



HAL
open science

AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?

Martin Ragot, Nicolas Martin, Salomé Cojean

► **To cite this version:**

Martin Ragot, Nicolas Martin, Salomé Cojean. AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?. CHI '20: CHI Conference on Human Factors in Computing Systems, Apr 2020, Honolulu, United States. 10.1145/3334480.3382892 . hal-03641111

HAL Id: hal-03641111

<https://hal.science/hal-03641111v1>

Submitted on 20 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?

Martin Ragot^a, Nicolas Martin^a, & Salomé Cojean^b

^aIRT b<>com
1219, Av des Champs Blancs
35510 Cesson-Sévigné, France
martin.ragot@b-com.com, nicolas.martin@b-com.com

^bLaboratory of Psychology LPPL (EA 4638)
University of Angers
11, boulevard Lavoisier
Angers, 49045, France
salome.cojean@univ-grenoble-alpes.fr
DOI: <https://doi.org/10.1145/3334480.3382892>

Abstract

Via generative adversarial networks (GANs), artificial intelligence (AI) has influenced many areas, especially the artistic field, as symbol of a human task. In human-computer interaction (HCI) studies, perception biases against AI, machines, or computers are generally cited. However, experimental evidence is still lacking. This paper presents a wide-scale experiment in which 565 participants are asked to evaluate paintings (which were created by humans or AI) on four dimensions: liking, perceived beauty, novelty, and meaning. A priming effect is evaluated using two between-subject conditions: artworks presented as created by an AI, and artworks presented as created by a human artist. Finally, the paintings perceived as being drawn by human are evaluated significantly more highly than those perceived as being made by AI. Thus, using such a methodology and sample in an unprecedented way, the results show a negative bias of perception towards AI and a preference bias towards human systems.

Author Keywords

Arts; Artificial Intelligence; AI; Authorship; Bias; CAN; Computational Creativity; GAN; Painting.

Introduction

For many years, technological progress has wrought profound changes in the creative industry, as in the fields of music (e.g., [18, 19]), parametric design (e.g., [41]), generative fashion design (e.g., [34]), and paintings (e.g., [23]). This has given rise to the emergence of new research fields: computational aesthetics [31], attractiveness computing (see Chu [14] for a literature review), Creative Adversarial Networks (CAN) [23] and computational creativity [16]. Computational creativity is a generic field of artificial intelligence (AI) that studies both artificial and human creativity [56] and has grown considerably. Indeed, in recent years, AI and especially its subfield machine learning (ML) have gained popularity [38] thanks to

research breakthroughs [37]. In 2014, Goodfellow and collaborators [27] proposed a new ML paradigm to generate synthetic data called a generative adversarial network (GAN)¹. Artists have appropriated these models have been in recent years to the point that they have become a new art market [20]. Thus, a first art painting generated by a GAN was sold for \$432,500 by Obvious [14]². With the increase in quality and ubiquity of creative systems such as GANs, studying these creativity support systems seems to be necessary [17, 29]. Evaluating the relevance of recently produced artistic works consists of assessing the perception of the public towards these artworks. To improve the algorithms, one approach is to try to understand user perceptions. There is also a need to understand the acceptance of works produced by AI [16]. Currently, this perception seems to be rather negative. Indeed, Gaut [26] has postulated a “negative value” towards machines or computers. Colton [15] is also concerned about the rejection of creative computers. Nevertheless, the evidence for such negative bias remains unclear. It seems crucial to study human perceptions of these AI-generated artworks. For this purpose, different methodologies exist. First, in many studies, a modified Turing test (TT) is used to analyze humans’ capacity to distinguish human or AI artworks (e.g., [23, 45, 51]). However, the TT seems insufficient. Indeed, this type of test does not make it possible to study audience perceptions. This method has also been criticized for encouraging imitation and pastiches of existing human artworks [46]. In parallel, to the best of our knowledge, few researchers have studied the differences in public perceptions between AI-generated or human artworks to understand the potential human bias. In their survey, Moffat and Kelly [40] found significant bias against computer-generated music. Other researchers have tried to reproduce these results with a small sample but without success [25, 42, 44]. In addition, Hong [32] has analyzed the differences in perceptions of computer art within focus groups. For the same piece of art is more considered to be “art” when it is perceived to be of human origin than AI origin. In the field of art and paintings, Elgammal [23] has found conflicting results with a positive bias towards machines. Despite this, the bias against computers is widespread and often cited [36]: “This suggests that there is much more to investigate about our relation with artificially generated art” ([20], p. 159).

With increasing exposure to “more high-quality computer generated artifacts,” the following research questions have become “more pressing” ([17], p. 276). Art is a symbol of an activity long regarded as a human task. Thus, AI raises important questions in terms of acceptability and human-and-machine relationships, which must be studied in the field of HCI. Can people distinguish computer-generated paintings? Do people prefer systems (or paintings) created by humans? Do people have a negative bias towards AI? Finally, we seek to learn how assumptions about the identity of the painters (AI or human) of the same artworks influences their evaluation. This paper presents a large-scale experiment to assess the existence of a potential human bias towards works generated by AI by using the technique of priming effect (e.g., [3, 55]). The priming effect represents the influence of a first stimulus (the primer) on the processing of a subsequent stimulus (the target). Thus, our study evaluates the influence of the exposure of the primer (i.e., the declared identity of the painter, AI vs. human) on the perception of artistic targets (i.e., AI-generated paintings vs. paintings drawn by humans).

¹ Generative Adversarial Network (GAN) [25] consists in train jointly two separate models: a generator and a discriminator. The generator is train to generate the most realistic data (e.g., images, text, signal) and to fool the discriminator. The discriminator is train to discriminate real data (i.e., data from the training dataset) and fake data (i.e., data from the generator).

² <https://obvious-art.com/>

Method

Participants

565 participants (M = 32.80 years; S.D. = 9.91; 234 females and 331 males) were involved. The participants were recruited without specific criteria on Amazon Mechanical Turk (AMT) in order to select more diverse profiles and obtain a larger sample [11]. The reward paid to each worker was \$0.75.

Measures

As Amirshahi [1] has recommended with lead subjective tests, many proprieties related to the paintings were studied. Inspired by the works of Berlyne [5], Jordanous [33], and Elgammal [23], many dimensions were evaluated: liking ("I like this painting"), beauty ("This painting looks beautiful"), novelty ("This painting seems novel"), and meaning ("I perceive the meaning of the painting"). To evaluate subjective dimensions, questionnaires with Likert scales are used in this survey. The optimal number of rating bars seems to be seven [13]. For these four dimensions, participants were asked to indicate how much they agreed according to seven-point Likert scales (1 = totally disagree; 7 = totally agree). Each Likert had antagonist anchors without labels between.

Material

Following Pasquier [44], the selection of pieces of art was influenced by their similarity in order to conceal their authorship. For this purpose, 40 impressionist-style paintings made by Piet Mondrian, Claude Monet, Robbie Barrat, and the collective Obvious were selected³ (see **Erreur ! Source du renvoi introuvable.** for examples).

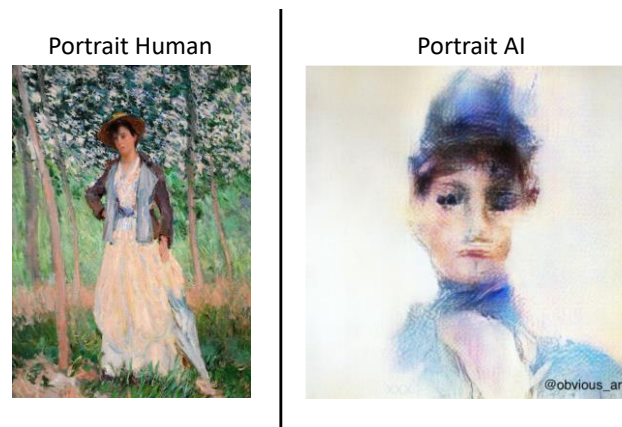


Figure 1. Examples of used art paintings (Human: The Stroller - Claude Monet; AI: Madame De Belamy - @obvious_art)

Protocol

An online survey was used to control the experimental conditions in this type of study [1]. The participants who accepted the task on AMT were redirected to the SurveyMonkey platform. A mixed-subject design was used; see **Erreur ! Source du renvoi introuvable.** Similar to [25],

³ Due to space limitation, the complete list of paintings is not presented. Please contact the first author for details.

participants were randomly assigned to one of the two conditions (i.e., Human or AI condition, the between-subject variable). Thus, participants saw one of the two primings and instructions presented in **Erreur ! Source du renvoi introuvable.** All participants saw portraits and landscapes actually made by humans and AI (within-subject design).

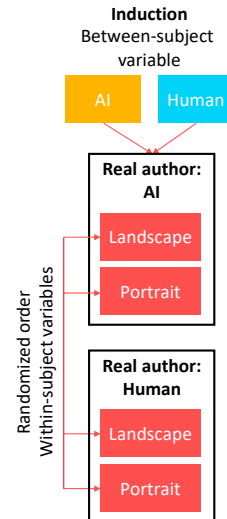


Figure 2. Experimental design

The artworks were randomly selected from 40 paintings (10 portraits by AI, 10 landscapes by AI, 10 portraits by humans, and 10 landscapes by humans) to improve the generalization of the results, to reduce experiment duration, and to avoid the repetition of painting styles. Next, for each painting, participants had to complete several items (i.e., liking, perceived beauty, novelty, and meaning). Finally, in order to ensure the effectiveness of the priming of the declared author type (i.e., AI or human), and following previous studies (e.g., [42]), an item of manipulation check was introduced (i.e., "At the beginning of the study, the identity of the painters was defined. Do you remember the identity of the painters?"). Afterward, it was explained that the origin of the painting had been manipulated. Participants then had to guess the origin of four paintings (one painting was randomly selected from each category). This modified TT was proposed at the end of the study to avoid any bias in evaluation and the priming effect. Lastly, two demographic questions about age and gender were asked.

Instruction: AI condition

8 paintings created by some Artificial Intelligence will be presented. You will be asked to rate them. There are no right or wrong answers. Only your opinion counts. Please respond spontaneously according to your feelings.

Instruction: Human condition

8 paintings created by some artists will be presented. You will be asked to rate them. There are no right or wrong answers. Only your opinion counts. Please respond spontaneously according to your feelings

Table 1: Instructions

Results

Mixed model analyses were used for each dependent variable. Thus, the participant and the paintings⁴ were considered random factors. The induction condition was considered a between-subject factor. The type of paintings and the real authors were considered within-subject factors. The statistical analysis was performed using R [47], lme4 [4], and afex [52]. The effect sizes were based on [10] (measured as d).

Participants who responded "I don't know" to the manipulation check were excluded (i.e., 79 out of 565 participants, or 13.98% of the sample). Moreover, for "Human condition" and "AI condition," 82% and 62%, respectively, of participants remembered their type of induction. Thus, to control the effect of this variable, answering correctly or incorrectly on this item was introduced as a random factor similar to [1].

The effects of the induction condition, the type of paintings, and the real authors were evaluated on declared liking, perceived beauty, perceived novelty, and meaning. All descriptive statistics are presented in **Erreur ! Source du renvoi introuvable.**, **Erreur ! Source du renvoi introuvable.**, Table 4, and Table 5.

Concerning *declared liking* (see **Erreur ! Source du renvoi introuvable.**), the analyses show a main effect of induction ($F_{1,481.25} = 17.67$, $p < .001$, $d = 0.06$), the type of painting ($F_{1,30.74} = 26.69$, $p < .001$, $d = 0.15$), and the real authors ($F_{1,30.74} = 82.44$, $p < .001$, $d = 0.25$).

	Human	AI
Induction	5.12 (1.56)	4.80 (1.66)
	Human	AI
Real Authors	5.18 (1.44)	4.72 (1.76)
	Land	Portrait
Painting type	5.35 (1.39)	4.55 (1.74)

Table 2. Descriptive statistics for *declared linking*. Land = landscape

⁴ To consider the variability between paintings, the painting is included as a random factor. Thus, the variability in perception between the paintings is absorbed in the model fitting.

Regarding *perceived beauty* (see **Erreur ! Source du renvoi introuvable.**), the analyses show a main effect of induction ($F_{1,481.23} = 17.74, p < .001, d = 0.05$), the type of painting ($F_{1,32.14} = 22.18, p < .001, d = 0.16$), and the real authors ($F_{1,32.14} = 64.81, p < .001, d = 0.27$).

	Human	AI
Induction	5.12 (1.53)	4.81 (1.67)
	Human	AI
Real Authors	5.20 (1.38)	4.71 (1.78)
	Land	Portrait
Painting type	5.37 (1.36)	4.54 (1.73)

Table 3. Descriptive statistics for *perceived beauty*. Land = landscape

In terms of *perceived novelty* (see Table 4), the analyses show a main effect of induction ($F_{1,481.28} = 13.31, p < .001, d = 0.11$), the type of painting ($F_{1,25.72} = 16.33, p < .001, d = 0.08$), and the real authors ($F_{1,25.72} = 3.47, p = .074, d = 0.04$).

	Human	AI
Induction	4.90 (1.46)	4.63 (1.59)
	Human	AI
Real Authors	4.87 (1.47)	4.63 (1.58)
	Land	Portrait
Painting type	4.81 (1.48)	4.70 (1.58)

Table 4. Descriptive statistics for *perceived novelty*. Land = landscape

Concerning *perceived meaning* (see Table 5), the analyses showed a main effect of induction ($F_{1,481.34} = 15.09, p < .001, d = 0.13$), the type of painting ($F_{1,28.81} = 19.61, p < .001, d = 0.10$), and the real authors ($F_{1,28.81} = 52.96, p < .001, d = 0.16$).

	Human	AI
Induction	4.81 (1.56)	4.48 (1.69)
	Human	AI
Real Authors	4.80 (1.52)	4.47 (1.74)
	Land	Portrait
Painting type	4.90 (1.50)	4.37 (1.73)

Table 5. Descriptive statistics for *perceived meaning*. Land = landscape

Lastly, participants had to distinguish paintings made by humans from generated-AI paintings. The analyses show an effect of the painting type on the recognition rate ($F_{1,1717.44} = 17.67, p < .001, \eta_p^2 = 0.34$) and the author's type ($F_{1,1717.44} = 64.41, p < .001, \eta_p^2 = 0.34$): see Table 6. Thus, paintings drawn by humans (compared to AI-generated paintings) and the portraits (compared to the landscapes) are the best-recognized artworks with the lowest percentage of recognition errors.

Type of Paintings / Real authors	Recognition rate
Landscape	53 %
Portrait	69 %
Human	66 %
AI	56 %

Table 6. Recognition rate (in percent) according to painting type and author type

Discussion

A phase with a manipulation check was proposed to participants to verify the relevance of the proposed induction. The responses suggest that a large majority of participants believed and remembered their type of induction (i.e., AI or human condition). These results seem consistent with those of Friedman [25]. In their first experiment, 83.87% and 65.27%, respectively, believed in the assertion in "Human Condition" and "AI condition." Unlike in the previous studies, here, the use of a large sample (565 participants), combined with the introduction of the answers in the manipulation check as a random factor, resulted in precise statistical analyses. Indeed, the results show significant effects of both the induced and real author. More specifically, the artworks presented as AI-generated paintings were significantly less liked and were perceived as less beautiful, novel, and meaningful than paintings presented as drawn by a human. The same pattern, with significant effect, is also apparent for the real type of authors of the artworks presented. Indeed, AI-generated paintings were less well evaluated (in terms of liking, beauty, novelty, and meaning) than paintings made by humans. These results support the first results obtained by Moffat and Kelly [40] on the perception bias towards computer-generated music. To the best of our knowledge, these results had never been replicated on a large sample, with the previous methodological precautions presented. Moreover, the modified TT, in which participants have to guess the real author of the paintings, showed a better recognition of human paintings (66%) than AI-generated paintings (56%). These results are consistent with Burnett [12] for composed-computer music and are opposed to results of Moffat and Kelly [40]. In Elgammal [23], 75% of respondents thought that the AI-generated paintings were made by humans. These differences could be explained by the improvement of the techniques in computational creativity, especially with the emergence of GAN. In parallel, it should be noted that participants are better able to identify the origin of the author for portraits (69%) than for landscapes (53%). Similar to Boden [8], with a recognition

rate close to chance (50%), people have difficulty distinguishing the type of authors. These results may suggest that the technical proprieties of AI-generated paintings are currently more advanced for the generation of landscapes than portraits. A possible explanation of the existing bias against AI-generated paintings could be that participants evaluated the pieces of art with an intergroup bias. AI may be anthropomorphized and thus considered as the out-group [21]. Intergroup bias can be defined as the tendency people have to evaluate their own group (i.e., the in-group) in a more positive manner than another group (i.e., the out-group) [6, 30]. Social changes relative to the out-group may bring a sense of threat and negative feelings that reinforce the intergroup bias [9, 24, 30]. The aim of the out-group devaluation is to maintain a positive social identity and superiority in some areas of competence [54]. To further explain the existence of a negative perception bias towards AI, several concepts can be cited, such as technophobia ([43]; see [49] for meta-analysis), anxiety towards machines, and reactive devaluation [50]. AI can be seen as a potential source of danger because it could replace humans in many areas [43]. To combat this bias, some authors have raised the question of framing information on user perceptions [7, 15, 35, 48, 53]. Indeed, “[f]raming information might be useful to help humans understand an agent who is unlike them; but this remains to be tested” ([36], p. 28:24), even if the experimental results seem to be unclear [39]. Thus, for Lamb ([36], p. 28:24), “a more subtle question is if humans are biased toward familiar and humanlike forms of creativity,” which could be linked to the previously mentioned intergroup bias.

Future works

Within this paper, a general bias against AI-generated artwork can be highlighted. In future works, the interaction effects across different parameters (e.g., induction and painting type) may be explored. Moreover, it may be interesting to include a measure of self-reported expertise in Art. Hekkert and Wieringen [28] have analyzed the differences in the art evaluation among experts and non-experts; they found a correlation between perceived originality and quality that was significantly higher among experts than non-experts. In the same way, Moffat and Kelly [40] have found a bias against computers that was significantly higher in musicians than non-musicians. Nevertheless, in a modified version of the TT, non-musicians were surprisingly better than musicians at recognizing music generated by computers or humans. In parallel, in order to make our results even more generic, the evaluation of a general bias against AI-generated artworks could be made with different styles of paintings (e.g., [23]) or types of art (e.g., music, poem). In this paper, public perception bias is questioned based on dichotomic human or AI priming. However, the frontiers between AI and human-generated artworks may appear more and more faint. For Elgammal [22], AI has even blurred the definition of artists. More generally, this raises the issue of authorship: “Can an Artificial Intelligence make Art without artists?” [2]. An initial answer could be in the eye or mind of the viewer.

Conclusion

The idea of a negative perception bias against machines, computers, and/or AI is widely shared, particularly in computational creativity. However, there has been a lack of experimental proof that demonstrates evidence of this bias. Unlike previous studies, the current large-scale experiment with 565 participants focused on AI and AI-generated paintings with different methodological precautions (large sample, randomization, manipulation check as random factor, etc.). In addition to the large sample, the main contributions of this paper are the methodological precautions and the robustness of the results, which seem to represent significant progress in the field of HCI and computational creativity to demonstrate the existence of a negative perception bias towards AI at the expense of human-made systems. Thus, depending on the perceived identity of the author (Human vs. AI), the same artworks were evaluated differently. Moreover, the results show that real artworks made by humans are also evaluated more highly than real AI-generated artworks. Alongside technological advances in AI, this paper paves the way for other studies to further consider perception bias in computational creativity and to analyze the impact of AI on human representations.

Acknowledgement

This study was carried out within b<>com, an institute of research and technology dedicated to digital technologies. It received support from the Future Investments program of the French National Research Agency (grant no. ANR-07-A0-AIRT).

References

- [1] Seyed Ali Amirshahi, Gregor Uwe Hayn-Leichsenring, Joachim Denzler, and Christoph Redies. 2015. JenAesthetics Subjective Dataset: Analyzing Paintings by Subjective Scores. *Computer Vision - ECCV 2014 Workshops*, Springer International Publishing, 3–19.
- [2] Sofian Audry and Jon Ippolito. 2019. Can Artificial Intelligence Make Art without Artists? Ask the Viewer. *Arts* 8, 1: 35.
- [3] John A. Bargh, Mark Chen, and Lara Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71, 2: 230–244.
- [4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1: 1–48.
- [5] D. E. Berlyne. 1971. *Aesthetics and psychobiology*. Appleton-Century-Crofts, East Norwalk, CT, US.
- [6] Michael Billig and Henri Tajfel. 1973. Social categorization and similarity in intergroup behaviour. *European journal of social psychology* 3, 1: 27–52.
- [7] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. “It’s Reducing a Human Being to a Percentage”: Perceptions of Justice in Algorithmic Decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, ACM Press, 1–14.
- [8] Margaret A. Boden. 2010. The Turing test and artistic creativity. *Kybernetes*.
- [9] Marilyn B Brewer. 2001. Ingroup identification and intergroup conflict. *Social identity, intergroup conflict, and conflict reduction* 3: 17–41.
- [10] Marc Brysbaert and Michaël Stevens. 2018. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition* 1, 1: 9.
- [11] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1: 3–5.
- [12] Adam Burnett, Evon Khor, Philippe Pasquier, and Arne Eigenfeldt. 2012. Validation of Harmonic Progression Generator Using Classical Music. 126–133.
- [13] M. Y. Cai, Y. Lin, and W. J. Zhang. 2016. Study of the optimal number of rating bars in the likert scale. *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services - iiWAS ’16*, ACM Press, 193–198.
- [14] Wei-Ta Chu, Hideto Motomura, Norimichi Tsumura, and Toshihiko Yamasaki. 2019. A Survey on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia. *ITE Transactions on Media Technology and Applications* 7, 2: 60–67.
- [15] Simon Colton. 2008. Creativity Versus the Perception of Creativity in Computational Systems. *AAAI Spring Symposium: Creative Intelligent Systems*.
- [16] Simon Colton. 2012. The Painting Fool: Stories from Building an Automated Painter. In J. McCormack and M. d’Inverno, eds., *Computers and Creativity*. Springer Berlin Heidelberg, Berlin, Heidelberg, 3–38.
- [17] Simon Colton, Alison Pease, and Rob Saunders. 2018. Issues of Authenticity in Autonomously Creative Systems. *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.*, 272–279.
- [18] David Cope. 2005. *Computer models of musical creativity*. MIT Press, Cambridge, Mass.
- [19] David Cope and Douglas R. Hofstadter. 2001. *Virtual music: computer synthesis of musical style*. MIT Press, Cambridge, Mass.
- [20] Antonio Daniele and Yi-Zhe Song. 2019. AI + Art = Human. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society - AIES ’19*, ACM Press, 155–161.
- [21] Chad Edwards, Autumn Edwards, Brett Stoll, Xialing Lin, and Noelle Massey. 2019. Evaluations of an artificial intelligence instructor’s voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior* 90: 357–362.

- [22] Ahmed Elgammal. 2019. AI Is Blurring the Definition of Artist. *American Scientist* 107, 1: 18.
- [23] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms. *arXiv:1706.07068 [cs]*.
- [24] Victoria M Esses, Lynne M Jackson, and Tamara L Armstrong. 1998. Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of social issues* 54, 4: 699–724.
- [25] Ronald S. Friedman and Christa L. Taylor. 2014. Exploring emotional responses to computationally-created music. *Psychology of Aesthetics, Creativity, and the Arts* 8, 1: 87–95.
- [26] Berys Gaut. 2010. The Philosophy of Creativity: Philosophy of Creativity. *Philosophy Compass* 5, 12: 1034–1046.
- [27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. 2014. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*.
- [28] Paul Hekkert and Piet C. W. Van Wieringen. 1996. Beauty in the Eye of Expert and Nonexpert Beholders: A Study in the Appraisal of Art. *The American Journal of Psychology* 109, 3: 389–407.
- [29] Beth A. Hennessey and Teresa M. Amabile. 2010. Creativity. *Annual Review of Psychology* 61, 1: 569–598.
- [30] Miles Hewstone, Mark Rubin, and Hazel Willis. 2002. Intergroup bias. *Annual review of psychology* 53, 1: 575–604.
- [31] Florian Hoenic. 2005. Defining Computational Aesthetics. *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Eurographics Association, 13–18.
- [32] Joo-Wha Hong. 2018. Bias in Perception of Art Produced by Artificial Intelligence. *Human-Computer Interaction. Interaction in Context*, Springer International Publishing, 290–303.
- [33] Anna Jordanous. 2014. Stepping Back to Progress Forwards: Setting Standards for Meta-Evaluation of Computational Creativity. .
- [34] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-Aware Fashion Recommendation and Design with Generative Image Models. *arXiv:1711.02231 [cs]*.
- [35] René F. Kizilcec. 2016. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, ACM Press, 2390–2395.
- [36] Carolyn Lamb, Daniel G. Brown, and Charles L. A. Clarke. 2018. Evaluating Computational Creativity: An Interdisciplinary Tutorial. *ACM Comput. Surv.* 51, 2: 28:1–28:34.
- [37] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553: 436–444.
- [38] Jay H. Lee, Joohyun Shin, and Matthew J. Realf. 2018. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering* 114: 111–121.
- [39] Stephen McGregor, Matthew Purver, and Geraint Wiggins. 2016. Process Based Evaluation of Computer Generated Poetry. *Proceedings of the INLG 2016 Workshop on Computational Creativity in Natural Language Generation*, Association for Computational Linguistics, 51–60.
- [40] David C Moffat and Martin Kelly. 2006. An investigation into people’s bias against computational creativity in music composition. *Proceedings of the 3rd International Joint Workshop on Computational Creativity (ECAI06 Workshop)*.
- [41] Danil Nagy, Damon Lau, John Locke, et al. 2017. Project discover: an application of generative design for architectural space planning. Society for Computer Simulation International, 7.
- [42] David Norton, Derrall Heath, and Dan Ventura. 2015. Accounting for Bias in the Evaluation of Creative Computational Systems: An Assessment of DARCI. *Proceedings of the Sixth International Conference on Computational Creativity*, 31–38.

- [43] Changhoon Oh, Taeyoung Lee, Yoojung Kim, SoHyun Park, Sae bom Kwon, and Bongwon Suh. 2017. Us vs. Them: Understanding Artificial Intelligence Technophobia over the Google DeepMind Challenge Match. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, ACM Press, 2523–2534.
- [44] Philippe Pasquier, Adam Burnett, and James Maxwell. 2016. Investigating Listener Bias Against Musical Metacreativity. *Proceedings of the Seventh International Conference on Computational Creativity (ICCC 2016)*, Sony CSL.
- [45] Marcus T. Pearce and Geraint A. Wiggins. 2001. Towards A Framework for the Evaluation of Machine Compositions. .
- [46] Alison Pease and Simon Colton. 2011. On impact and evaluation in computational creativity: a discussion of the Turing Test and an alternative proposal. *Proceedings of AISB '11: computing and philosophy*, Society for the Study of Artificial Intelligence and Simulation of Behaviour, 15–22.
- [47] R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [48] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, ACM Press, 1–13.
- [49] Larry D. Rosen and Phyllisann Maguire. 1990. Myths and realities of computerphobia: A meta-analysis. *Anxiety Research* 3, 3: 175–191.
- [50] Lee Ross. 1995. Reactive Devaluation in Negotiation and Conflict Resolution. In K. Arrow, R. Mnookin, L. Ross, A. Tversky, and R.B. Wilson, eds., *Barriers to Conflict Resolution*. New York.
- [51] Oscar Schwartz and Benjamin Laird. 2019. bot or not. Retrieved July 27, 2019 from <http://botpoet.com>.
- [52] Henrik Singmann, Ben Bolker, Jake Westfall, and Frederik Aust. 2019. *afex: Analysis of Factorial Experiments*. .
- [53] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*, ACM Press, 830.
- [54] Henri Tajfel and John Turner. 2004. An integrative theory of intergroup conflict. In M.J. Hatch and M. Schultz, eds., *Organizational identity: A reader*. Oxford University Press, 56–65.
- [55] Endel Tulving, Daniel L. Schacter, and Heather A. Stark. 1982. Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, 4: 336–342.
- [56] Geraint A. Wiggins. 2006. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19, 7: 449–458.