

Putting the Art in Artificial: Aesthetic responses to computer-generated art

Abstract

As artificial intelligence (AI) technology increasingly becomes a feature of everyday life, it is important to understand how creative acts, regarded as uniquely human, can be valued if produced by a machine. The current studies sought to investigate how observers respond to works of visual art created either by humans or by computers. Study 1 tested observers' ability to discriminate between computer-generated and man-made art, and then examined how categorisation of art works impacted on perceived aesthetic value, revealing a bias against computer-generated art. In Study 2 this bias was reproduced in the context of robotic art, however it was found to be reversed when observers were given the opportunity to see robotic artists in action. These findings reveal an explicit prejudice against computer-generated art, driven largely by the kind of art observers believe computer algorithms are capable of producing. These prejudices can be overridden in circumstances in which observers are able to infer anthropomorphic characteristics in the computer programs, a finding which has implications for the future of artistic AI.

Keywords: computer art, aesthetics, image statistics, intentionality, embodiment

Introduction

The creation of beautiful and emotive artworks that are commensurable with man-made artworks would represent a significant milestone in the development of artificial intelligence (AI). The computer artist Harold Cohen captured the challenge of creating an agent whose work can match up to the complexity of man-made artworks in the following quote:

Would it be possible, for example, for the machine to produce a long series of drawings rather than a single drawing, different from each other in much the same way that the artist's would be different, unpredictable as his would be unpredictable, and changing in time as his might change? (p.1, Cohen, 1973)

The unpredictable nature of human creativity as articulated by Cohen is the basis of the Lovelace Test for evaluating AI. In the Lovelace Test a computer program is deemed intelligent only if it creates a routine that it was not initially engineered to create (Bringsjord et al., 2003). Critically, for intelligence to be established, the designers of the AI must not be able to explain how the agent came to create the new routine. While to date no computationally creative system has been deemed intelligent by Lovelace, computer generated drawings, paintings, poetry, and music have begun to establish themselves as legitimate forms of art. An important and understudied psychological question relating to this phenomenon is the extent to which individuals are willing to accept computer art as having the same worth and aesthetic value as that of a human artist, regardless of whether it passes such stringent tests of human-level intelligence.

A brief history of computer art

The birth of computer art can be traced back roughly to the creation of the computer itself. Computer art was initially seen as a means of accessing objectivism in art and as a result early programs often focused on form over content. An early example is 'Hommage a Paul

Klee' by Fredier Nake ¹. To create this work Nake programmed randomly specified instances of variables which allowed the computer to make formal choices based on probability theory (Nake, 2005). At the same time, Noll created a computer-generated Mondrian-like artwork, which when shown next to a reproduction of a real Mondrian "Composition with Lines" (1917) was indistinguishable and often preferred over the true Mondrian (Noll, 1966). The artist Harold Cohen has been experimenting with computer simulations of the cognitive processes underpinning drawing and painting in a computer program named AARON since the 1970s². The AARON program does not perceive the world through direct observation, instead Cohen writes the structure of the stimuli that it paints or sketches into the code (McCorduck, 1990). Over the past few decades, computer-generated art has improved on its techniques and expanded its toolbox to include new learning algorithms and evolutionary computing as a means of generating novel artworks. Despite the swift advances in and wider availability of computer-generated art, there is little psychological research concerning its impact in terms of human computer interaction (HCI) and aesthetics. Such research is critical to understanding whether and how computer-generated art can be assimilated into society. At present it is unclear whether people are able to distinguish between computer and man-made artwork, and if they can, what impact this has on their aesthetic impressions of the artwork.

The computer-art bias

Whilst the ability of individuals to discern between computer and man-made art has not been directly empirically addressed, it is known that audiences are sensitive to the provenance of artworks in other contexts. Hawley-Dolan and Winner (2011) tested the proposition that abstract expressionist artworks created by professional artists cannot be differentiated by

¹ Hommage a Paul Klee by Fredier Nake: <http://dada.compart-bremen.de/item/agent/68>

² <http://www.aaronshome.com/aaron/gallery/IF-recent-gallery.html>

those created by animals or young children. The authors presented works of art by both artist groups (professional/child-animal) and measured participants' evaluation and preference for the artworks. Participants consistently preferred and valued the professionally-made artworks. This effect was moderated by the educational background of participants: art students preferred art works by professional artists more than non-art students. The authors suggest that participants made their preference and judgment decisions based on perceived intentionality: the 'mind' behind the art. In a follow-up study Snapper *et al.* (2015) provided further support for this conclusion by showing that participants make discriminations between the two types of artworks on the basis of perceived intentionality and the emergence of structure in abstract painting.

With regard to the impact of source knowledge on perceived aesthetic value, some empirical evidence sheds light on whether the provenance of a work of art impacts on aesthetic and value appraisals. Kirk *et al.* (2009) presented images to participants that were labelled as originating from an art gallery or generated by the experimenter in Photoshop. Images that were labelled as Photoshopped were rated as less aesthetically pleasing even though they were visually identical to those images labelled as being from an art gallery. Moffat and Kelly (2006) conducted a similar study using musical pieces composed by either humans or a computer. They found that participants could differentiate pieces of music composed by computer from those composed by humans. Furthermore, participants preferred music composed by humans but these freely made preferences were not altered by labelling the pieces of music as composed by human or computer. Musicians showed a greater bias against computer-generated music pieces than non-musicians. Labelling the artwork as made by a professional artist or a child/animal also had little impact on participants' preference and value judgments for the artworks in Hawley-Dolan and Winner's (2011) study. Together these studies suggest that there is a bias against computationally derived works of art, but the

root of this bias remains unclear. It could be that lower perceived aesthetic value ratings are a consequence of a high-level cognitive judgment that computer art is less valuable (explicit prejudice) or inherent visual characteristics of computer generated art that are disliked (implicit prejudice).

Factors affecting value, aesthetic and categorisation judgments of art

A number of factors may affect categorisation and judgments of works of art. Art philosophers have argued that observers assess artwork as the ‘end point of a performance’(Dutton, 2003). Under this view what is known about the process that governed the creation of the work is just as important as the final product in determining its aesthetic and artistic value. Hawley-Dolan and Winner (2011) emphasised the role of intentionality judgments in appraising an artwork, later clarifying that objective appraisals (i.e. measures of value) are more likely to be impacted by intentionality than subjective measures (i.e. aesthetic preference; Hawley-Dolan and Young, 2013). Newman and Bloom (2011) found that the financial value of an art object is determined by the degree to which it is viewed as a unique creative act and by the amount of physical contact that the original artist has with the art object. Similarly, Kruger *et al.* (2003) and Jucker *et al.* (2013) demonstrated that artworks that appeared to take more time and effort to produce were rated highly for quality, value and liking. Given that there are several different means of production for computer-based artwork, and that relatively little is known about the artistic processes involved (Colton, 2008), issues of intentionality, authenticity and effort each have relevance for acceptance and appreciation of computer-generated art.

Observers may rely on surface level indicators in the artworks to form assumptions of intentionality, authenticity, and effort in order to guide their evaluations. For example, the presence of physical brush strokes may invoke a greater sense of uniqueness and physical

contact between the artist and the artwork. It has recently been shown in marketing research that giving a product the quality of being hand-made increases its perceived attractiveness (Fuchs *et al.*, 2015). Consequently, the convincing production of surface effects (non-photorealistic rendering) is a priority for computer graphics. Isenberg *et al.* (2006) conducted an investigation of computer-generated and hand-drawn stippling effects. They found that participants could distinguish between computer-rendered and hand-drawn stippling effects, although this did not lead to one form being valued more highly than the other. Maciejewski *et al.* (2007) suggest that the precision of computer-generated images often leads to a sense of rigidity in the resulting images whereas hand-drawn images are less sterile and may possess statistical properties that imply self-similarity, much like natural surfaces.

Artworks are often a reflection of the natural world, and studies have shown that the image statistics (measurable properties of images such as spatial frequency, edge strength, etc.), are similar between the real world and artistic representations of it. Both artworks and natural scenes are characterised by their scale-invariant (Graham & Field, 2007; Redies *et al.*, 2007; Alvarez-Ramirez *et al.*, 2008) and fractal properties (Taylor, 2002; Taylor *et al.*, 2011). These statistics have also been shown to impact aesthetic evaluation. Self-similarity, the property of the whole having the same appearance as its parts (similar to fractality) (Amirshahi *et al.*, 2012, 2013; Mallon *et al.*, 2014; Redies *et al.*, 2012), complexity, the regularity or heterogeneity of the pattern (Forsythe *et al.*, 2011; Redies *et al.*, 2012), and anisotropy, the uniform distribution of oriented edges (Koch *et al.*, 2010; Melmer *et al.*, 2013; Redies *et al.*, 2012) have all been shown to modulate aesthetic experience and characterise aesthetic art forms. Therefore, low-level image properties, such as the appearance of lines or brushstrokes, could constitute important visual information which guides the classification and aesthetic appraisal of computer-generated images.

The importance of embodiment in art

Robotic art is a special class of computer-generated art. Robotic artists function beyond the conventional plotter style output devices used in many forms of computer art as they have varying amounts of affordance and can respond to ongoing feedback during the generation of the artwork (Tresset & Leymarie, 2013; Deussen *et al.*, 2012), tying the embodiment process explicitly into the process of creation. The ability of the observer to perceive the embodiment of the artist through their artwork is thought to be a vital part of the esthetic response (Freedberg & Gallese, 2007). The connection between action systems (particularly the mirror neuron system) and perception of artworks is supported by cognitive neuroscientific evidence which shows that mu rhythm suppression (a proxy for mirror neuron activation) as well as activation of motor and premotor cortices is elicited when participants observe the dynamic artworks of Lucio Fontana and Franz Kline (Umilta *et al.*, 2012; Sbriscia-Fioretti *et al.*, 2013). This ability to have the movements and experiences of the artists projected into our minds and bodies triggers an empathic response in the observer. Increased empathy is said to enhance the spectator's emotional response to the piece by allowing for a direct understanding of the inner world of the artist. In support of this, participants who were primed for particular brushstrokes or who covertly generated particular brushstrokes while viewing works of art preferred artworks made using similar brushstrokes, suggesting an aesthetic perception-action congruency effect (Leder *et al.*, 2012; Ticini *et al.*, 2014). As the mirror neuron system is equally activated when watching humans and robots perform actions (Gazzola *et al.*, 2007), it is likely that seeing a robot produce a work of art may elicit a similar empathic aesthetic response to that elicited by the simulation of human artistic actions.

While insight into the actions necessary to create a work of art may increase the aesthetic impact of robotic art, aspects of robotic actions may increase the impression of intentionality

within the robotic system, thereby heightening the aesthetic experience of the generated artworks. It has been shown that individuals show more empathy toward more anthropomorphic robots (Riek et al., 2009; Darling, 2015). Convincing simulation of humanlike hand and eye movements may induce implicit impressions of animacy and anthropomorphism, potentially attributing greater value and pleasure to robotic art through empathic responses (Bartneck et al., 2009). Thus, the level of anthropomorphism granted to robotic artists may be an important factor in how the artwork produced by the robot is evaluated.

Aims

The ability to classify and assign aesthetic value to computer-generated art may be governed by a number of factors. They may be impacted by the embodiment of the system, by perceived intentionality within the system, by surface properties of the artwork or by higher-level cognitive biases concerning the inherent personal, and societal value of computer-generated art. In two studies we set out to explore these factors in more detail. In the first study we allowed participants to generate their own labels for computer-generated and man-made artworks and asked them to provide perceived aesthetic value ratings for them either before or after classifying them. We then assessed whether surface properties of the images (low-level image statistics) predicted the classification and perceived aesthetic value ratings attributed to them. In the second study we assessed the impact of embodiment of a computer-generated art system by measuring responses to artworks made by a robot with or without the robot present.

Study 1

The aim of the first study was to investigate aesthetic bias against computer-generated art. A previous study found that if images were labelled as being generated in Photoshop they were

rated less aesthetically pleasing than those labelled as originating from an art gallery context (Kirk *et al.*, 2009). In the current study the definition of computer art was broadened and the effect of participants providing their own labels for the artworks was studied in the context of genuinely computer-generated images. The definition of computer-generated art provided to participants was, ‘any work of art (either abstract or representational) that uses digital technology as an *essential* part of the creative or presentation process.’ This could encompass a wide range of artworks including those generated with accompanying visual and haptic feedback in robotics (e.g. eDavid, Deussen *et al.*, 2012), those in which image structure is pre-coded and then randomly composed (e.g. AARON, Cohen, 1973), algorithmic art defined using mathematical principles (e.g. artist and designer Max Bill, Bill, 1993) and work that takes in multimodal input transformed into action (e.g. artist Benjamin Grosser, Grosser, 2011). Participants were given no further information about how a computer-generated work of art could be made. A between-subjects blocked design was used in which categorisation of a group of artworks took place either before or after the same artworks were aesthetically rated. Image type (computer-generated/man-made) was manipulated within participants. It was predicted that those participants that categorised the artworks first would later show a bias by rating those artworks that they categorised as computer-generated as less aesthetically pleasing than those they categorised as man-made. By contrast, it was predicted that those participants who aesthetically rated the artworks first and then categorised them would not show an aesthetic bias against computer-generated art, as they would assume all the works were man-made. To test the assumption that visual properties of the images would drive categorisation and perceived aesthetic value rating of the images, the role of image statistics was investigated in relation to the dependent variables. The image statistics used were: the slope of the Fourier spectrum, anisotropy, self-similarity, and complexity derived from the Pyramid Histogram of Oriented Gradients (PHOG; Amirshahi *et al.*, 2012; Redies *et al.*,

2012). It was predicted that those images whose low-level statistics were closer to that of natural scenes would be classified more readily as man-made and would receive higher perceived aesthetic value ratings. In addition, it was hypothesised that observers with an art-educational background would be better at identifying man-made artworks and would show a greater bias against computer-generated art, in line with the findings of Hawley-Dolan and Winner (2011) as well as Moffat and Kelly (2006).

Method

Participants. Participants were recruited online from the KU Leuven student and staff community and from an international mailing list for a drawing research network. Data collection took place over three weeks. Participation in the online study was voluntary. The final sample (N=65) consisted of 20 art-educated participants (9 female, mean age 42.65 years (SD=20.40)) and 45 non-art-educated participants (28 females, mean age 28.02 (SD=12.86)).

Materials and Procedure. The online study took approximately 20 minutes to complete. Participants first completed an online consent form and a demographic questionnaire that included questions about their background in art and design (Appendix 1). Participants were then shown a randomised series of 60 artistic images. Computer-generated artworks (n=30) were selected from computer art databases online. They were then broadly matched by the authors for mode of production (physical paint on canvas, ink, digital etc.) as well as content (landscape, portrait, abstract shape, etc.) with man-made counterparts (n=30). Half of the image set was abstract and half was representational, creating four image types: 15 Abstract computer-generated, 15 Representational computer-generated, 15 Abstract man-made, 15 Representational man-made (see Appendix 2 for a list and links).

Each participant was randomly allocated to either a ‘categorise first’ (n=34) or ‘rate first’ (n=31) condition. In the ‘categorise first’ condition for each of the 60 images participants were asked to judge whether the image was man-made or computer-generated in a two-alternative forced choice. In the ‘rate first’ condition participants were asked to rate how much they liked each image on a scale of 1-7 (1=very unattractive, 7=very attractive). Each participant then completed the alternate task (rate/categorise) in the second phase of the study. At the end of the study, participants were asked to report how they made their categorisation judgment in a free-response format. They were asked, ‘how did you decide if a work was computer-generated?’ and ‘how did you decide if a work was man-made?’. At the end of the study participants received feedback on the number of computer-generated images they correctly identified.

Results

Categorisation performance. Participants’ categorisation performance was calculated by producing an average of correct responses (Figure 1). Out of the total 65 participants, 39 got more than 50% of the items correct. Mean accuracy across all participants was low at 52.49% (SD=6.09%), but was significantly different from chance, $t(64) = 3.29, p < .01, 95\% \text{ CI } [50.98, 54.00]$, Cohen’s $d = 0.41$. Performance was then split for man-made art trials and computer-generated art trials. It was revealed that man-made art was successfully detected at 64.66% (SD=12.12%) accuracy, in comparison to computer-generated art which yielded a much lower accuracy rate of 40.31% (SD=11.27%). Accuracy for the two types of images was significantly different from one another, $t(64) = 9.83, p < .001, 95\% \text{ CI } [19.41, 29.31]$, Cohen’s $d = 1.22$. Further analysis revealed that participants were biased to respond that images were man-made, choosing the man-made category 62.18% of the time. This was split by image type (representational/abstract); participants chose the man-made category 72.56%

of the time for representational images, which was statistically more frequent than would be predicted by chance, $t(29) = 7.08, p < .001, d = 1.29$, while participants only chose the man-made category 51.79% of the time for abstract images, which was not significantly different from chance, $t(29) = 0.34, p = 0.73, d = 0.06$. This indicates that participants were biased to believe that representational images were man-made which is reflected in a break-down of the accuracy data by image type (Figure 1). Art-educated participants ($M = 54.42\%$, $SD = 5.62\%$) performed better than non-art-educated participants ($M = 51.63\%$, $SD = 6.78\%$) but the difference in accuracy between the two groups was not significant, $t(32) = 1.61, p = .11$, 95% CI [0.06, -0.01], Cohen's $d = .40$. Age significantly predicted accuracy, $r(63) = .42, p < .001$, 95% CI [0.19, 0.60]; such that older participants were better at the task. As art-educated participants also tended to be older, the relationship between accuracy and subgroup membership could be driven by age differences. Indeed, when art education was included as a predictor in a linear regression between age and accuracy, the model fit was significant, $F(2, 62) = 6.56, p < .01, R^2 = 0.15$, and age remained a significant predictor in the model, $\beta = 0.39, p < .01$, whereas art education subgroup did not contribute significantly to the variance in accuracy, $\beta = 0.05, p = 0.67$. There was however no significant correlation between the age of participants and how aesthetically pleasing they found the images, $r(63) = 0.14, p = 0.27$. A k-means cluster analysis on accuracy per image gave rise to a single cluster, suggesting that there were no particular images that drove performance above chance. Accuracy for computer-generated and man-made artworks can be seen in Figure 2.

=====

INSERT FIGURE 1 HERE

=====

=====

INSERT FIGURE 2 HERE

Effect of image properties and choice on perceived aesthetic value. To assess whether aspects of the images predicted perceived aesthetic value, image type (abstract/representational) and image source (computer/man-made) were assessed in relation to perceived aesthetic value ratings in a 2X2 within-subjects ANOVA. There was no significant impact of image type, $F(1, 64) = 2.78, p = 0.10, \eta_p^2 = 0.04$, and image source, $F(1, 64) = 1.25, p = 0.27, \eta_p^2 = 0.02$, on perceived aesthetic value ratings and no significant interaction between source and type, $F(1, 64) = 2.46, p = 0.12, \eta_p^2 = 0.04$ (Figure 3).

INSERT FIGURE 3 HERE

To calculate the impact of participants' categorisation of the images on perceived aesthetic value, mean aesthetic ratings were calculated for those trials that participants selected as computer-generated and those trials that participants selected as man-made. A within-subjects ANOVA was then conducted in which choice (computer/man-made) and image type (representational/abstract) were independent variables. There was a main effect of choice, $F(1, 64) = 7.65, p < .01, \eta_p^2 = 0.11$, but no effect of image type, $F(1, 64) = 2.72, p = 0.10, \eta_p^2 = 0.04$, and no significant interaction between choice and image type, $F(1, 64) = 3.77, p = 0.06, \eta_p^2 = 0.06$, implying that the aesthetic bias toward computer generated images was not driven by their content (abstract or representational). A mixed-model ANOVA was then conducted in which choice (computer/man-made) was the within-subjects variable and order (rate first/categorise first) was the between-subjects variable. There was a main effect of choice, F

(1, 63) = 26.69, $p < .001$, $\eta_p^2 = 0.30$ but no effect of order, $F(1, 63) = 0.02$, $p = 0.90$, $\eta_p^2 < .001$, and no interaction between order and choice, $F(1, 63) = 0.08$, $p = 0.78$, $\eta_p^2 = 0.001$ (Figure 4), implying that the order in which the blocks were given did not impact perceived aesthetic value ratings.

=====
INSERT FIGURE 4 HERE
=====

Effects of expertise on computer choice aesthetic bias. To investigate whether expertise moderated the impact of image choice on aesthetics, a mixed-model ANOVA was conducted with choice (computer/man-made) as a within-subjects variable and art-educated group (art-educated/non-art-educated) as the between-subjects variable. There was a significant main effect of choice; images categorised as computer-generated were given lower aesthetic ratings than images categorised as man-made, $F(1, 63) = 26.66$, $p < .001$, $\eta_p^2 = 0.30$. There was no main effect of group, $F(1, 63) = 2.02$, $p = 0.16$, $\eta_p^2 = 0.03$, and there was no significant interaction between group and choice, $F(1, 63) = 0.02$, $p = 0.88$, $\eta_p^2 < .001$ (Figure 5).

=====
INSERT FIGURE 5 HERE
=====

Relationship between image statistics and categorisation. Self-similarity, Fourier slope, complexity, and anisotropy were calculated from a PHOG for each image (for more information on image statistic calculation, see Amirshahi et al., 2012, Braun et al., 2013; Redies et al., 2012). A series of repeated measures *t*-tests were conducted to compare image

statistics for the different image types (Table 1). The only significant result was a difference in anisotropy between abstract and representational images. Representational images were less anisotropic than abstract images. This indicates that orientations of brush strokes/lines were more uniformly distributed in the representational images compared to the abstract images. (For examples of low and high anisotropy see Appendix 2 for work by Verostko and Rosen, respectively.)

=====
INSERT TABLE 1 HERE
=====

The values for these image statistics were then correlated with the degree to which each image was categorised as computer-generated or man-made and with each image's aesthetic rating (Table 2). None of the image statistics predicted aesthetic ratings. It was found that anisotropy of images predicted the extent to which they were categorised as man-made. Less anisotropic images were more likely to be categorised as man-made. As anisotropy predicted both categorisation and image type (representational/abstract) and participants more frequently selected representational works as man-made, the link between anisotropy and categorisation could be attributed to differences in anisotropy between the image types. Therefore, a logistic regression was performed to ascertain the effects of anisotropy and image type on the likelihood that an image would be selected as man-made. The logistic regression model was statistically significant, $\chi^2(2) = 261.50, p < .001$, and the model explained 19.11% (McFadden R^2) of the variance in image choice. Anisotropy, $e^\beta < .001$, $p < .01$, and image type, $e^\beta = 1.90, p < .01$, were both independent predictors of image choice. This suggests that observers used increased dispersion of line orientations and representational content to inform their categorisation decisions.

=====

INSERT TABLE 2 HERE

=====

Analysis of free-response. The free-responses provided at the end of the study were qualitatively analysed using thematic analysis (Braun & Clarke, 2006). Key phrases and words were extracted from the free responses and then placed into themes and subthemes. The overarching themes that observers used to inform their categorisation decisions were: surface, structure, content, and intentionality. The number of times each keyword was used per participant was calculated and the data are presented in Table 3. The most common categorisation justification for man-made images was that they had the appearance of being hand-made; often evidenced by the presence of brush strokes. Computer-generated images were classified predominantly on the basis of having bright or artificial colours and rigid, straight, regular forms, and lines. In comparison, there were relatively few references to content and intentionality in participants' justifications, with the main focus being on the surface and structural content of the artworks.

=====

INSERT TABLE 3 HERE

=====

Discussion

The aim of the first study was to investigate whether observers could categorise man-made and computer-generated artworks accurately and whether categorisation of artworks impacted on perceived aesthetic value. Accuracy on the source categorisation task was low suggesting

that the provenance of computer-generated art cannot be easily detected from surface and content characteristics alone. Statistical properties of the images did not predict accuracy, nor did they differ according to the ground-truth categories, suggesting that computer-generated images, like man-made images, have similar characteristics as natural scenes. This is perhaps not surprising given the broad range of computer-art used in the current study and the fact that artworks were matched for mode of production (e.g. paint or digital) and content (semantic meaning) with their man-made counterparts. However, the matching procedure used to balance the high level semantics of the images may have resulted in a subset of man made images that are not representative of the entire genre. Therefore, it is important to note that these findings should only be interpreted with respect to the current image set used in the experiment and cannot speak to the overall comparison of human and computer made images.

Low accuracy in the categorisation task was in part due to participants' bias to believe that the majority of the artworks in the sample were man-made. This bias effect was driven by the fact that representational images were disproportionately categorised as man-made. Whilst in the free-response task participants rarely cited representational content as a reason for selecting works as man-made, this exposes participants' bias to believe that computer-generated art deals for the most part with abstract representation. This is most likely due to the notoriety of abstract and geometric computer art created during the 1960s. A representational bias withstanding, there was no corresponding bias for abstract artworks to be categorised as computer-generated. In the instance of abstract artworks, it appeared that participants drew upon surface properties of the artworks like visible brushstrokes and irregular, imperfect lines to spot man-made qualities. Contrary to the findings of Hawley-Dolan and Winner (2011), intentionality was not frequently cited as a justification for categorising an artwork as man-made rather than computer-generated. Bias and low accuracy

was present in both the art-educated and non-art-educated participants suggesting that these effects are not moderated by expertise.

It was hypothesised that due to the previously reported bias observed against machine made artworks, those images that participants categorised as man-made would receive higher perceived aesthetic value ratings than those that were categorised as computer-generated, specifically in those participants that categorised the artworks before aesthetically rating them. This hypothesis was driven by the assumption that participants would remember which category they initially assigned the artwork to and subsequently their bias would cause them to rate these images as less aesthetically pleasing. This hypothesis was partially supported such that participants did rate those images they categorised as computer-generated as having a lower aesthetic rating than those they categorised as man-made, supporting previous findings by Kirk et al. (2009). However, this effect was present both in the group that rated the images first and in the group that categorised the images first. Those subjects that categorised the images first may have suspected that some of the images were computer generated based on the categorisation dichotomy while those that rated first would not share these suspicions; however, images that were categorised as computer-generated were rated as visually less pleasing, regardless of the order of rating and categorisation.

This unexpected finding could have multiple interpretations. First, it suggests that rather than the assumption of provenance influencing the perceived aesthetic value rating, bottom-up aspects of the artwork (i.e. visual information contained in the piece) influenced aesthetic perception of artworks categorised as computer-generated. These visual cues can be elucidated in terms of their image statistics. The distribution of oriented lines in the artworks (the an/isotropy of the image) predicted categorisation; more isotropic (greater distribution of orientations) images were categorised as man-made. This measure could relate to the surface

characteristics of the images which were frequently cited as a categorisation justification in the free response task. Greater isotropy in the images could result from visible brushstrokes running in multiple different directions, as well as visual forms that are more random and irregular with fewer cardinal orientations depicted. Anisotropy also differed according to the image type (representational/abstract) but this was independent of its impact on image choice, suggesting that it was not just a side effect of participants' bias to select representational works as man-made. An additional interpretation of the lack of order effects between the two tasks is that perceived aesthetic value is obtained prior to provenance categorization. While categorization is a phenomenon that occurs rapidly and automatically (Grill-Spector & Kanwisher, 2005; Greene & Fei-Fei, 2014), studies of aesthetic response also demonstrate an initially rapid stage of aesthetic response (Mullin, Hayn-Leichsenring & Wagemans, 2015; Locher et al., 2007), followed by a slower more in depth aesthetic evaluation (Leder et al., 2004). In the current situation, an initial positive aesthetic response might have lead participants to assume that the source of the artwork is human, rather than categorising it as human and then finding the piece more aesthetically favourable.

The impact of image choice on perceived aesthetic value ratings was not moderated by artistic expertise, although art-educated participants provided higher ratings for all artworks, a finding that is prevalent in previous literature (Lindell & Mueller, 2011; Leder et al., 2012; van Paasschen et al., 2015). This contrasts with the findings of Hawley-Dolan and Winner (2011), who found that aesthetic responses to professionally made artworks were moderated by the art educational background of the participants. The discrepancy between these findings could be explained by the fact that experts in Hawley-Dolan and Winner's study would have had more experience with abstract expressionist art than non-experts, whereas the experts in the current study may not have had greater exposure to computer art than non-experts.

It is important to note that these findings may be shaped by the intentionally broad definition of computer generated art used in the instructions to participants. One of the goals of this experiment was to test our participant's concepts of computer generated art and its associated visual cues. We therefore intentionally kept the definition of computer generated art as broad as possible to avoid directing participants' assumptions about the kinds of art machines could produce. It is possible that a more specific or directed definition of computer generated art would change the visual cues participants were using in their categorisations and this should be investigated in follow-up studies.

Study 2. Evaluations of robotic artwork

Given that previous findings have demonstrated an aesthetic bias against computer generated art (Kirk et al., 2009; Moffat & Kelly, 2006), it is perhaps not surprising that participants in the current study tended to prefer artworks believed to be generated by human artists.

However, the nature of what causes this bias remains unclear.

It has been previously argued that the full impact of an aesthetic experience is partially conveyed by aspects outside of the physical artwork, and is influenced by the observer's assumptions regarding the process that created it (Dutton, 2003; 2009). Kruger et al. (2004) demonstrated the *effort heuristic* in artworks, which states that if a painting appears to have taken longer to paint, it is considered more aesthetically pleasing and more valuable than one that appears to have been painted quickly and with minimal effort.

Given that typically robots and other AI systems are designed to reduce effort and labor for humans, and as no information was provided regarding the process of artistic production for either the computer or human generated art in Study 1, this combination may have falsely lead observers to assume the production of machine made art took less effort and/or was farther removed from human effort resulting in the negative bias observed. This potential

assumption due to lack of production knowledge may be functioning in much the same way that Jackson Pollock's paintings were considered unimpressive by art critics until his painstaking methods were revealed (Kruger et al., 2004). By this logic, giving participant more information as to how the artworks from *Study 1* were created may alleviate this bias against computer-generated art.

In addition to how knowledge of the process may influence the aesthetic judgments of computer generated art, it is important to consider the embodiment (physical presence) of an artificial system in creating visual art. Previous work on human-robot-interactions has revealed that the embodiment of the system affects the perception of the output it generates. A relevant example by Ogawa et al. (2012) compared audience responses to poetry read by a humanoid robot to a box with a speaker inside it. Their findings revealed that watching the humanoid robot intensified the impact of the poem for the audience compared to the box. This demonstrates how embodiment, even within artificial agents, can generate a different level of artistic experience. Moreover, painting/drawing in the physical world differs greatly from digital environments given the natural constraints the physical world places on the artist such as velocity, pressure, amount of paint on the brush, etc. All of these physical constraints must be carefully considered by human and robotic artists as they can directly affect the aesthetic appeal of the final product (Tresset & Deussen, 2014).

Given that in *Study 1*, participants cited surface qualities and brush strokes as informative features for their source categorisation, we opted to investigate the relationship between aesthetic judgment and artistic value on artworks created by artistically skilled embodied agents. In brief, we collected drawings from robotic portrait artists and displayed them under three different viewing conditions in a between-groups design. In *condition 1*, observers were in the presence of the robots as they created the drawings and could interact with them by

having their own portrait drawn. *Condition 2* consisted of displaying the drawings and telling observers that they were created by robots but not allowing them to see or interact with the robots. Finally, *condition 3* consisted of displaying the drawings and giving observers no information as to how they were created. In all three conditions participants filled out a questionnaire detailing their aesthetic response and artistic value of the drawings. Based on the results of *Study 1* and the existing literature regarding computational creativity and human-robot-interactions, we predicted that informing participants of the drawings' true source would negatively affect the perceived aesthetic and artistic value compared to the condition where no source knowledge was given. However, given the opportunity to see, and interact with, the robots, we hypothesized that this negative bias would be reduced or even reversed.

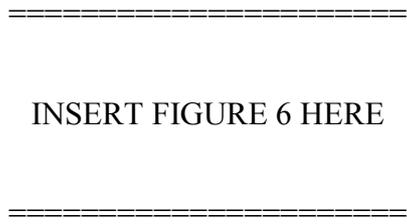
Methods

Stimuli: The robots and the drawings

Study 2 used portrait drawings from a single source, the installation of “5 Robots named Paul”, created by artist/roboticist Patrick Tresset¹ and displayed at the BOZAR centre for fine arts in Brussels, Belgium from April 30th to May 4th, 2015. These robots, collectively known as Paul, are each composed of a left-handed planar robotic arm, with a black Biro pen as an end-effector and an actuated pan and tilt webcam attached to a vertical wooden stick bolted to an old school desk that acts as a body. The robots are positioned around a chair where the subject is seated, much like a real drawing class (see Figure 6). A broad overview of the drawing cycle involves: localizing the face of the sitter, taking a picture and converting it to a

¹ Patrick Tresset is an artist who creates theatrical installations with robotic agents as actors. Tresset's installations use computational systems that aim to introduce artistic, expressive and obsessive aspects to robots' behaviour. Tresset also develops and uses robots to produce series of drawings, paintings and animations. Tresset's work is internationally exhibited including in association with major institutions such as Pompidou center (France), Victoria and Albert Museum (UK), Tate Modern (UK), MMCA (KR), Museum of Israel (IL), Foundation Prada (IT). Drawings produced by his robots are in public collections (Prada fondation, Maison d'Ailleurs, Victoria and Albert Museum), in a number of private collections and owned by hundreds of individuals.

grey-level imaging and applying a contrast curve, drawing salient lines, performing shading behaviour and executing the signing script (for full details on Paul, see Tresset & Leymarie, 2013). In addition to the drawing cycles, these robots also perform a number of behaviours that, while unnecessary for the creation of the drawing, are implemented as theatrical effects to give the impression to the sitter and audience that the robots are somewhat "alive". For example, it would appear that the robots are looking at the sitter, scanning their face with saccades and then examining the progress of their drawing or the eye following the hand's movements in the same manner as a human drawing would, when in fact the drawing is based on a single picture taken at the initial state of the cycle.



When the drawings are complete for a given sitter, they are displayed on the walls around or behind the installation (see Figure 7). Each of the five robots exhibits their own idiosyncratic behaviours that manifest in their drawings. This is achieved on a software level, by controlling the execution and speed of the arm movements to simulate varying levels of nervousness. Some observers also project different traits onto the robots, for example the robot with the fastest movements is often qualified as "nervous", "annoyed" or "a bit mad" by members of the audience. As a result, it would appear that each robot has its own distinctive style that can be distinguished from the others.

Condition 1: Interaction. 145 participants took part in the *Interaction condition* (Age: M=37.81, SD=14.27; 64 female). The data collection for this condition took place over four days at the BOZAR Centre for Fine Arts in Brussels, Belgium. During this time observers

were free to examine the robots and their drawings as they wished for as long as they liked. During this time they were approached by one of the researchers and asked if they would be interested in filling out a questionnaire regarding their impressions on the drawings and the robots. The questionnaire consisted of basic demographic information, as well as questions regarding the participant's art education and interest in art. Then participants answered a series of questions on their aesthetic response to the drawings, their impressions of the artistic value of the drawings and their impressions of the robots themselves on Likert-type scales and semantic differential scales (see Appendix 3). There was also a free response section where participants could add any additional comments regarding their experiences. The questionnaire was available in Dutch, French, and English.

Condition 2: Source Information. 97 participants took part in the *Source Information condition* (Age: M=26.97, SD=13.25; 49 female). Data collection for this condition took place over one week in mid-August, 2015 in the gallery space of the KU Leuven Library in Leuven, Belgium. A grid of 4 (identities) x 5 (robots) drawings produced by the 5 Robots Named Paul during their residency at Bozar was used as stimuli for this condition (similar to those shown in Figure 7). Note that the robots themselves were not present during this condition as its purpose was to assess the perception of the drawings when participants knew that they were created by an artificial source but were not given the chance to observe or interact with them. Observers were free to examine the drawings for as long as they liked. During this time they were approached by one of the researchers and asked if they would be interested in filling out a questionnaire. The questionnaire was prefaced with information about the source of the drawings:

“The artworks are made by a robot who draws portraits. The robot has no semantic knowledge of what a face is. The robot starts each drawing by orienting a camera on the subject's face. It then takes a picture of the face and

converts it to a grayscale image. It uses an edge detection technique similar to the kinds of processes we have in the visual areas of our brain to pick out salient lines on the face at many different spatial scales. A robotic arm with a biro then goes to work creating the drawing based on the salient lines that have been selected. The robot then takes another picture of the subject for comparison to the drawing in order to refine it and add shading. Through this feedback, the robot is able to draw using the equivalent of an artist's stylistic signature based on a number of processes mimicking drawing skills and technique.”

The rest of the questionnaire was the same as the *Interaction condition* but did not contain any questions regarding participant’s impressions of the robots. It was available in both Dutch and English.

=====

INSERT FIGURE 7 HERE

=====

Condition 3: No Source Information. 107 participants took part in the *No Source Information condition* (Age: M=26.13, SD=10.64; 53 female). This condition was identical to Condition 2, however the questionnaire did not contain any information regarding the source of the drawings. Given no source information, participants were likely to assume that the drawings were made by human sources.

Demographic differences between conditions.

As the samples for the three experimental conditions were drawn from two different locations, it is likely that there were differences in age, gender, and art experience between the conditions. A chi-squared test revealed no significant differences in gender distribution between the three subsamples, $\chi^2(2) = 0.67, p=0.71$. There were significant differences in age

between the three subsamples, $F(2, 334) = 31.21, p < .001, \eta_p^2 = .20$ (Interactive: $M = 37.81$; Source Information $M = 26.97$; No source information: $M = 26.13$) driven by a difference between the interactive condition and the other two conditions (source information and no source information). Art experience and interest is known to impact on aesthetic attitudes (Lindell & Mueller, 2011; Leder et al., 2012; van Paasschen et al., 2015) and therefore differences in these characteristics could account for differences in aesthetic ratings between the experimental conditions rather than the experimental manipulation itself. To explore this possibility, a demographic analysis was conducted for the three experimental samples. To calculate the prevalence of art education in our samples, all participants who had responded yes to the question ‘did you receive an art education?’ in the questionnaire were labelled as art-educated participants (Appendix 1). There were more art-educated participants in the interactive condition ($n = 56$) than in the source information ($n = 20$) and the no source information condition ($n = 18$). To calculate interest in art, participants had to have reported that they visit art galleries 2-3 times per year or more and that they have more than 10 books on art. There were more art-interested participants in the interactive condition ($n = 57$) than in the source information ($n = 1$) and the no source information condition ($n = 0$). The difference in distribution of art education and interest was significantly different between the conditions (art education: $\chi^2(2) = 17.59, p < .001$; art interest: $\chi^2(2) = 91.68, p < .001$).

Results

Factor analysis of response to robot drawings. Eleven questions posed to all participants concerning their response to the drawings were analysed using Principal Components Analysis (PCA) with varimax rotation. Responses to the statement ‘the drawings are hand-crafted’ were omitted from the PCA as they did not correlate strongly with any other responses (Table 4).

INSERT TABLE 4 HERE

On inspection of the scree plot, PCA analysis implied a one component solution which explained 35% of the variance. The factor loadings of the questions onto the primary component are presented in Appendix 4. As all factor loadings were in a small range (0.23-0.36) and for ease of interpretation of the results, mean ratings for questions included in the PCA were calculated to produce average perceived aesthetic value ratings for each participant. This variable was then used as the dependent variable in a between-groups ANOVA, assessing the impact of condition (interaction/no source information/source information) on perceived aesthetic value (Figure 8). There was a significant impact of condition on compiled ratings of aesthetic value, $F(2,346) = 42.16, p < .001, \eta_p^2 = .20$. Post-hoc Tukey tests revealed that there was a significant difference in aesthetic value between the interactive ($M = 3.58$) and the no source information ($M = 3.19$) conditions at $p < .001$, 95% CI [0.24, 0.53]. There was a significant difference between the no source information condition and the source information ($M = 3.02$) condition, $p < .05$, 95% CI [0.009, 0.33]. Finally, there was a significant difference between the interactive and the source information conditions, $p < .001$, 95% CI [0.41, 0.71]. There was no significant correlation between age of the sample and perceived aesthetic value in the drawings, $r(335) = -0.07, p = 0.18$, suggesting that even though there were age differences between the subsamples, that this did not drive the differences between conditions.

INSERT FIGURE 8 HERE

Analyses for participants interested and educated in visual arts. As there are known differences in aesthetic ratings provided by art experts and non-experts (Lindell & Mueller, 2011; Leder et al., 2012; van Paasschen et al., 2015) and given the difference in level of art experience in the three sub-samples in the current study (see Method: Demographic differences between conditions), a series of analyses were conducted to assess the impact of interest and education in art on the assessment of the robot drawings. First, between-groups *t*-tests were performed to determine if there were differences between individuals with interest and education in the arts and those without for aesthetic ratings (Table 5). Those participants who were interested in the visual arts provided higher aesthetic ratings for the robot drawings. There was no significant difference in perceived aesthetic value for art-educated participants.

INSERT TABLE 5 HERE

As there was a difference in the number of participants with an interest in art in each experimental condition and these participants gave higher aesthetic ratings, the results of the repeated measure ANOVA (Figure 8) could be due to the presence of more art-interested individuals in the interactive condition. Therefore, the ANOVA was performed again omitting the art-interested participants from the dataset (it was not possible to conduct a 3 X 2 between-subjects ANOVA as the number of participants interested in the arts in the no source information and source information conditions was too low). The pattern of data was the same in this case: condition had a significant impact on perceived aesthetic value ratings, $F(2,$

284)=30.90, $p<.001$, $\eta_p^2=.18$, post-hoc Tukey tests revealed there was a significant difference between the interactive condition ($M=3.55$), and the source information ($M=2.99$), $p<.001$, 95% CI [0.39,0.73], between the interactive and the no source information conditions ($M=3.20$), $p<.001$, 95% CI [0.18, 0.51], and between the source information and no source information conditions, $p<.01$, 95% CI [0.05,0.37]².

Analysis of data in the interactive condition. Participants in the interactive condition completed extra questions concerning their responses to the robots themselves in relation to common Human Robot Interaction (HRI) metrics and questions concerning the artistic process which were specifically designed for this study (Figure 9). As can be seen from Figure 9a, participants responded most positively to questions pertaining to the robots drawing from observation and possessing their own individual style. By contrast, participants were unsure whether the robots were creative or were the authors of the artworks. Responses to the HRI scales (Figure 9b) also reveal interesting dichotomies. Observers reported that they thought that the robots were machine-like and artificial while appearing intelligent.

=====

INSERT FIGURE 9 HERE

=====

Correlations between responses to questions about the robots were then computed to investigate overall data structure (Table 6). The questions in the HRI section generally correlated well with one another, and correlated highly with whether participants thought the

² One reviewer observed that the criteria for the art-interested participants were quite strict (they must own >10 art books) and may introduce a socio-economic confound in the grouping variable as art books are sometimes prohibitively expensive. However, if the criteria for art-interested participants is widened to include participants who report having 5-10 art books the pattern of results remained the same with respect to the main effect and post-hoc comparisons but in this case perceived aesthetic value did not differ between the art-interested and the non-art interested participant groups.

robots were being creative. Perceptions of whether the robots had an individual style, were drawing from observation and were the authors of the drawings were to a large extent independent of other measures. On this basis, a composite score for the HRI measures was calculated by taking the mean across questions, corresponding broadly to a measure of perceived ‘anthropomorphism’. The artistic process questions were treated independently due to their low correlations with one another.

=====

INSERT TABLE 6 HERE

=====

Responses to the HRI and artistic process questions were then correlated with the perceived aesthetic value of the robot drawings. Table 7 and Figure 10 show that anthropomorphism correlated positively with the perceived aesthetic value of the robot drawings. Aesthetic value was also correlated with whether participants felt the robots were drawing from observation, whether each had an individual style, whether they were creative, and whether they were the authors of the drawing. Anthropomorphism was not significantly correlated with whether participants thought the robots were drawing from observation.

=====

INSERT TABLE 7 HERE

=====

=====

INSERT FIGURE 10 HERE

=====
Anthropomorphism did not differ significantly between art-educated (M=2.81) and non-art-educated (M=2.81), and art-interested (M=2.73) and non-art-interested (M=2.86) participants.

Discussion

Study 2 has elaborated upon the finding that computer-generated art is not afforded the same aesthetic status as man-made art, demonstrating that this bias can be moderated by interaction with the agents of the artwork, in line with a previous study involving computer-generated poetry (Ogawa et al., 2012). The study involved three participant samples, only one of which was able to view the process of artistic creation; in this case the artists took the form of robotic arms with camera ‘eyes’ attached. The presence of the robotic artists had a strong positive impact on aesthetic evaluations of the resulting artworks. This interactive condition elicited greater aesthetic responses compared to conditions in which participants viewed only the final artworks and were either told the work was made by robots or given no source information. A difference was also found in aesthetic responses to the robot drawings in the no source information condition relative to the source information condition. This finding supports the findings of *Study 1* as well as previous work (Kirk et al., 2009; Moffatt & Kelly, 2006) in which knowledge that an artwork is computer-generated impacts negatively upon aesthetic ratings. Here the difference in ratings in the two conditions cannot be explained by surface properties of the images as exactly the same images were shown to participants in the source information and no source information conditions.

Analysis of responses to the robots in the interactive condition of this study revealed interesting conceptual dichotomies. The robots were generally perceived as being intelligent but were not perceived as being human or life-like. The robots in Patrick Tresset’s installation are deliberately created and displayed in a way that does not provide them with

humanoid visual characteristics, contrasting them with other drawing robots (Calinon, Epiney, & Billard, 2005). Therefore, it is predictable that the robots are described as machine-like. However, some participants noted that the installation did have the feel of a drawing class with intelligent agents systematically observing and recording. This sense of examining the collective process of the robots, moving independently and seemingly evaluating their work is exemplified in the strong positive response to the perceived intelligence scale. This data illustrates that participants in the current study acknowledge a contrast between the outward appearance of the robot (machine-like) and the embodiment of an artistic agent (intelligent). This may explain why anthropomorphism correlated positively with aesthetic response in this study. In this case anthropomorphic characteristics were not surface level but were more concerned with the dynamics of the robotic actions and their implied intentional states. For instance, the robot's "looking behaviour" provided the observer with a subjective insight into the robots "intentions". Appropriate looking is crucial for robots to be accepted by the human they are interacting with (Andrist et al., 2014). Future investigations into the appreciation of machine made art should explore role of anthropomorphism through comparison with more humanoid robots.

In terms of artistic process, the participants did not think that the robots were creative or that they were the authors of the drawings. This can be set in the context of Newman and Bloom's (2011) proposition that the value of an art object depends upon whether it is seen as a unique creative act and is supported by the fact that aesthetic ratings for the robot drawings were positively correlated with responses to these two questions. This represents a remaining barrier to valuing a robotic work of art on the same level as a man-made one. It was not possible in the current study to contrast watching a live artist drawing against watching a robot drawing but a follow-up investigation including a human drawing condition would be

able to test the contention that perceived creativity and ownership is specific to human art and significantly affects the aesthetic experience.

While in general the participants did not describe the robots as human-like, the degree to which they anthropomorphised the robots correlated with perceived aesthetic value. This is the first study to demonstrate the anthropomorphism of an agent impacts positively on aesthetic appraisal. This finding suggests that perceived intentionality does play a role in the aesthetic appreciation of artworks, in line with previous findings in the domain of child and animal art (Hawley-Dolan & Winner, 2011; Snapper et al., 2015). The impact of anthropomorphism could be investigated further in a controlled experimental paradigm in which the experimenter could modify aspects of the robots' behaviour (like the frequency of the camera movements between the subject and the drawing and the fluidity of the robotic hand movements) to assess the impact this has on aesthetic response. It is important to note that the anthropomorphic behaviours of these robots did not, in fact, provide any insight into the artistic production process. Additional follow up work should investigate whether the legitimacy of the anthropomorphic cues effects aesthetic response. The degree of anthropomorphism perceived in the robots did not differ in the subgroups of participants that had an art education or were interested in the visual arts. This suggests that while participants interested in art liked the robotic artworks more than those less interested in art, this was not due to the degree to which they perceived humanlike characteristics in the robotic artists.

A limitation to Study 2 was the inability to test all three participant groups in exactly the same setting. Due to the constraints of the installation at BOZAR and the researchers access to other comparable institutions, the interaction condition was tested at an art gallery and the knowledge and no-knowledge conditions were tested in a gallery section of the university library. This difference in experimental setting coupled with the researchers

approaching potential participants to fill out the survey could have resulted in participant selection bias. As might be expected, we found more art-interested participants at the gallery than at the library. However, after controlling for the number of art-interested participants in the interaction condition the effect of context was unchanged. Therefore, while the selection of participants was a limitation of this study, the data demonstrate the same pattern of results remained after controlling for the artistic engagement of the participant samples such that interacting with the robotic agents ameliorated the bias to their artworks regardless of how engaged with art the observers were.

Conclusion

These studies have sought to investigate how observers respond to works of art known to be generated by a computer or robotic agent. The findings indicate that observers are influenced by surface properties of computer-generated artworks (the degree to which they look hand-crafted and imperfect) as well as cognitive biases about the negative value of computer-generated art and its ability to function as compelling representational work. *Study 1* found evidence for both of these contributions to the negative bias for computer art, but future studies incorporating a larger set of predominantly abstract works of art may be able to elaborate on these issues to understand when and where they come into play. In addition to surface properties and cognitive biases, aesthetic responses to artworks are now known to be affected by the embodiment of the system (*Study 2*). Viewing the creation of a work of art by a robot increased aesthetic appreciation for it by giving observers an insight into the complex creation process and physical constraints of the system. In addition, observers' aesthetic responses to robotic art were moderated by how anthropomorphic the robotic agent appears to the observer. On the basis of these findings, it can be suggested that increasing the anthropomorphic qualities of robotic and computational art will increase societal engagement and likely decrease hostility toward future manifestations of artistic AI. In addition, improved

technology for rendering the dynamics of human movement will produce more convincing artworks for observers to aesthetically engage with. It has become clear that computational art has not yet passed the Lovelace test, both objectively and in the eye of the beholder.

Perceived creativity in the artistic agent may indeed represent the final frontier for the true acceptance of computer art. Until then, computational art may still be regarded as merely an artificial form of one of the most human modes of expression.

References

- Alvarez-Ramirez, J., Ibarra-Valdez, C., Rodriguez, E., & Dagdug, L. (2008). 1/f-Noise structures in Pollocks's drip paintings. *Physica A: Statistical Mechanics and its Applications*, 387(1), 281-295.
- Amirshahi, S. A., Koch, M., Denzler, J., & Redies, C. (2012). PHOG analysis of self-similarity in esthetic images. *Proc. SPIE (Human Vision and Electronic Imaging XVII)*, 8291, 82911J.
- Amirshahi, S. A., Redies, C., & Denzler, J. (2013,). How self-similar are artworks at different levels of spatial resolution?. In *Proceedings of the Symposium on Computational Aesthetics* (pp. 93-100). ACM.
- Andrist, S., Tan, X. Z., Gleicher, M., & Mutlu, B. (2014, March). Conversational gaze aversion for humanlike robots. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 25-32). ACM.
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta- analysis of the relationship between the five- factor model of personality and Holland's occupational types. *Personnel Psychology*, 56(1), 45-74.
- Bartneck, C., Croft, E., Kulic, D. & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1, 71-81.
- Bill, M. (1993). The mathematical way of thinking in the visual art of our time. In *The visual mind* (pp. 5-9). MIT Press.

- Bringsjord, S., Bello, P., & Ferrucci, D. (2003). Creativity, the Turing test, and the (better) Lovelace test. In *The Turing Test* (pp. 215-239). Springer: Netherlands.
- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Braun, J., Amirshahi, S. A., Denzler, J. & Redies, C. (2013) Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, 4.
- Calinon, S., Epiney, J. & Billard, A. (2005). A Humanoid Robot Drawing Human Portraits. *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids), Tsukuba, Japan, 161-166.*
- Cohen, H. (1973) Parallel to perception: some notes on the problem of machine-generated art. *Computer Studies*, 4.
- Darling, K. (2015). “‘Who’s Johnny?’ Anthropomorphic Framing In Human-Robot Interaction, Integration, and Policy.” *We Robot 2015 Proceedings*. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2588669.
- Deussen, O., Lindemeier, T., Pirk, S., & Tautzenberger, M. (2012, June). Feedback-guided stroke placement for a painting machine. In *Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging* (pp. 25-33). Eurographics Association.
- Dutton, D. (2003). Authenticity in art. In J. Levinson (Ed.), *The Oxford handbook of aesthetics* (pp. 258–274). New York, NY: Oxford University Press.

- Dutton, D. (2009). *The art instinct: Beauty, pleasure and evolution*. New York, NY: Bloomsbury Press.
- Fayn, K., MacCann, C., Tiliopoulos, N., & Silvia, P. J. (2015). Aesthetic Emotions and Aesthetic People: Openness Predicts Sensitivity to Novelty in the Experiences of Interest and Pleasure. *Frontiers in Psychology*, 6.
- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11(5), 197-203.
- Forsythe, A., Nadal, M., Sheehy, N., Cela- Conde, C. J., & Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art. *British Journal of Psychology*, 102(1), 49-70.
- Fuchs, C. Schreier, M., van Osselaer, S. M. J. (2015) The handmade effect: what's love got to do with it? *Journal of Marketing*, 79, 98-110.
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, 35(4), 1674-1684.
- Graham, D. J. & Field, D. J. (2007). Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spatial Vision*, 21, 149-164.
- Greene, M. R., & Fei-Fei, L. (2014). Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *Journal of vision*, 14(1), 14-14.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition as soon as you know it is there, you know what it is. *Psychological Science*, 16(2), 152-160.

- Grosser, B. (2011, Jan, 20). What are art/music machines made for? Retrieved from <http://bengrosser.com/blog/what-are-art-music-machines-for/>
- Hawley-Dolan, A. & Winner (2011). Seeing the Mind Behind the Art: People Can Distinguish Abstract Expressionist Paintings From Highly Similar Paintings by Children, Chimps, Monkeys and Elephants. *Psychological Science*, (22), 435-441.
- Hawley-Dolan, A. & Young, L. (2013). Whose Mind Matters More: The moral agent or the artist? The role of intent in ethics and aesthetics. *PLoS One*, 8(9).
- Isenberg, T., Neumann, P., Carpendale, S., Costa Sousa, M. & Jorge, J. A. (2006) Non-photorealistic rendering in context: an observational study. *Proceedings of the 4th international symposium on Non-photorealistic animation and rendering*, Annecy, France.
- Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009) Modulation of aesthetic value by semantic context: an fMRI study. *NeuroImage*, 44 (3), 1125-1132.
- Koch, M., Denzler, J., & Redies, C. (2010). 1/f² characteristics and isotropy in the Fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. *PLoS One*, 5(8), e12268.
- Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, 40(1), 91-98.
- Leder, H., Bar, S. & Topolinski, S. (2012) Covert painting simulations influence aesthetic appreciation of artworks. *Psychological Science*, 23 (12), 1479-1481.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489-508.

- Locher, P., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, 21(1), 55-77.
- Lindell, A. K., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology*, 23(4), 453-475.
- Maciejewski, R., Isenberg, T., Andrews, W. M., Ebert, D. S., & Sousa, M. C. (2007). Aesthetics of Hand-Drawn vs. Computer-Generated Stippling. In *Computational Aesthetics* (pp. 53-56).
- Mallon, B., Redies, C., & Hayn-Leichsenring, G. U. (2014) Beauty in abstract paintings: perceptual contrast and statistical properties. *Frontiers in Human Neuroscience*, 8.
- McCorduck, P. (1990). *Aarons Code*. New York, NY: W.H.Freeman & Co Ltd.
- McManus, I. C., & Furnham, A. (2006). Aesthetic activities and aesthetic attitudes: Influences of education, background and personality on interest and involvement in the arts. *British Journal of Psychology*, 97(4), 555-587.
- Melmer, T., Amirshahi, S. A., Koch, M., Denzler, J., & Redies, C. (2013). From regular text to artistic writing and artworks: Fourier statistics of images with low and high aesthetic appeal. *Frontiers in Human Neuroscience*, 7.
- Moffat, D. & Kelly, M. (2006) An investigation into people's bias against computational creativity in music composition. *Presented at The Third Joint Workshop on Computational Creativity (ECAI'06), Riva del Garda, Italy*.
- Mullin, C., Hayn-Leichsenring, G., & Wagemans, J. (2015). There is beauty in gist: An investigation of aesthetic perception in rapidly presented scenes. *Journal of vision*, 15(12), 123-123.

- Nake, F. (2005) Computer art. A personal recollection. *Proceedings of C&C'05*, April 12-15, London, UK.
- Newman, G. E. & Bloom, P. (2011) Art and authenticity: the importance of originality in judgments of value. *Journal of Experimental Psychology: General*, 141(3), 558-569.
- Noll, A. M. (1966). Human or machine: A subjective comparison of Piet Mondrian's "Composition With Lines"(1917) and a computer-generated picture. *The psychological record*.
- Ogawa, K., Taura, K., & Ishiguro, H. (2012, September). Possibilities of Androids as poetry-reciting agent. *In RO-MAN, 2012 IEEE* (pp. 565-570). IEEE.
- Redies, C., Amirshahi, S. A., Koch, M., & Denzler, J. (2012, October). PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. *In Computer Vision–ECCV 2012. Workshops and Demonstrations* (pp. 522-531). Springer Berlin Heidelberg.
- Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: similarity to natural scenes. *Spatial Vision*, 21(1), 137-148.
- Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009, March). How anthropomorphism affects empathy toward robots. *In Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 245-246). ACM.
- Sbriscia-Fioretto, B., Berchio, C., Freedberg, D., Gallese, V., & Umiltà, M. A. (2013). ERP modulation during observation of abstract paintings by Franz Kline. *PloS One*, 8(10), e75241.

- Snapper, L., Oranc, C., Hawley-Dolan, A., Nissel, J. & Winner, E. (2015) Your kid could not have done that: even untutored observers can discern intentionality and structure in abstract expressionist art. *Cognition*, 137, 154-165.
- Taylor, R. P. (2002). Order in Pollock's chaos. *Scientific America*, 116–121.
- Taylor, R. P. (2011). Physics, perception, and physiological of Jackson Pollock's fractals. *i-Perception*, 2(4), 284-284.
- Taylor, R. P., Spehar, B., Van Donkelaar, P., & Hagerhall, C. M. (2014). Perceptual and physiological responses to Jackson Pollock's fractals. *Brain and Art*, 43.
- Ticini, L. F., Rachman, L., Pelletier, J., & Dubal, S. (2014). Enhancing aesthetic appreciation by priming canvases with actions that match the artist's painting style. *Frontiers in Human Neuroscience*, 8.
- Tresset, P., & Deussen, O. (2014). Artistically skilled embodied agents. *Proceedings of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB-50)*, London, UK.
- Tresset, P., & Leymarie, F. F. (2013). Portrait drawing by Paul the robot. *Computers & Graphics*, 37(5), 348-363.
- Umilta, M. A., Berchio, C., Sestito, M., Freedberg, D., & Gallese, V. (2012). Abstract art and cortical motor activation: an EEG study. *Frontiers in Human Neuroscience*, 6, 311.
- van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The Influence of Art Expertise and Training on Emotion and Preference Ratings for Representational and Abstract Artworks. *PloS One*, 10(8), e0134241.

Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: highly individual taste for abstract, but not real-world images. *Journal of Vision*, *10*(2), 18-18.

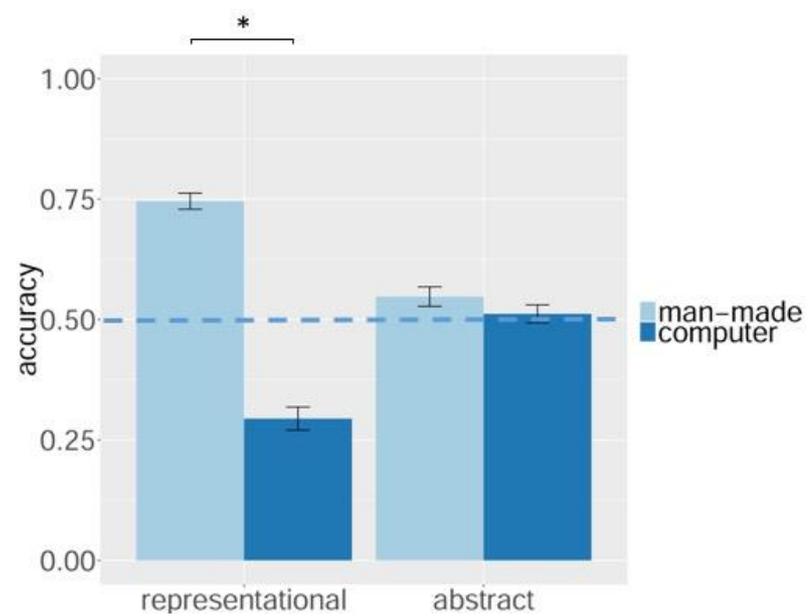


Figure 1. Mean accuracy by image source (computer/man-made) and type (abstract/representational). Error bars represent +/- 1 standard error of the mean. Dotted line represents 50% accuracy.

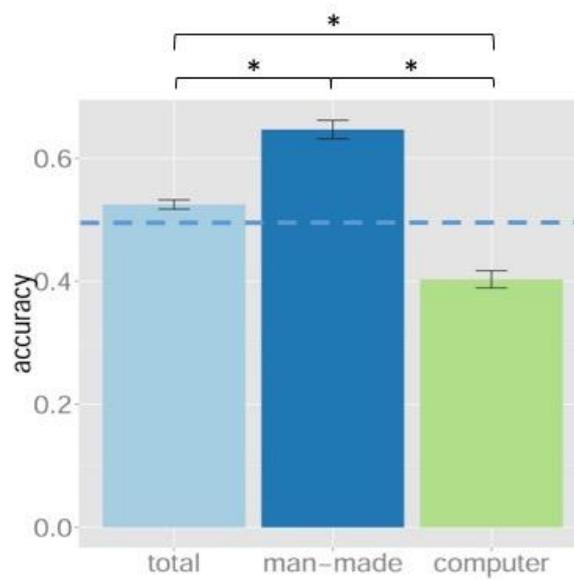


Figure 2. Overall accuracy across image types and accuracy for ground truth man-made and computer-generated image categories. Error bars represent +/- 1 standard error of the mean. Dotted line represents 50% accuracy.

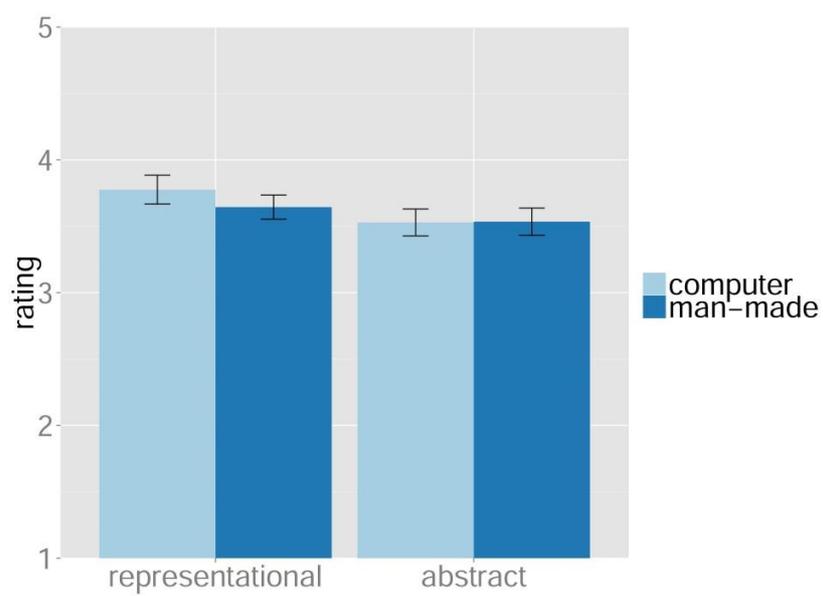


Figure 3. Impact of image type and source on aesthetic ratings. Error bars represent +/- 1 standard error of the mean.

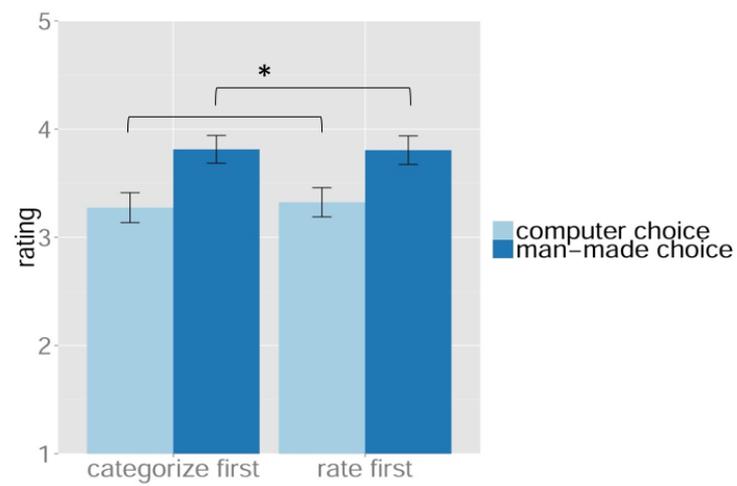


Figure 4. Impact of image choice on aesthetic rating in 'rate first' and 'categorise first' participant subgroups. Error bars represent +/- 1 standard error of the mean.

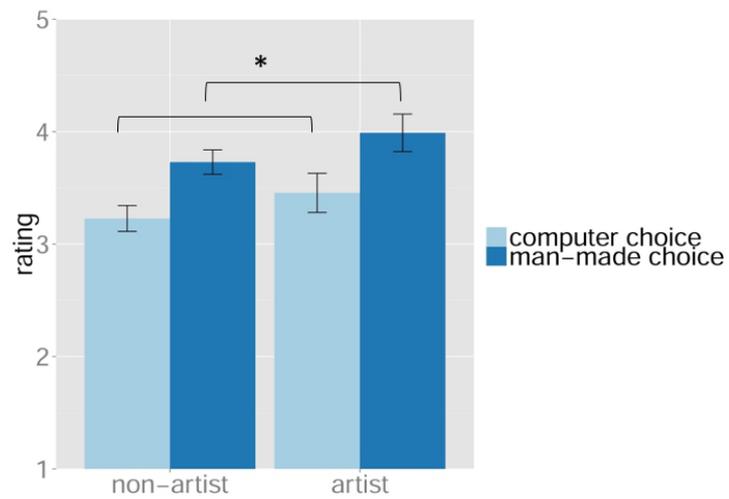


Figure 5. Impact of image choice on aesthetic ratings for art-educated and non-art educated participants. Error bars represent +/- 1 standard error of the mean.



Figure 6. The installation “5 Robots Named Paul” at the Bozar Centre for Fine Arts in Brussels, Belgium.



Figure 7. Example images displayed during both Condition 2: Source information and Condition 3: No Source Information. The images used in the study were created by the installation “5 Robots Named Paul” but did not correspond specifically to these identities.

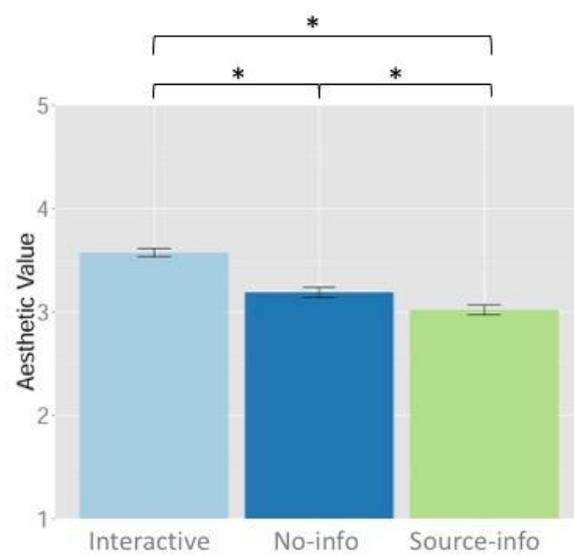
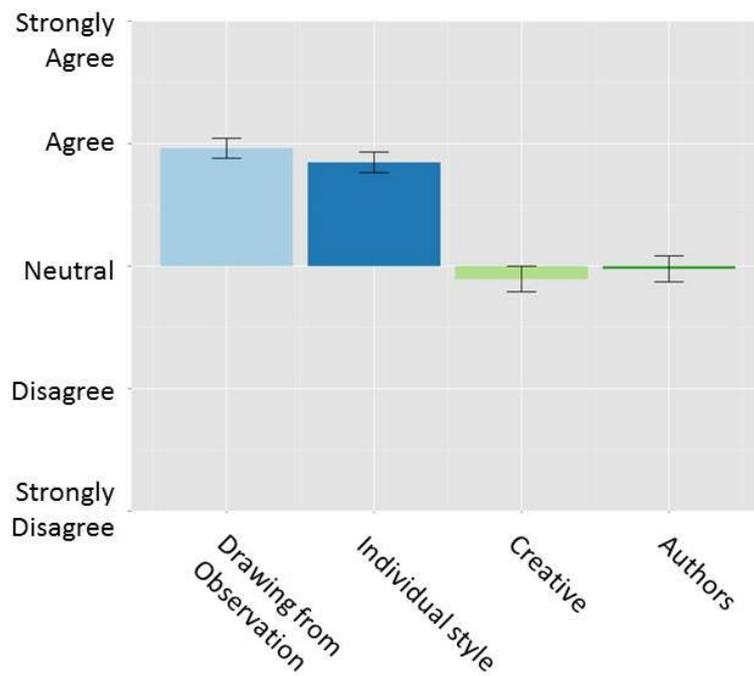
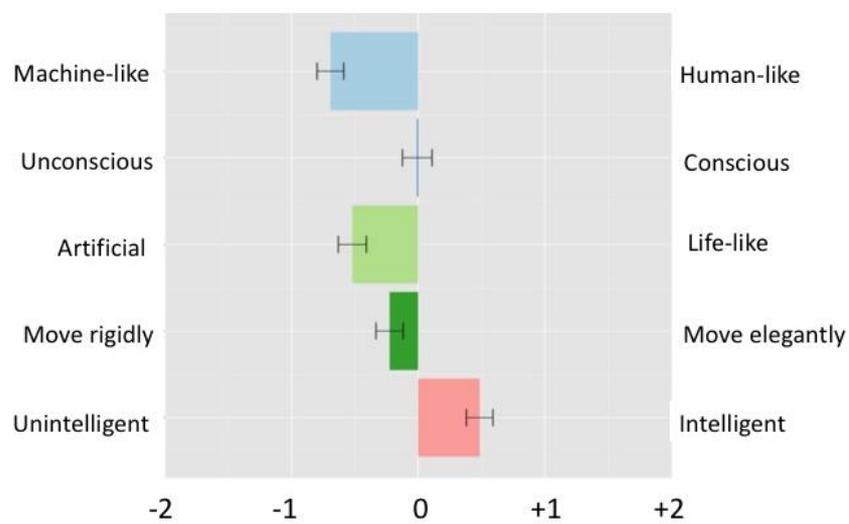


Figure 8. The impact of condition on mean aesthetic ratings of robotic drawings. Error bars represent +/- 1 standard error of the mean.



a.



b.

Figure 9. Response to questions pertaining to the a) artistic process of the robots b) the anthropomorphic properties of the robots. Error bars represent +/- 1 standard error of the mean.

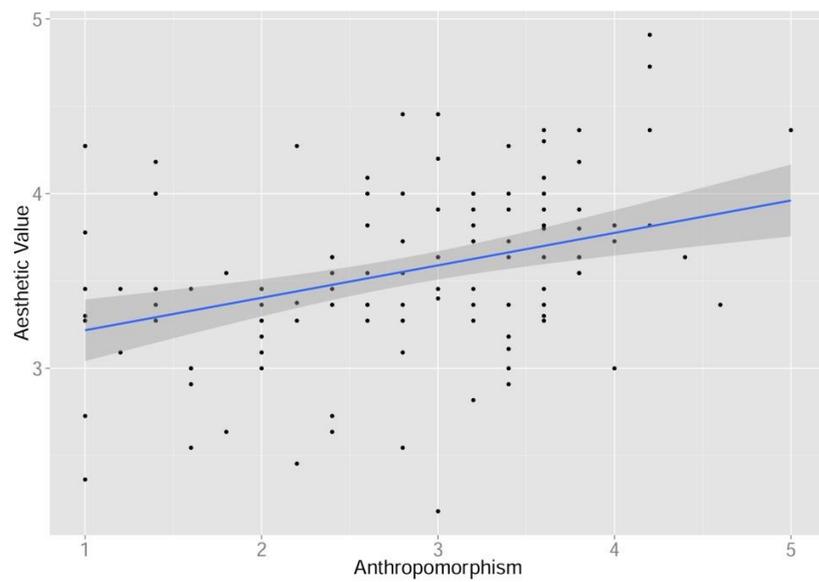


Figure 10. The relationship between perceived anthropomorphism and aesthetic rating. Shaded area represents 95% CI of linear regression line.

Table 1. Difference in image statistics for abstract/representational and computer-generated/man-made images.

	Abstract/Representational		Computer-generated/Man-made	
	95% CI of Mean Difference	<i>t</i> -test	95% CI of Mean Difference	<i>t</i> -test
Self-similarity	-0.04, 0.08	0.58	-0.90, 0.03	-0.91
Complexity	-1.69, 6.83	1.21	-6.95, 1.57	-1.27
Anisotropy	0.00007, 0.0003	3.08*	-0.0002, 0.00008	-0.91
Fourier Slope	-.010, 0.60	1.94	-0.42, 0.20	-0.71

Notes: * $p < .006$ (Bonferroni corrected for multiple comparisons)

Table 2. Correlation between image categorisation, aesthetics, accuracy and image statistics.

	Mean aesthetic rating per image	Proportion of trials in which image was categorised as man-made	Mean accuracy per image
Self-similarity	0.04	0.24	0.13
Complexity	-0.08	-0.01	-0.01
Anisotropy	-0.25	-0.43*	0.02
Fourier Slope	-0.20	-0.16	0.06

Notes: * $p < .001$

Table 3. Thematic analysis of participants' free response when asked how they made their categorisation judgements

Theme	Subtheme	Computer-generated		Man-made	
		Keywords	Frequency	Keywords	Frequency
Surface	Colour	Bright/Artificial	13	Natural/Varied	8
	Line & Shape	Rigid/Perfect/Straight	30	Imperfect	8
	Mark-making	Not hand-made	5	Hand-made/brush strokes	37
Structure	Regularity	Uniform/Symmetrical	6	Random/Irregular	8
	Repetition	Identical/Repetitive	6	Different forms	2
	Complexity	Detailed/Complex	7	Simple	2
	Depth	Lack of depth	1	Depth	2
Content	Abstraction	Abstract	4	Figurative	2
Intentionality	Emotion	Lack of expression	2	Feeling/expression	3
	Cognition	Lack of reason	1	Reason	2
	Uniqueness	Uninspired	1	Unique	2

Table 4. Correlations between responses to robotic drawings

	2	3	4	5	6	7	8	9	10	11	12
1. Like drawings	.26*	.46*	.17	.12	.33*	-.07	.42*	.43*	.26*	.18	.24*
2. Fetch money	-	.27*	.30*	.25*	.27*	.01	.32*	.24*	.26*	.14	.12
3. Visually pleasing		-	.25*	.26*	.31*	.03	.46*	.46*	.38*	.11	.23*
4. Clear intentions			-	.12	.16	.08	.21*	.21*	.33*	.19	.19
5. Time and effort				-	.28*	.14	.28*	.21*	.21*	.21*	.18
6. Appreciated					-	.15	.40*	.34*	.31*	.28*	.27*
7. Hand-crafted						-	.09	-.04	.16	.06	-.08
8. Willing to pay							-	.39*	.32*	.21*	.25*
9. Interesting								-	.33*	.27*	.39*
10. Objective quality									-	.19	.19
11. Creative										-	.31*
12. Thought-provoking											-

Notes: N=353, *Bonferroni corrected $p < .0008$

Table 5. Aesthetic ratings grouped by art interest and art education with significance tests for difference between the group means.

		Mean	SD	<i>t</i> -value	95% CI	Cohen's D	p-value
Art- interested	No	3.24	.53	5.10	[0.23,0.53]	.55	<.001
	Yes	3.61	.45				
Art- educated	No	3.28	.53	1.28	[-0.04,0.21]	.14	.20
	Yes	3.36	.57				

Table 7. Correlations between responses to HRI and artistic process questions

	2.	3.	4.	5.	6.	7.	8.	9.
1.Human-like	.49*	.53*	.39*	.26	.04	.18	.30*	.17
2.Conscious	-	.55*	.24	.59*	.17	.13	.32*	.25
3.Life-like		-	.32*	.42*	.14	.24	.40*	.18
4.Moving elegantly			-	.35*	.10	.17	.31*	.14
5.Intelligent				-	.23	.17	.36*	.21
6.Drawing from observation					-	.13	.09	.11
7.Individual style						-	.41*	.12
8.Robots Creative							-	.40*
9.Robots Authors								-

Notes: N=129, *Bonferroni corrected $p < .001$

Table 8. Correlations between measures of subjective and objective value for drawings and perceptions of anthropomorphism and artistic process in drawing robots

	Anthropomorphism	Drawing from observation	Individual style	Robots creative	Robots authors
Aesthetic value	.35*	.25*	.35*	.51*	.36*
Anthropomorphism	-	.19	.25*	.46*	.28*

Notes: $N=125$, *Bonferroni corrected $p<.01$