biasing moral acceptability judgments in either positive or negative directions (Cone et al., 2017). Social psychology research further highlights the impact of contextual framing on decision-making, showing that framing risks often results in negative biases, whereas highlighting benefits or successes leads to positive biases (Kahneman & Tversky, 1984; Nelson

et al., 1997). Additionally, our hypotheses draw on research in source credibility and persuasion, which has indicated that information is more impactful from a successful and credible source (Smith, De Houwer, & Nosek, 2013). Building on this, we expect that the financial success and artistic recognition of AI-generated images may enhance their credibility, therefore increasing their influence on moral acceptability judgments.

3.2. Method

The overall methodology was similar to Experiment 1, with key similarities and differences highlighted below.

3.2.1. Pre-registration

The pre-registration for Experiment 2 is accessible at https://osf.io/xd6np. As before, raw data, stimuli, and analysis code are openly available on the open science framework (https://osf.io/d4zme/).

Deviation from pre-registration. We acknowledge a deviation from the pre-registered data analysis plan due to an observed order effect, where ratings varied depending on the sequence of conditions. As such, we focus exclusively on reporting data from block 1, which remains unaffected by these order effects. For transparency and completeness, we report data in full in the supplementary materials (Figs. S12-S14).

3.2.2. Participants

Participants in Experiment 2 were also recruited online via Prolific (https://www.prolific.com/) underwent similar pre-screening as in Experiment 1. The sample size was determined based on the same criteria as Experiment 1. As in Experiment 1, we aimed to test 100 participants, however, to minimize variability arising from individual differences and enable clearer comparisons across conditions, we implemented a within-subject design instead of the between-subjects design (Experiment 1). Also, due to high exclusion rates in Experiment 1 from participants' inability to maintain attention, and given the focus here on varying information doses, we updated the attention checks such that participants were now required to answer two multiple-choice questions per condition to assess their understanding and engagement with task instructions. A total of eight attention checks were used (two percondition), and we pre-registered to exclude participants who failed three or more of the eight checks. One hundred participants (55 females, 45 males, $Mean_{age} = 37.96$, $SD_{age} = 9.71$) were recruited and no participants were excluded for failing more than 2 attention check questions.

3.2.3. Stimuli, design, tasks and procedure

Experiment 2 used the same 40 AI-generated art images from Experiment 1, which included both landscapes and people. However, due to the lack of meaningful effects of image type in Experiment 1, here we did not compare landscapes and people. In addition, building on the findings of Experiment 1, here we focused on three moral judgments, specifically on financial gain, prestige, and art status, as these have demonstrated clear effects in Experiment 1. Unlike Experiment 1, Experiment 2 used a within-participant design with four conditions: factual, low, medium, and high impact information. In the factual condition, participants received the same information about how AI-generated art images are generated as in the information group of Experiment 1 and were asked to rate moral acceptability on 10 AI-generated art images across all three DVs (financial gain, prestige, and art status). This condition was fixed for all participants to establish a baseline for further judgments comparisons.

The low, medium, and high impact information conditions were counterbalanced, with participants rating 10 AI-generated images across all three DVs in each condition. These conditions represented varying levels of artistic and financial success, operationalized as progressively higher "doses" of artistic recognition and monetary value success. Specifically, the low-impact condition discussed a local gallery exhibition with images sold for $100 each, the medium-impact condition presented a state-level art gallery exhibition with images sold for $1000 each, and the high-impact condition featured an internationally renowned auction house where images were sold for $10,000 each. The diagram below illustrates the information presented in each experimental condition (Fig. 5). This design allowed for a systematic examination of participants' moral judgments across increasing levels of AI art success. The order of images and judgment questions (financial gain, prestige, and art status) was randomized across participants. No images were repeated, and each condition featured a unique set of image exemplars, ensuring that participants evaluated different images across all conditions.

As in Experiment 1, all tasks were administered via Qualtrics (https://www.qualtrics.com/), and ratings were collected on a 5-point Likert scale (1–5; not at all – absolutely). AI art images stayed on the screen until participants provided a response. Each participant rated 120 AI images in total. Similar to Experiment 1, all participants completed a series of questionnaires after the rating task. The exploratory results are reported in the Supplementary Material (S17-S24).

3.2.4. Data analyses

We preregistered a similar multilevel Bayesian analytical approach as in Experiment 1, and used to following model formula to address our research questions:

```
bf(mvbind(financial gain, prestige, art status) | thres(4, gr=item) ~
  1 + condition +
  (1 |p| item) +
  (1 + condition |a| participant)))
```

Notes: condition represents the experimental manipulation that reflects the type and degree of information, which has four levels of information (factual; low; medium; high). Condition 1 (factual information) acts as the baseline and is, therefore, the natural reference category; item represents the AI-generated art images across all conditions.

While we have specified the whole model, we are primarily interested in just two specific hypotheses:

1) To test hypothesis 1, we will estimate each level (low, medium, high) to the baseline reference category (factual information). To support hypothesis 1, we expected judgments to be higher in one, two or all levels (low, medium, high) compared to the reference category.
2) To test hypothesis 2, we will compute paired comparisons between the posterior estimates of low, medium and high levels of condition. We expected an increasing impact (low > medium > high) to support a clear dose-response effect.

To assess these effects, we will calculate 95 % quantile intervals and interpret an estimate above zero as an effect in the predicted direction.

3.3. Results

Rating summary data for all three dependent variables (financial gain, prestige, art status, across varying information doses (factual, low, medium, high) for AI-generated art images are shown below (Fig. 6).

The main results, namely the posterior estimates across all three dependent variables for Block1, are shown in Fig. 7 and Supplementary Tables (Exp.2, Section B, Tables S4-S5). We discuss the main pre-registered parameters of interest that address our key hypotheses. Across all information conditions (low, medium, and high) and dependent variables (DVs), the 95 % quantile intervals of the posterior distributions overlapped with zero. This indicates that different levels of information about the artworks' success had no effect on moral acceptability ratings when compared to the baseline (Panel A). However, an interesting trend was observed for the art status DV under the medium information condition, where the 95 % quantile interval

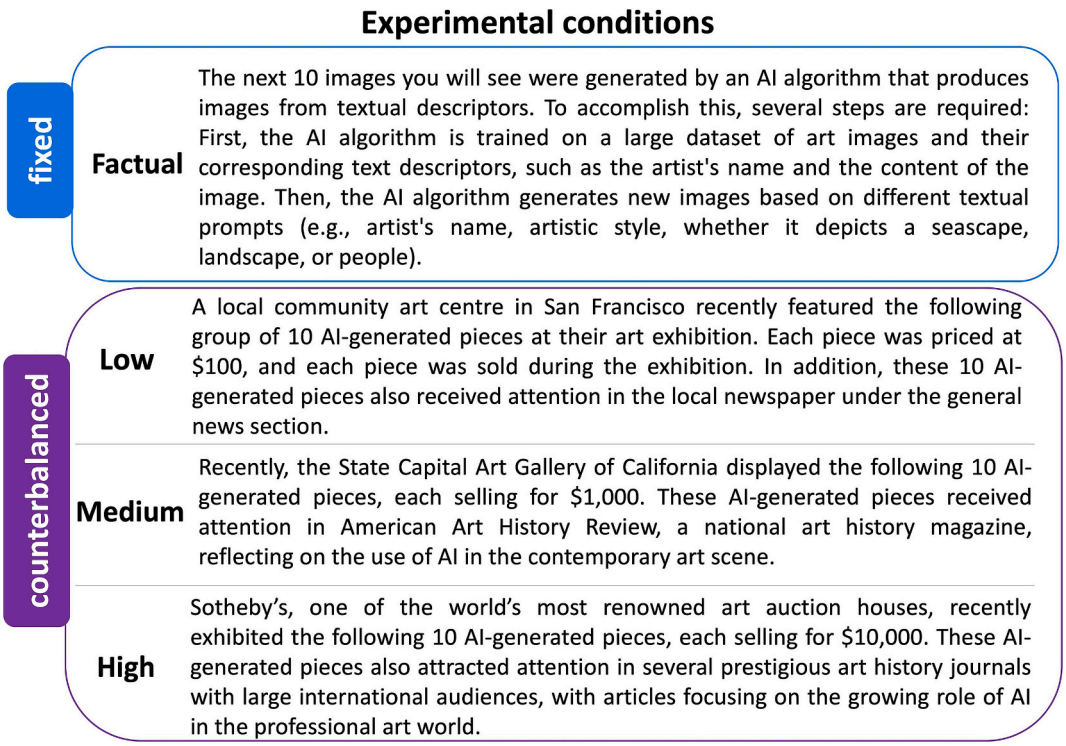

Fig. 5. Example of experimental conditions in Experiment 2.

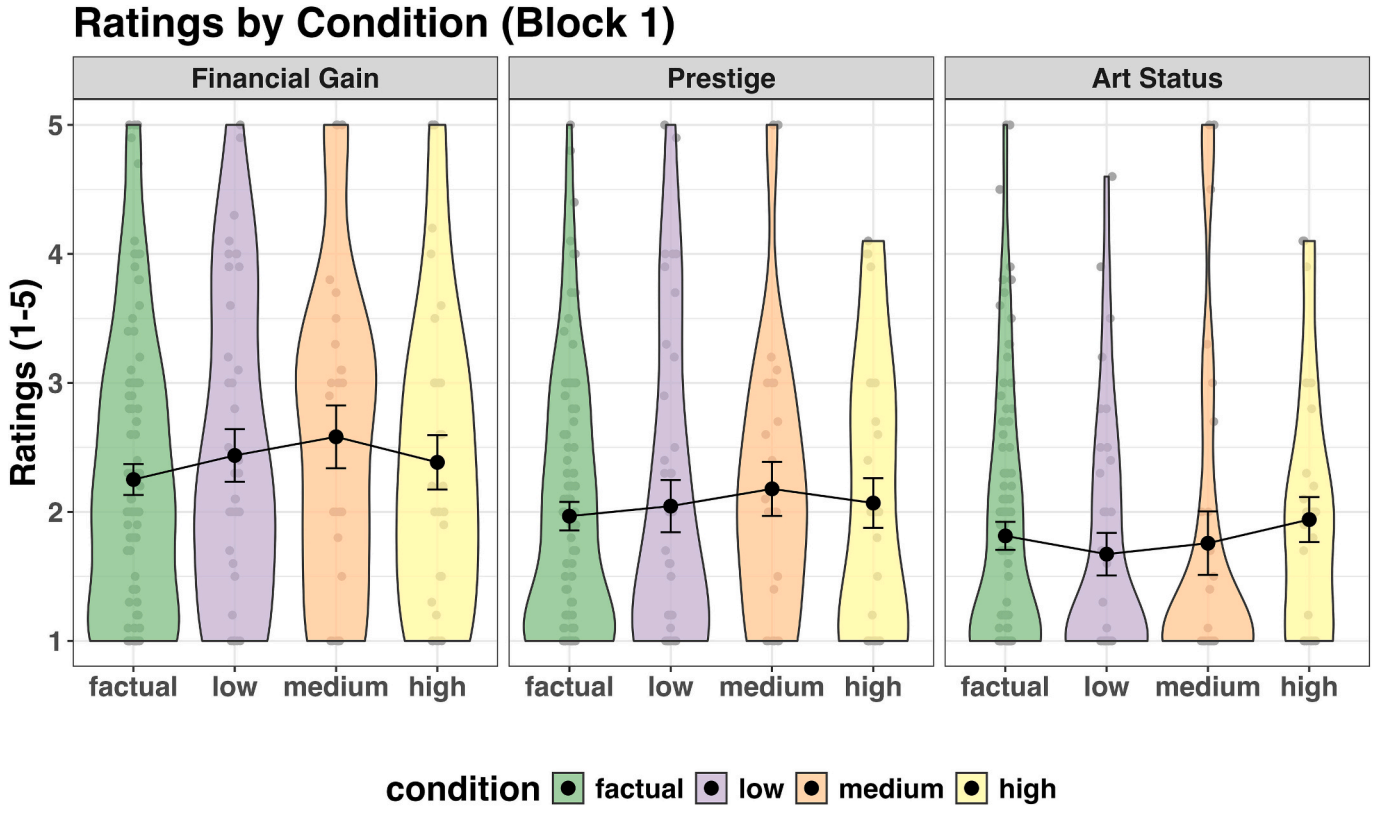

Fig. 6. Ratings across factual, low, medium, high information doses for all three DVs (Block 1). The three columns illustrate our main DVs: financial gain, prestige, art status The ratings are reported on a 5-point Likert scale (1 = not at all morally acceptable to 5 = absolutely morally acceptable). Error bars represent 95 % confidence intervals. The black markers (circles) and interval estimates represent the group mean average, whereas the grey markers represent the individual participants.

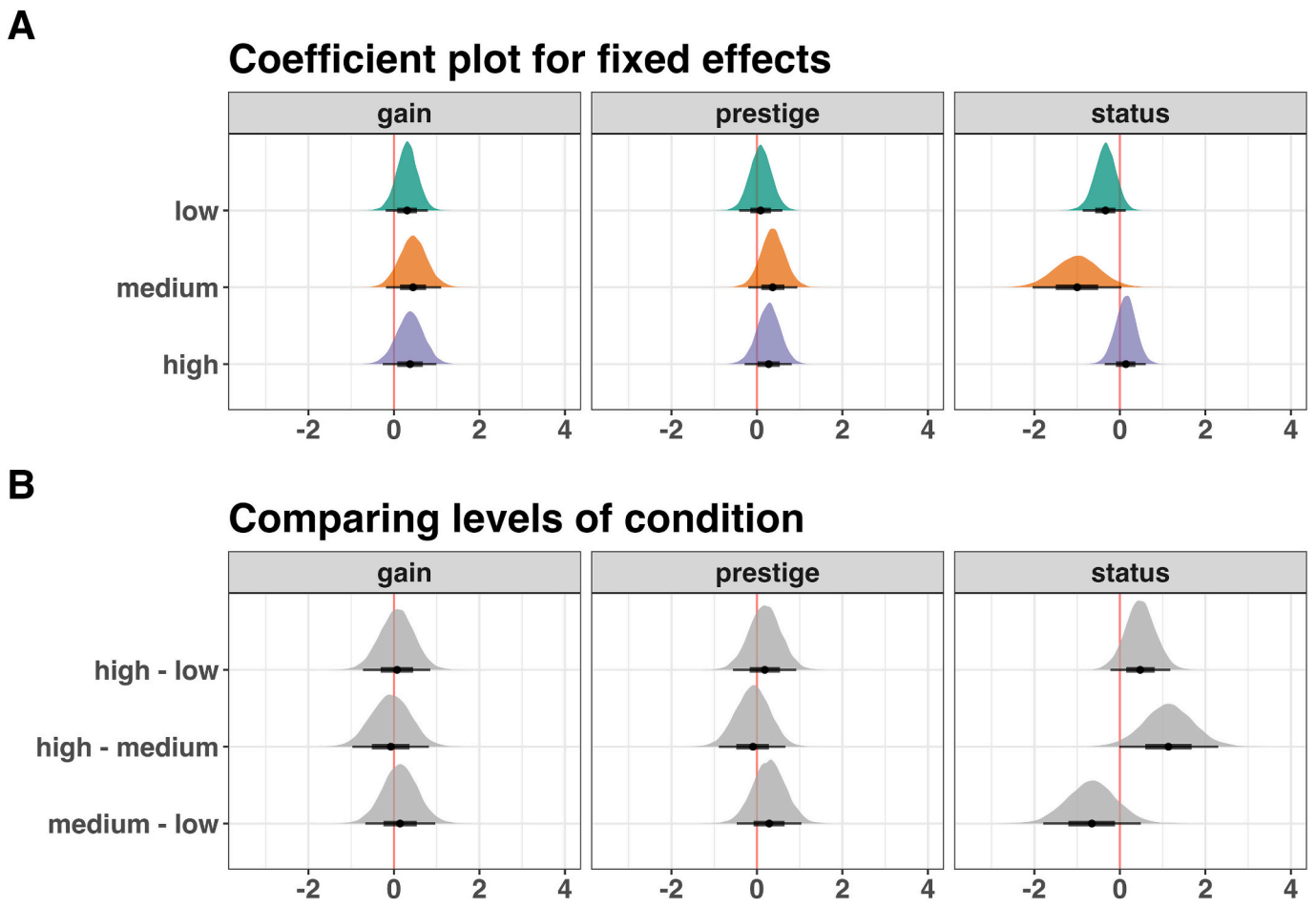

Fig. 7. Multivariate coefficient plot for the full model (Block 1). Panel A shows the posterior distribution for fixed effects across all three DVs (gain = financial gain, prestige, status = art status) and for the low, medium, and high information dose conditions. The vertical red line represents zero, serving as a reference for the effects relative to the baseline (factual condition, not explicitly shown in the figure). The coloured half-eye plots indicate the posterior (density) distribution for each condition (low in green; medium in orange; high in purple). The width and height of the plots show the probability density of the responses at each level. The point estimate (black dots) represents the median of the posterior distribution for each category. Error bars represent 66 % quantile intervals (thick black lines) and 95 % quantile intervals (thin black lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

showed lower moral acceptability ratings compared to the baseline. Further analysis using the 66 % quantile intervals revealed slight increases in moral acceptability ratings for financial gain and prestige DVs.

Regarding the paired comparisons between different information levels (Panel B), the 95 % quantile intervals of the posterior distributions across all DVs included zero, indicating no clear evidence for differences between conditions. The only exception was for the art status DV, where the 95 % quantile intervals showed a tendency for a higher difference of high-medium comparison. When examining a narrower 66 % quantile intervals for art status, there are stronger differences between conditions. Specifically, the high condition showed a greater difference than the medium condition. Also, the medium condition showed lower trends compared to the low condition. Overall, these results suggest that varying levels of information (low, medium, high) did not systematically impact participants' moral acceptability judgments in a positive direction. While the narrower 66 % quantile intervals provide a nuanced view of trends in the data, they do not point to conclusive effects. The lack of consistent effects across information levels, combined with overlaps with zero, highlights the modest nature of these effects.

Panel B shows the posterior distributions for pairwise comparisons between the low, medium, and high information dose conditions for the same DVs. All other details stay the same as in Panel A.

3.4. Discussion

Experiment 2 provided no clear evidence that different information doses about AI artworks' artistic and financial success compared to factual information shaped participants' moral judgments. Specifically, providing information beyond factual details about how AI art is generated did not change the moral perceptions, so that AI-generated art would be perceived as more morally acceptable, even when the artworks' success was deliberately systematically emphasised. The results highlight the complexity of moral judgments surrounding AI-generated art, revealing that its moral acceptability is not easily influenced by information regarding its broader public artistic reception nor by its financial success. The broader implications of these findings are discussed below.

4. Experiment 3

4.1. Introduction

We next sought to explore further questions concerning the moral implications of AI systems in producing art through the use of an implicit speeded reaction time task. While explicitly reported or ranked moral judgments provide a valuable research tool to quantify explicit bias in different moral situations (Graham, Haidt, & Nosek, 2009; Greene et al.,

2009), they are unable to measure implicit biases due to their reliance on self-reported responses. The Go/No-go Association Task (GNAT; Nosek & Banaji, 2001) provides the means to evaluate implicit associations between a target category and an attribute category. Across various experimental settings, GNAT has been demonstrated to be a viable tool in moral psychology research, in measuring implicit associations between morality and food (Lakritz et al., 2022), morality and religion (Pirutinsky, Carp, & Rosmarin, 2017), human and non-human traits (Loughnan & Haslam, 2007) or negative attitudes to substance use disorder (Ashford, Brown, & Curtis, 2019). Using a GNAT tool in the context of AI-generated art is important because it helps uncover implicit associations that may not be captured through explicit measures. Specifically, the GNAT tool could provide novel and valuable insights into how the general public perceives the inherent "good" or "bad" of AI-generated and whether is a perceptual difference to human-made art. For example, if AI art is implicitly associated with "bad" attributes while human art is associated with "good," this could indicate a deeper cultural resistance to accepting AI in the creative process. Such findings would highlight not only biases in how we view different forms of art, but also broader societal concerns about the authenticity, creativity, and moral value of AI-driven works.

Compared to Experiments 1 and 2, the aim of the Experiment 3 was to assess implicit moral associations between human-made art and AI-generated art. Specifically, by using a modified version of GNAT, we investigated the extent to which Impressionist art is implicitly associated with "good" word attributes while Impressionist AI-generated art is associated with "bad" word attributes. We expected:

a) greater sensitivity to Impressionist art and "good" word attributes than Impressionist art and "bad" word attributes.
b) greater sensitivity to Impressionist AI-generated art and "bad" word attributes than Impressionist AI-generated art and "good" word attributes as a function of the quality type of AI-generated art. Specifically, low-quality rather than high-quality of AI-generated art will lead to increased sensitivity to the association between Impressionist AI-generated art and the "bad" words than Impressionist AI-generated art and the "good" words.

4.2. Method

4.2.1. Pre-registration

Similar to Experiments 1 and 2, before data collection started, the research questions, hypotheses, planned analyses, sample sizes and exclusion criteria were pre-registered. The pre-registration for Experiment 3 is available at (https://osf.io/ruhg2). Also, we have made the raw data, stimuli, and analysis code openly available on the open science framework (https://osf.io/d4zme/).

4.2.2. Participants

Participants in Experiment 3 were also recruited online through Prolific (https://www.prolific.com/), and they went through similar pre-screening procedures as Experiments 1 and 2. The sample size was determined on the same considerations as in Experiments 1 and 2. We aimed to test 50 participants per group (high-quality AI art images vs. low-quality AI art images), which would total 100 participants after exclusions. We pre-registered exclusions similar to Nosek and Banaji (2001), such as excluding RT $\leq$ 300 ms trials as they might not reflect a genuine response but an accidental key-press. To remove the possibility that participants ignore the task and instead make random keypresses, we also pre-registered excluding participants who have accuracy rates for hits (<55 %) or false alarms (>45 %) close to chance performance (50 %).

Participants were randomly recruited to one of the two experimental groups. Fifty participants (25 females, 25 males, $Mean_{age}$ = 35.72, SD_{age} = 8.26) were recruited for the high-quality AI art images group. An additional fifty participants (32 females, 18 males, $Mean_{age}$ = 36.28, SD_{age} = 8.50) were recruited for the low-quality AI art images group. Based on RT criteria we excluded 16,617 trials (out of 32,000 trials in total). Also, according to low accuracy hits criteria, we excluded one participant from the low-quality images group. After exclusions, the final sample for the low-quality AI images group included forty-nine participants (31 females, 18 males, $Mean_{age}$ = 36.57, SD_{age} = 8.33).

4.2.3. Stimuli, design, tasks and procedure

Stimuli. The visual images were either target or distracter. The target stimuli for Impressionist AI Art consisted of 16 AI-generated images of high-artistic quality (8 depicting people, 8 landscapes) and 16 of low-artistic quality (8 depicting people, 8 landscapes; see Fig. 3 for some examples). The high-quality images were the same AI art images used in Experiment 1. All Impressionist AI Art images were generated using OpenAI DALL-E3 with textual prompts based on Sorolla's Impressionist style. For human-made Impressionist Art target stimuli, we used 16 human-made art images by the Impressionist Spanish artist Joaquín Sorolla (8 people, 8 landscapes). The Impressionist art stimuli were taken from previously validated stimuli dataset (Bara et al., 2023, 2024). The stimuli are available here (https://osf.io/d4zme/).

When generating Impressionist AI Art images based on Sorolla's art prompts, the initial outputs from the DALL-E3 algorithm often showed low artistic quality compared to subsequent iterations. These early images were categorised as low artistic quality group, while the later, more refined versions mimicking closely Sorolla's style, represented the high artistic quality group. For more information, please see Supplementary Material (Exp.3, Section A) An example of AI art low and high artistic quality, as well as Sorolla's work, can be seen in Fig. 8.

Following a similar approach to Nosek and Banaji (2001), the distracter stimuli were items from a superordinate category than the target. For example, when the target was Impressionist Art, the distracters were Art Photographs depicting nature and trains. Similarly, for Impressionist AI Art, the distracters were AI-generated Art Photographs of nature and trains, created using OpenAI's DALL-E3. We reasoned that art photography is a more general art category than Impressionist Sorolla's art paintings because it encompasses a broader range of artistic styles and subjects. While Impressionist Sorolla's paintings represent a specific historical period, style, and technique in visual art, art photography includes a diverse range of genres, techniques, and themes, making it more general. All photographic distracters had won photography awards at various art competitions, and by doing so, they met the photographic and artistic standards required by these art competitions. Due to copyright restrictions, we cannot directly provide the images. However, web-links to each image are available in the supplementary materials (Table S8).

The attributes stimuli for "Good" were 16 words. Some examples include words, such as beautiful, celebrating or excellent. The attributes stimuli for "Bad" were 16 words, such as disaster, disgusting or dislike. Both "Good" and "Bad" word attributes stimuli were taken from the original GNAT (Nosek & Banaji, 2001).

Design. Experiment 3 used a mixed within- and between-participant design. Participants completed the modified GNAT test in a between-participant manipulation (2 x AI art type: AI art low-quality vs AI art high-quality). Participants were randomly assigned to one of two AI quality art type groups. At the end of GNAT test, all participants filled in the same short questionnaires as in Experiments 1 and 2. Sensitivity (d-prime) was our main dependent variable and calculated by combining accuracy data rates for each pairing (e.g., Impressionist Art + "Good"), following the same algorithm as Nosek and Banaji (2001). D-prime values of 0 or below suggest that participants either could not differentiate any signal from noise or did not adhere to the task instructions.

Tasks and Procedure. The modified GNAT evaluated the strength of association between a target category (e.g., Impressionist Art, Impressionist AI Art) and two moral poles of an attribute dimension ("Good", "Bad"). The GNAT involved a training phase and a test phase. In the training phase, participants categorised attributes (e.g.," bad", "good"

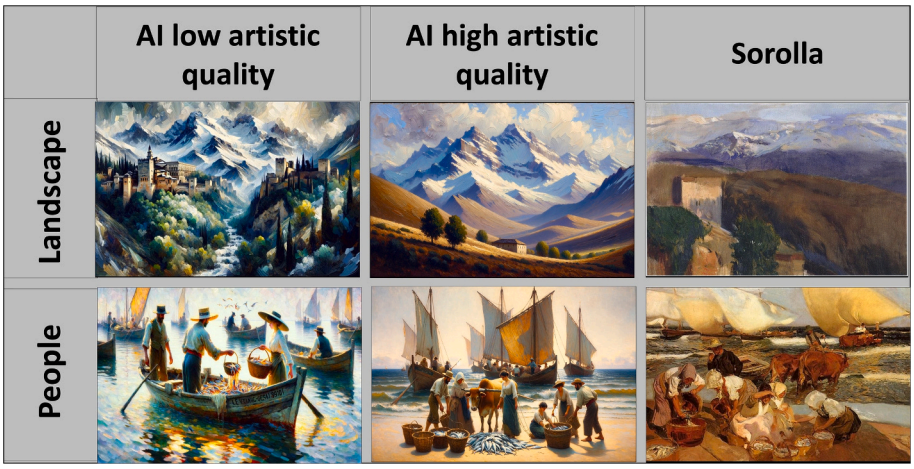

Fig. 8. A representation of the different stimuli used in Experiment 2. The columns show examples of low-quality AI art, high-quality AI art, and their corresponding Sorolla's artwork. The rows show landscapes and people.

words) and target items (e.g., Impressionist Art, Impressionist AI Art) into predetermined categories via keystroke presses. The basic task was to press the Spacebar as quickly and as accurately as possible if an item (e.g., "excellent") belonged to the category being tested (e.g., "Good") and to do nothing if it did not. The training phase included four training blocks (Impressionist Art, Impressionist AI Art, "Good", "Bad") of 20 trials each with 1000 ms response time, therefore, 80 trials in total. The block order was determined randomly.

The test phase combined paired targets and attributes (e.g., "Impressionist Art or Good", "Impressionist Art or Bad", "Impressionist AI Art or Good", "Impressionist AI Art or Bad"). When an item belonged to either one of these two categories, participants were instructed to press the Spacebar as quickly and as accurately as possible and do nothing if they did not. There were 4 test blocks in total with 750 ms response time. The block order was determined randomly. Each block included 16 'practice trials' followed by 80 test trials. As we followed a similar approach to Nosek and Banaji (2001), only the test phase blocks were included in the analysis. The distracters (noise) were items from a superordinate category relative to the target. For example, when Impressionist Art was the target, the distracters were art photography items depicting nature and trains. Furthermore, when Impressionist AI Art was the target, the distracter was AI art photography items depicting nature and trains. The distracters for word attributes were represented by the alternate attribute (e.g., when good-related words are signal, bad-related words are distracters).

The GNAT was produced in PsychoPy (v2024.1.3, Peirce et al., 2019) and ran online using Pavlovia (https://pavlovia.org/). An example of GNAT experimental conditions is provided in Fig. 9.

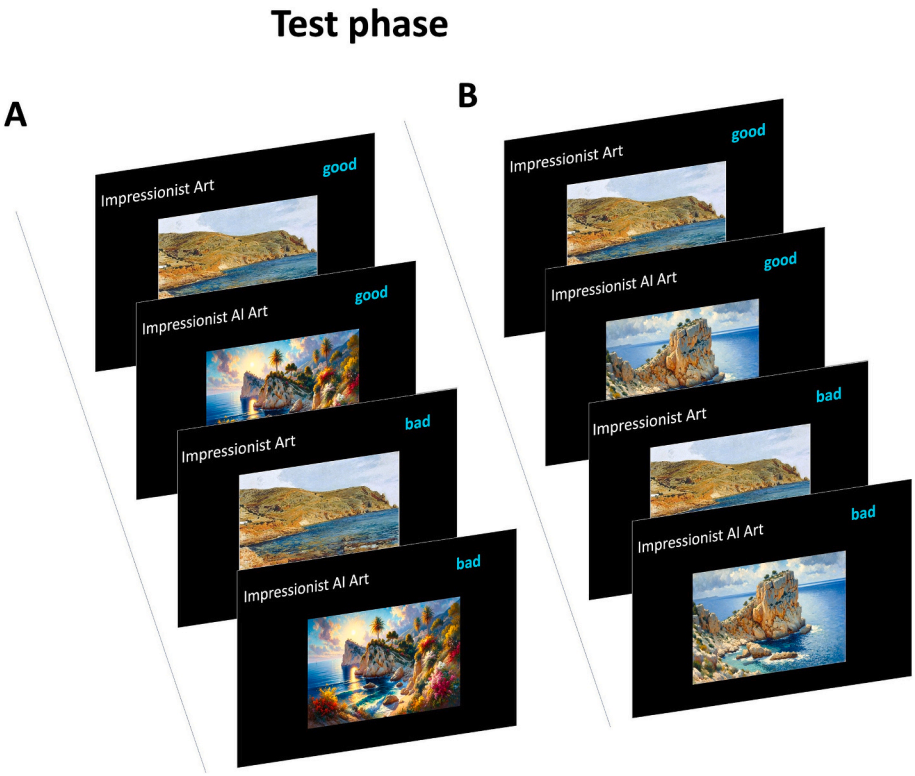

Fig. 9. An example of GNAT test phase by AI art quality group. Panel A shows the target stimuli in the low-quality Impressionist AI Art, whereas panel B shows the high-quality Impressionist AI Art group.

Following the GNAT task, all participants completed a series of questionnaires. We report these exploratory results in the Supplementary material (Exp.3, Section C, Figs. S32-S37).

4.2.4. Data analyses

As in Experiments 1 and 2, we preregistered a similar multilevel Bayesian analytical approach. For more information, please see the Supplementary Material (Exp. 3, Section A). We addressed our research questions using the following model formula:

```
brm(d-prime ~ 1 + target * attribute * group +
(1 + | participant))
```

Note: target = Impressionist Art, Impressionist AI Art; attribute = good, bad words; group = low-quality Impressionist AI Art, high-quality Impressionist AI Art.

We report weakly informative priors in the Supplementary material (Exp.3, Section A, Table S6).

4.3. Results

Summary d-prime data by target, attribute and group are shown below (Fig. 10).

The posterior distribution for the full model is shown in Fig. 11 and Tables S7 (Supplementary Material). While we visualise parameter estimates from the full model, we discuss only discuss the pre-registered parameters of interest that address our key hypotheses.

Regarding our pre-registered interactions, we first discuss the two-way interactions between target and attribute. The posterior distribution revealed similar sensitivity to both "good" and "bad" attributes for Impressionist Art and a mirrored pattern for Impressionist AI Art, as both the 95 % and 66 % quantile intervals overlapped with zero. This indicates a lack of implicit associations between Impressionist Art and positive words and between Impressionist AI Art and negative words. Furthermore, regarding the three-way interactions between group, target, attribute, the posterior distribution showed largely overlapping

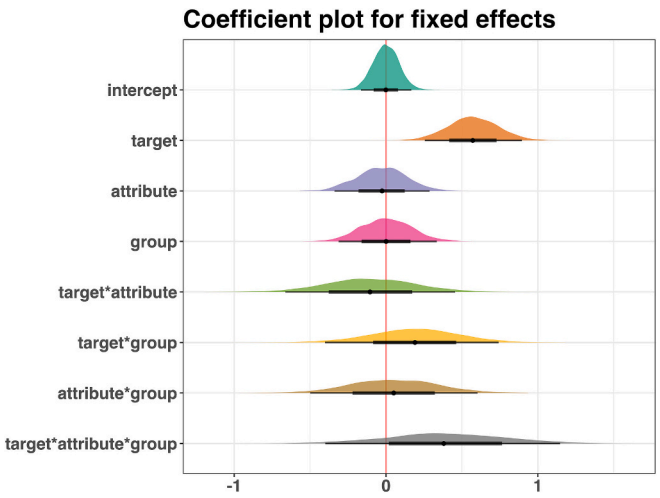

Fig. 11. Coefficient plot for the full model. It illustrates the posterior distribution across group (low AI art quality vs. high AI art quality), target (Impressionist AI Art vs. Impressionist Art), and attribute.("bad" vs. "good" words). The coloured half-eye plots indicate the posterior distribution for each response category. The width and height of these shapes show the probability density of the responses at each level. The point estimate (black dots) represents the median of the posterior predictions for each category. Error bars represent 66 % quantile intervals (thick black lines) and 95 % quantile intervals (thin black lines).

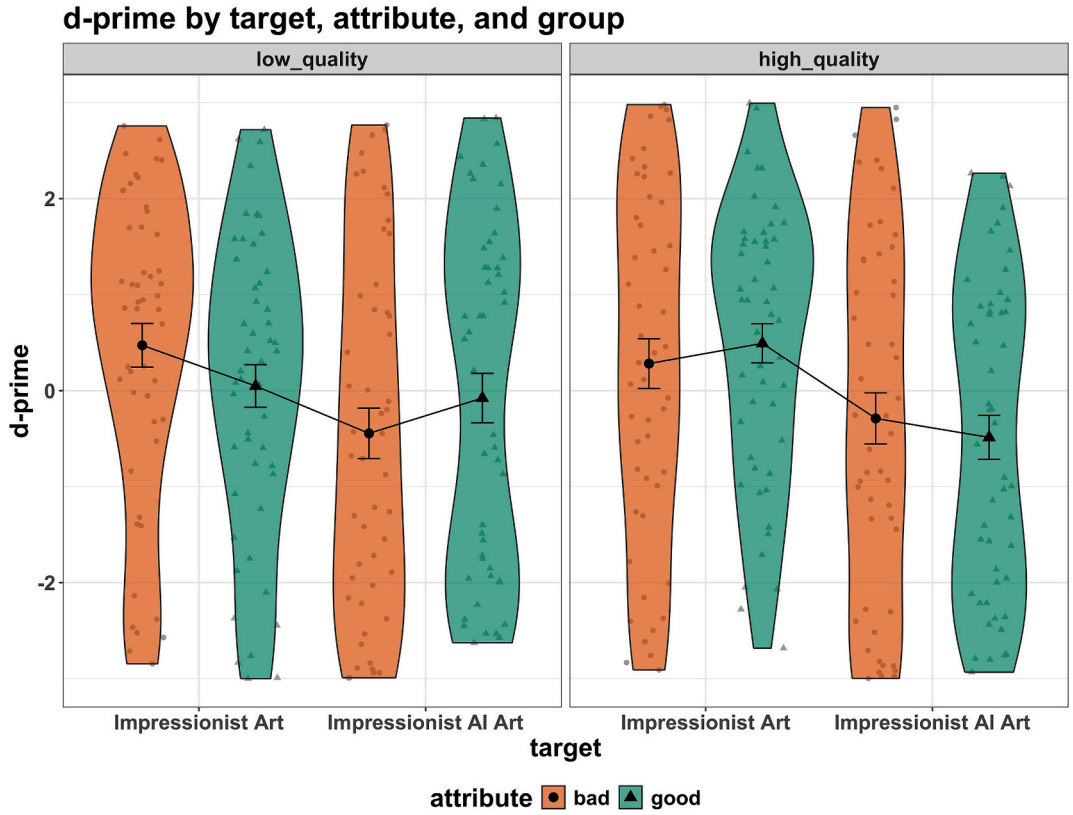

Fig. 10. d-prime summary data by target (Impressionist Art vs. Impressionist AI Art), group (low-quality vs high-quality), attribute. The panels show the target and group type, whereas the attribute is "bad" (orange), "good" (green). Error bars represent 95 % confidence intervals. The black markers (circles and triangles) and interval estimates represent the group mean average, whereas the grey markers represent the individual participants.

distributions, suggesting reduced sensitivity to our key interaction terms. In addition, while there is a trend toward higher sensitivity to high-quality Impressionist Art and positive attributes compared to low-quality and negative attributes (66 % quantile intervals exclude zero), respectively, the large overlap of the 95 % quantile interval distributions, suggests a reduced effect of the interaction terms.

And finally, an exploratory finding revealed a positive main effect of the target (Impressionist Art > Impressionist AI Art) regardless of group and attribute. This suggests that Impressionist Art as a target was detected more accurately than Impressionist AI Art.

4.4. Discussion

In Experiment 3 we found no evidence for implicit associations between Impressionist Art and notions of "good" nor between Impressionist AI Art and notions of "bad". In contrast to our main hypothesis, this suggests that observers do not automatically associate Impressionist Art with positive attributes, nor do they associate AI-generated Impressionist Art with negative attributes. While a tendency emerged for higher sensitivity toward high-quality Impressionist Art and positive attributes, and lower sensitivity toward low-quality and negative attributes, this distinction was relatively weak. Overall, our results suggest that recognition of Impressionist Art is more effective than Impressionist AI art, while the perceptions of quality and attribute associations show similar sensitivity, rather than clear-cut distinctions. This finding highlights the complex nature of implicit associations, particularly when comparing human art with different quality levels of AI-generated art.

5. General discussion

Across three pre-registered experiments using both explicit and implicit paradigms and Bayesian modelling, our findings offer new insights into the moral and aesthetic implications of AI-generated art. Overall, our findings suggest that: (1) factual information about AI systems reduces moral acceptability and aesthetic judgments, regardless of the content depicted; (2) moral acceptability of AI-generated art remains largely unaffected by different information doses about its success beyond the factual information about AI systems; and (3) implicit associations between human art and AI-generated art and their perceived attributes are similar. The implications of these findings are discussed below.

5.1. Implications for contextual information processing theories

Our findings extend prior research on the role of contextual information in shaping moral and aesthetic judgments. Previous studies have shown that moral judgments are impacted by contextual information, such as different social scenarios or word framings (Earp, McLoughlin, Monrad, Clark, & Crockett, 2021; Simpson, Laham, & Fiske, 2016). This supports the theory that moral judgments are flexibly updated and context-dependent (Andrejević et al., 2020; Bartels et al., 2015; Schein, 2020; Simpson, 2017; Slothuus, 2008). Our work shows that information about AI-backend operations is critical when evaluating the moral acceptability of AI-generated art. Our findings are therefore important for advancing future discussions about the ethical implications and societal acceptance of AI-generated art. Our results suggest that moral acceptability judgments of AI-generated art are grounded in foundational knowledge about AI systems, with success-related framing having minimal impact. This is consistent with research showing that relevant information shapes moral understanding (Andrejević et al., 2020; Mann & Ferguson, 2015). However, unlike findings from social psychology, where success framing often results in positive bias (Cone et al., 2017; Kahneman & Tversky, 1984), success-related information about AI-generated artworks appears peripheral in moral judgment contexts. This limited influence may indicate a deeper skepticism regarding AI's artistic capabilities (Bellaiche et al., 2023), as framing AI art as successful could highlight concerns about potential deception, and the undermining of traditional artistic values and practices. These findings overall emphasize the critical role of foundational knowledge in shaping moral evaluations, suggesting that moral judgments of AI systems prioritise understanding over contextually positive framing.

Moreover, these findings complement prior work on the impact of contextual information on the aesthetic appeal of AI-generated art. While existing research has focused on how knowing about an artwork's artificial origins affects aesthetic appreciation (Chamberlain et al., 2018; Di Dio et al., 2023; Hong et al., 2021), we show that technical details about AI operations can further diminish the aesthetic appeal. This is particularly important given that the general public often lacks an understanding of AI processes (Edelman AI Center of Expertise, 2019). Our study also highlights the need for greater transparency and AI literacy to foster more informed and realistic public attitudes, paving the way for future research on AI education and art production.

5.2. Implications for moral foundations theory

Our findings revealed important moral concerns regarding AI-generated art, particularly on financial gain, prestige and its art status, when AI backend operations are known. While it remains for future research to disentangle the relationship between AI-generated art and Moral Foundations Theory (MFT; Haidt & Graham, 2007), it could be that AI-generated art may violate moral perceptions about just and fair financial practices. Research has indicated that unethical consumer behaviour, such as profiting from questionable actions, is linked to fairness and sanctity moral foundations (Chowdhury, 2019). Considering possible perceived minimal creative effort in AI-generated art and the lack of compensation for human artists whose artworks contribute to AI databases, obtaining financial incentives from AI-generated art may challenge moral standards of equity, justice, and respect for societal and legal norms.

Furthermore, gaining prestige from exhibiting AI-generated art in traditional art museums can be linked to ethical values of social and cultural acclaim, based on authentic personal merit and professional skill (Cheng, 2020). Empirical evidence has shown that prestige strongly predicts moral foundations of care, fairness, loyalty, and sanctity (Khanipour, Pourali, & Atar, 2021), suggesting that ethical means are essential for attaining cultural prestige. Since AI-generated art is derived from databases of human-made artworks gaining public acclaim could be linked to violating moral values of fairness and justice.

Similarly, AI-generated art may violate moral perceptions concerning its status as a legitimate form of art. For example, labelling AI-generated art as human-made art may conflict with the moral pillars of loyalty, authority, and purity (Graham et al., 2013; Haidt & Graham, 2007; Simpson, 2017; Zakharin & Bates, 2023). Moral loyalty involves respect to cultural or artistic traditions and AI-generated art challenges conventional notions of authorship and the admiration people may feel toward human art. Questions of moral authority regarding AI-generated art may relate to who holds the prerogative to define art, as AI art dissents from traditional norms and institutions that have defined creativity throughout history. Furthermore, concerns about artistic purity may stem from doubts over the genuineness and integrity of AI-generated art, particularly when it is perceived as lacking authentic artistic intent or emotional depth. Overall, our findings can inform future investigations involving ethical concerns for fairness and justice in the AI-generated art market. By doing so, future research might provide a basis for guiding policies, and practices that uphold ethical standards in the evolving landscape of AI-generated art and creative industries.

5.3. Implications for implicit measures

Our findings do not support the hypothesis that human-created Impressionist Art is implicitly associated with positive attributes,

while AI-generated Impressionist Art is associated with negative attributes, particularly in the low-quality AI art category. Since our GNAT task used general "good" and "bad" attribute words as Nosek and Banaji's (2001), future research may examine the moral sensitivity to AI-generated art using attribute words that better reflect current AI's moral concerns, such as fair vs. dishonest or authentic vs. fake. For example, (Lakritz et al., 2022) used a more specialised GNAT approach to investigate the moralisation of food and found that specific food categories were implicitly tied to distinct moral dimensions (e.g., purity vs. impurity). This suggests that tailoring positive and negative attributes to address specific questions can provide more precise insights.

One interesting trend in our findings indicated that participants were better at detecting and distinguishing Impressionist Art compared to Impressionist AI Art, irrespective of the group quality or attribute type. This is intriguing given that most existing studies have reported low accuracy rates in detecting AI-generated art compared to human art (Chamberlain et al., 2018; Darda & Cross, 2022; Gangadharbatla, 2021; Samo & Highhouse, 2023). While we acknowledge differences in methodologies, future research might build upon our findings, as they may represent a more up-to-date perception of AI detection due to the growing everyday exposure to AI-generated creative outputs. Importantly, our use of implicit measures, such as the GNAT, provides a valuable complement to self-reported judgments by capturing automatic, unconscious associations and less susceptible to social desirability biases. In the context of AI-generated art, as exposure to AI in creative fields continues to grow, the GNAT's ability to track subtle, evolving changes in implicit associations becomes increasingly valuable, deepening our understanding of the changing dynamics of human-AI interactions in art and their broader societal implications.

5.4. Constraints on generality

Across all three experiments, we used Impressionist Art as the main reference stimuli category. It therefore remains for future studies to explore of role of different artistic styles on moral and aesthetic judgments. In addition, our sample was predominantly UK-based, which requires caution when generalising these results to other cultural or geographic settings. Since moral judgments are sensitive to normative moral beliefs updating in response to new information (Hogarth & Einhorn, 1992; Monroe & Malle, 2019), future work might consider moral updating judgments in the context of AI-generated art. That approach has also the potential to account for possible order effects by tracking how individuals dynamically adjust and update their moral reasoning over time in response to evolving information. Finally, our work reflects the current moral and aesthetic perceptions of AI-generated art. By contrasting different qualities of AI-generated art, our work highlights its present real-world relevance, acknowledging that viewers are likely to experience a variety of AI-generated artistic outputs in everyday settings – from impossible artistic renderings to closely resembling human-made art. At the same time, we recognise that AI-generated art is an evolving domain, and these moral and aesthetic attitudes are likely to change over the years. It is, therefore, essential to continuously update our understanding of the dynamic interplay between technology, morality, and aesthetics.

CRediT authorship contribution statement

Ionela Bara: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Richard Ramsey:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Emily S. Cross:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Acknowledgments

We thank Dr. Andrew Wildman for his help with the PsychoPy task, and Hannah Megan Curwen and Nea Mazouz for their assistance with AI-generated art.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106063.

Data availability

All data is availble on the open science framework (https://osf.io/d4zme/).

References

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review, 55*(1), 589–656. https://doi.org/10.1007/s10462-021-10039-7

Andrejević, M., Feuerriegel, D., Turner, W., Laham, S., & Bode, S. (2020). Moral judgements of fairness-related actions are flexibly updated to account for contextual information. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-74975-0

Ashford, R. D., Brown, A. M., & Curtis, B. (2019). The language of substance use and recovery: Novel use of the go/no–go association task to measure implicit Bias. *Health Communication, 34*(11), 1296. https://doi.org/10.1080/10410236.2018.1481709

Bara, I., Cross, E. S., & Ramsey, R. (2023). *The Role of Art Knowledge Training on Aesthetic Judgements and Executive Functions*. preprint. https://doi.org/10.31234/osf.io/pvxzy.

Bara, I., Darda, K. M., Kurz, A. S., & Ramsey, R. (2021). Functional specificity and neural integration in the aesthetic appreciation of artworks with implied motion. *European Journal of Neuroscience, 54*(9), 7231–7259. https://doi.org/10.1111/EJN.15479

Bara, I., Ramsey, R., & Cross, E. S. (2024). *Art knowledge training shapes understanding, inspires creativity and stimulates thinking*. preprint. https://doi.org/10.31234/osf.io/sxqgc.

Barasch, M. (1990). *Modern theories of art: From impressionism to Kandinsky* (Vol. 2). NYU Press.

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. In , *478–515. The Wiley Blackwell Handbook of Judgment and Decision Making*. https://doi.org/10.1002/9781118468333.ch17

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications, 8*(1), 42.

Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology, 107*(1), 154–178.

Bonnefon, J.-F., Rahwan, I., & Shariff, A. (2024). The moral psychology of artificial intelligence. *Annual Review of Psychology, 75*(1), 653–675. https://doi.org/10.1146/annurev-psych-030123

Bornstein, R. F., & Pittman, T. S. (Eds.). (1992). *Perception without awareness: Cognitive, clinical, and social perspectives*. Guilford Press.

Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal, 10*(1), 395–411. https://doi.org/10.32614/rj-2018-017

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40*(1), 202–223. https://doi.org/10.1111/cogs.12231

Cameron, C. D., Payne, B. K., Sinnott-Armstrong, W., Scheffer, J. A., & Inzlicht, M. (2017). Implicit moral evaluations: A multinomial modeling approach. *Cognition, 158*, 224–241.

Carroll, N. (2000). *Theories-of art today*. The University of Wisconsin Press.

Cha, Y., Baek, S., Ahn, G., Lee, H., Lee, B., Shin, J., & eun, & Jang, D.. (2020). Compensating for the loss of human distinctiveness: The use of social creativity under Human–Machine comparisons. *Computers in Human Behavior, 103*, 80–90. https://doi.org/10.1016/j.chb.2019.08.027

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. https://doi.org/10.1037/aca0000136.supp

Chatterjee, A. (2022). Art in an age of artificial intelligence. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1024449

Cheng, J. T. (2020). Dominance, prestige, and the role of leveling in human social hierarchy and equality. *Current Opinion in Psychology, 33*, 238–244. https://doi.org/10.1016/j.copsyc.2019.10.004

Chowdhury, R. M. (2019). The moral foundations of consumer ethics. *Journal of Business Ethics, 158*(3), 585–601. https://doi.org/10.1007/s10551-017-3676-2

Collingwood, R. (1958). *The principles of art* (Vol. 11). Oxford University Press.

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology, 56*, 131–199. https://doi.org/10.1016/bs.aesp.2017.03.001

Coombe, R. J. (1998). *The cultural life of intellectual properties: Authorship, appropriation, and the law*. Duke University Press.

Danto Arthur, C. (2013). *What art is*. Yale University Press.

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

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

Davies, S. (2012). *The artful species: Aesthetics, art, and evolution*. OUP Oxford.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers, & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 11–28). Sage Publications, Inc.. https://doi.org/10.4135/9781412976237.n2

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347.

Di Dio, C., Ardizzi, M., Schieppati, S. V., Massaro, D., Gilli, G., Gallese, V., & Marchetti, A. (2023). Art made by artificial intelligence: The effect of authorship on aesthetic judgments. *Psychology of Aesthetics, Creativity, and the Arts.*. https://doi.org/10.1037/aca0000602

Dissanayake, E. (1995). *Homo aestheticus : Where art comes from and why*. University of Washington Press.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one, 18*(3), Article e0279720.

Dunn, S. (2020). Identity manipulation: Responding to advances in artificial intelligence and robotics. In *Suzie Dunn, "Identity Manipulation: Responding to Advances in Artificial Intelligence and Robotics"(2020) WeRobot, 2020, Conference Paper*.

Earp, B. D., McLoughlin, K. L., Monrad, J. T., Clark, M. S., & Crockett, M. J. (2021). How social relationships shape moral wrongness judgments. *Nature Communications, 12*(1). https://doi.org/10.1038/s41467-021-26067-4

Edelman AI Center of Expertise. (2019). *2019 Edelman Artificial Intelligence Survey*.

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review, 23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Epstein, Z., Levine, S., Rand, D. G., & Rahwan, I. (2020). Who gets credit for AI-generated art? *IScience, 23*(9), Article 101515. https://doi.org/10.1016/J.ISCI.2020.101515

Frith, C., & Frith, U. (2023). *What makes us social? 16 communicating and sharing meaning. What makes us social?* MIT Press. https://doi.org/10.7551/mitpress/10400.001.0001

Gangadharbatla, H. (2021). The role of AI attribution knowledge in the evaluation of artwork. *40*(2), 125–142. https://doi.org/10.1177/0276237421994697

Gaut, B., & McIver Lopes, D. (2001). *The Routledge companion to aesthetics*. London: Routledge.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*, 515–534. https://doi.org/10.1214/06-BA117A

Goldberg, B. (2018). First-ever auction of AI-created artwork set for Christie's gavel. *Reuters*. Retrieved from https://www.reuters.com/article/us-france-art-artificial-intelligence/first-ever-auction-of-ai-created-artwork-set-for-christies-gavel-idUSKCN1MX2WO.

Goldman, A. (2001). In B. Gaut, & D. McIver Lopes (Eds.), *The aesthetic (in the Routledge companion to aesthetic)*. London: Routledge.

Graham, G. (2005). *Philosophy of the arts: An introduction to aesthetics* (3rd ed.). Routledge.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Morality takes two: Dyadic morality and mind perception. In M. Mikulincer, & P. R. Shaver (Eds.), *(Eds.), the social psychology of morality: Exploring the causes of good and evil*. American Psychological Association.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364–371. https://doi.org/10.1016/J.COGNITION.2009.02.001

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464.

Guglielmo, S. (2015). Moral judgment as information processing: An integrative review. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01637

Gunkel, D. J. (2017). Rethinking art and aesthetics in the age of creative machines: Editor's introduction. *Philosophy and Technology, 30*(3), 263–265. https://doi.org/10.1007/s13347-017-0281-3

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hitti, N. (2018). *Christie's sells AI-created artwork painted using algorithm for $432,000*.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*(1), 1–55.

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

Hong, J. W., Peng, Q., & Williams, D. (2021). Are you ready for artificial Mozart and Skrillex? An experiment testing expectancy violation theory and AI music. *New Media and Society, 23*(7), 1920–1935. https://doi.org/10.1177/1461444820925798

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341.

Kay, M. (2020). *Tidybayes: Tidy data and Geoms for Bayesian models. R package version 2.3.0*. https://doi.org/10.5281/zenodo.1308151

Khanipour, H., Pourali, M., & Atar, M. (2021). Prestige and dominance as differential correlates of moral foundations and its clinical implications. *Practice in Clinical Psychology, 9*(1), 1–8. https://doi.org/10.32598/jpcp.9.1.63.7

Kuklinski, J. H., Quirk, P. J., Schwieder, D. W., & Rich, R. F. (1998). "Just the facts, ma'am": Political facts and public opinion. *The Annals of the American Academy of Political and Social Science, 560*(1), 143–154.

Kurz, A. S. (2020). *Statistical rethinking with brms, ggplot2, and the tidyverse (second)*. https://doi.org/10.5281/zenodo.4302550

Ladak, A., Loughnan, S., & Wilks, M. (2023). The moral psychology of artificial intelligence. *Current Directions in Psychological Science., 33*(1), 27–34. https://doi.org/10.1177/09637214231205866

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1). https://doi.org/10.1525/COLLABRA.33267

Lakritz, C., Tournayre, L., Ouellet, M., Iceta, S., Duriez, P., Masetti, V., & Lafraire, J. (2022). Sinful foods: Measuring implicit associations between food categories and moral attributes in anorexic, Orthorexic, and healthy subjects. *Frontiers in Nutrition, 9*. https://doi.org/10.3389/fnut.2022.884003

Latikka, R., Bergdahl, J., Savela, N., & Oksanen, A. (2023). AI as an artist? A two-wave survey study on attitudes toward using artificial intelligence in art. *Poetics, 101*. https://doi.org/10.1016/j.poetic.2023.101839

Lima, G., Zhunis, A., Manovich, L., & Cha, M. (2021). On the social-relational moral standing of AI: An empirical study using AI-generated art. *Frontiers in Robotics and AI*, 8. https://doi.org/10.3389/frobt.2021.719944

Loughnan, S., & Haslam, N. (2007). Animals and Androids: Implicit Associations between Social Categories and Nonhumans. *18*(2), 116–121.

Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology, 72*, 293–318. https://doi.org/10.1146/annurev-psych-072220

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*(2), 101–121.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108*(6), 823.

McElreath, R. (2020). *Statistical rethinking; A Bayesian Course with Examples in R and Stan* (2nd ed.). https://doi.org/10.4324/9780429029608

Midjourney. (2023). Midjourney (V5) Text-to-image model. https://www.midjourney.com/.

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior, 143*, Article 107707. https://doi.org/10.1016/J.CHB.2023.107707

Moffat, D., & Kelly, M. (2006). An investigation into people's bias against computational creativity in music composition. In *Paper Presentedat the third joint workshop on computational creativity (ECAI'06), Riva Del Garda, Italy*.

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology, 116*(2), 215.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Du Sert, P., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*. https://doi.org/10.1038/s41562-016-0021

Muszyński, M. (2023). Attention checks and how to use them: Review and practical recommendations. Ask. *Research & Methods, 32*. https://doi.org/10.18061/ask.v32i1.0001

Nelson, T. E., Oxley, Z. M., & Clawson, R. A. (1997). Toward a psychology of framing effects. *Political Behavior, 19*, 221–246.

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

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625–666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166–180.

OpenAI. (2024). ChatGPT. *Large language model*. Retrieved from https://chat.openai.com/chat.

Palumbo, L., Ruta, N., & Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PLoS One, 10*(10), Article e0140043.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*(3), 145–171.

Pirutinsky, S., Carp, S., & Rosmarin, D. H. (2017). A paradigm to assess implicit attitudes towards god: The positive/negative god associations task. *Journal of Religion and Health, 56*(1), 305–319. https://doi.org/10.1007/S10943-016-0303-Y/TABLES/4

Ramachandran, V. S., & Hirstein, W. (1999). The science of art A neurological theory of aesthetic experience. *Journal of Consciousness Studies, 6*(7), 15–51.
RCore Team. (2024). *R Core team. R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Retrieved from http://arxiv.org/abs/2112.10752.
Roose, K. (2022). *An A.I.-generated picture won an art prize. Artists aren't happy*. The New York Times.
Rossion, B. (2022). What makes us human? Face identity recognition. In *In The Routledge handbook of semiosis and the brain* (pp. 325–345). https://doi.org/10.4324/9781003051817-26
Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and open-AI models: A preliminary review. *Future Internet, 15*(6), 192. https://doi.org/10.3390/FI15060192
Samo, A., & Highhouse, S. (2023). Artificial intelligence and art: Identifying the aesthetic judgment factors that distinguish human and machine-generated artwork. *Psychology of Aesthetics, Creativity, and the Arts.*. https://doi.org/10.1037/aca0000570
Sawyer, R. K. (2012). *The science of human innovation: Explaining creativity. New York, NY*.
Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science, 15*(2), 207–215. https://doi.org/10.1177/1745691620904083
Schmitt, B. (2020). Speciesism: An obstacle to AI and robot adoption. *Marketing Letters, 31*(1), 3–6. https://doi.org/10.1007/s11002-019-09499-3
Shahriar, S. (2022). GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. *Displays, 73*, Article 102237. https://doi.org/10.1016/J.DISPLA.2022.102237
Shao, Y., Zhang, C., Zhou, J., Gu, T., & Yuan, Y. (2019). How does culture shape creativity? A mini-review. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.01219
Simpson, A. (2017). Moral foundations theory: Background, review, and scaffolding for future research. *Encyclopedia of Personality and Individual Differences*, 1–19.
Simpson, A., Laham, S. M., & Fiske, A. P. (2016). Wrongness in different relationships: Relational context effects on moral judgment. *Journal of Social Psychology, 156*(6), 594–609. https://doi.org/10.1080/00224545.2016.1140118
Slothuus, R. (2008). More than weighting cognitive importance: A dual-process model of issue framing effects. *Political Psychology, 29*(1), 1–28. https://doi.org/10.1111/j.1467-9221.2007.00610.x
Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin, 39*(2), 193–205.
Tepe, B., & Byrne, R. M. J. (2022). Cognitive processes in imaginative moral shifts: How judgments of morally unacceptable actions change. *Memory and Cognition, 50*(5), 1103–1123. https://doi.org/10.3758/s13421-022-01315-0
Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., … Le Google, Q. (2022). LaMDA: Language models for dialog applications. arXiv preprint. arXiv:2201.08239. Retrieved from https://arxiv.org/abs/2201.08239.
Tubadji, A., Huang, H., & Webber, D. J. (2021). Cultural proximity bias in AI-acceptability: The importance of being human. *Technological Forecasting and Social Change, 173*. https://doi.org/10.1016/j.techfore.2021.121100
Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines, 29*(3), 417–440. https://doi.org/10.1007/s11023-019-09506-6
Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*.
Yoder, K. J., Harenski, C., Kiehl, K. A., & Decety, J. (2015). Neural networks underlying implicit and explicit moral evaluations in psychopathy. *Translational Psychiatry, 5*(8), e625.
Young, J. O. (2001). *Art and knowledge*. Routledge.
Young, J. O. (2010). *Cultural appropriation and the arts*. John Wiley & Sons.
Zakharin, M., & Bates, T. C. (2023). Moral foundations theory: Validation and replication of the MFQ-2. *Personality and Individual Differences, 214*. https://doi.org/10.1016/j.paid.2023.112339