"What is human?" A Turing Test for artistic creativity.

Antonio Daniele, Caroline Di Bernardi Luft, and Nick Bryan-Kinns

Queen Mary University of London, London, UK

Abstract. This paper presents a study conducted in naturalistic setting with data collected from an interactive art installation. The audience is challenged in a Turing Test for artistic creativity involving recognising human-made versus AI-generated drawing strokes. In most cases, people were able to differentiate human-made strokes above chance. An analysis conducted on the images at the pixel level shows a significant difference between the symmetry of the AI-generated strokes and the human-made ones. However we argue that this feature alone was not key for the differentiation. Further behavioural analysis indicates that people judging more quickly were able to differentiate human-made strokes significantly better than the slower ones. We point to theories of embodiment as a possible explanation of our results.

Keywords: Drawing · Embodiment · Turing Test · Bias.

1 Introduction

This paper presents the design of the interactive art installation "Grammar#1" by Antonio Daniele [1] which was exhibited in May 2019 at the Albumarte gallery in Rome, as one of the 10 winners of the open call *Re:Humanism*. The art installation is presented to the audience as a Turing Test (TT) for artistic creativity [2] where the goal is to recognise human-made artefacts among AI-generated ones. The art work is described to the audience as informing both our scientific research as well as artistic enquiry using their behavioural data collected during the interaction. By questioning what is human and what is artificial, the artist invites the audience to reflect on their own nature as human beings in relation to their understanding of the concept of "artificial". The artistic enquiry of this work addresses the relation between humans and technology in the era of super-human AI and deep fakes by exploiting one of the most ancient and simplest form of human communication: the drawing of lines. In fact, glyphs and symbols are considered one of the earliest forms of human communication [10]. In addition, from more contemporary studies in experimental psychology, we know that abstract lines are successfully used to express moods [22] and spontaneous drawings can be used to interpret complex non-conscious states in clinical settings [6, 5]. While the art work invites reflection about the relation between *human* and

[1] Throughout the paper, the first author will also be referred to as the "artist".

technological in the arts, the study explores the cognitive implications of interfacing with artificially generated content. This problem was previously investigated and discussed from a theoretical perspective [12, 11]. In this paper, the approach is quantitative and focuses on the medium of the drawing, asking whether we perceive substantial differences between human-made and computer-generated strokes and how.

The first part of the paper describes the artist's creative and technical process for the making of the art installation. Initially, a collection of 300 abstract drawings were produced by using automatic drawing techniques [4]. Then, the individual strokes of these drawings were manipulated into a larger dataset and used to train SketchRNN [21] a Variational Autoencoder (VAE) capable of generating simple sketches. Finally, the interactive installation is designed as a TT for artistic creativity [2] where the audience is challenged to recognise human-made strokes amongst the AI-generated ones. The second part of this paper presents the results of the study consisting of the TT for artistic creativity and the analysis of the audience's behavioural patterns, more specifically, the audience response time in relation to their performance. Next, the strokes selected by the audience from both datasets (human-made and AI-generated) are analysed at a pixel level by comparing their respective average entropy and symmetry. Although we find a significant difference in symmetry between the two datasets, a deeper analysis of the audience behaviour shows that the visual feature alone is not crucial for the differentiation. We discuss our findings pointing at theories of embodiment [35, 18, 17].

2 Background

2.1 Turing Test

In 1950, Alan Turing [45] proposed the question "Can machines think?". He described this problem in terms of a game where an interrogator, located in an isolated room, has the objective to recognise the sex of a man and a woman by simply asking them questions. Then, Turing asks if the interrogator would guess wrongly more often if a machine took the part of the man. This game, originally named "the imitation game", became commonly known as the Turing Test, a way to assess a machine's ability to emulate human skills. In 2010, Margaret Boden [2, p.409] argued that

> for an artistic program to pass the TT would be for it to produce artwork which was:
> 1. indistinguishable from one produced by a human being; and/or
> 2. was seen as having as much aesthetic value as one produced by a human being.

Pease and Colton [34] argue that this test might not be accurate enough to evaluate creativity in its true complexity. However, the first point proposed by Boden represents a necessary and sufficient step for avoiding the bias against machine generated content during the evaluation. Any evaluation of an AI-generated artefact indistinguishable from a human-made one would be the result of a choice unbiased against technology. In the next section, we present some of the most relevant studies investigating this problem.

2.2 Bias against machine generated content

Artists and scientists have investigated how people react to computer-generated content since the early age of computer arts. The "Mondrian Experiment" conducted in the Bell Labs by the computer art's pioneer Michael Noll [33], is one of the first attempts to investigate how people compare art made with machines to art made by humans. In this case, Noll generated an image with a program instructed to replicate the patterns from the painting "Composition with Lines" by Piet Mondrian (1917). The results show that only 28% of people were able to recognise the image created in the Bell labs and that the 59% preferred the computer picture. As expected by Noll, the majority of subjects with a technical background were able to recognise the computer image. In contrast, the non-technical people were fooled by assuming that a computer would have built a more "ordered" image while humans would have expressed their creativity with more random patterns.

In Moffat and Kelly [30], the test was conducted on musicians and non-musicians, using music in the style of Bach or Jazz. The participants had to differentiate computer generated from human made music and give their preference to each track. Surprisingly, their results show that non-musicians were significantly better than musicians in recognising computer generated music. Furthermore, both groups preferred human made music, independently from knowing who or what created it.

A more recent experiment was conducted by Elgammal et al. [15] to test the efficiency of their generative model for images. The Creative Adversarial Network (CAN) is a generative model proposed by Elgammal et al. as a variation of the Generative Adversarial Network (GAN) [20]. The visual output generated with CAN was compared with the one generated by existing models (DCGAN) and with abstract art from famous painters or contemporary established artists from the Basel Art fair. Their results show that the participants rated CAN generated works more likeable than the ones produced with DCGAN as well as the works made by the emerging artists. Furthermore, the majority of the participants were tricked into believing that the works made by CAN were actually made by human artists.

In opposition to Elgammal results, Ragot and Martin [36] found a negative bias towards machine generated visual content. The experiment, conducted on a large group of 565 participants, showed images of landscapes and portraits generated by algorithms mixed with paintings of the same subjects made by human artists. According to their results, the 66% of participants were able to recognise human-made paintings and a significant majority preferred human paintings despite knowing who/what made them.

The experiments in this field usually focus on music or more complex visual representations like paintings, whereas the medium of the drawing is almost completely unexplored. To the best of our knowledge, the only study including artificially generated drawings is Chamberlain et al. [8]. The general findings of their larger study show a negative bias towards computer generated works. In the specific experiment with the drawings, the authors explore possible reasons for the negative bias by using the robotic system by the artist Patrick Tresset [43]. The results show that the audience evaluation was conditioned by how much "anthropomorphic" the robotic arm appeared to them, suggesting that the judgment was involving concepts such as "social engagement" or the "embodiment" of human gestures.

The following sections explain why drawings are important for our species and why they are a suitable medium to investigate the mechanisms involved in people's perception of artificially generated content versus human-made one.

2.3 Why Drawings?

From an evolutionary perspective, drawings have had a special role in shaping human cognition since prehistory. Drawing gestures allowed a physical embodiment of the inhabited environment, specifically, simple lines were used to illustrate the shapes of rivers and trees [10]. Drawings possess qualities that are processed by the human brain in a unique way, common to children, cavemen and monkeys [40]. They can be used to represent shapes, volumes and shadows with high accuracy and they are interpreted by our brain as realistic [7]. Furthermore, the action of drawing is essential during the developmental age [22, 14]. For instance, representational flexibility in children facilitates the association of symbols with new meanings [14]. The same cognitive mechanisms allow children as young as four years of age to express emotions and moods by using abstract lines [22]. Furthermore, drawings are often used in therapeutic settings to help patients express non-conscious states [6, 5].

2.4 Automatism and Automation

Automatic Drawings In the history of art, spontaneous activities capable of revealing non-conscious thoughts and feelings are regarded as *automatisms* and are traditionally associated with the Surrealist movement [27, 4]. More precisely, automatic drawings consist of a "pure and simple abandon to graphic impulse" [4, p. 274], an artistic approach that influenced art movements such as the American Abstract Expressionism as well as artists like Salvador Dali and Jean Miro, Jackson Pollock among others.

Let it Brain The abstract drawings used in this art installation and for training SketchRNN are created with an automatic technique named by the artist as "Let it Brain" (LIB). LIB drawing technique has been developed by the artist over the past 20 years in a spontaneous way, independent from traditional automatic techniques from the Surrealist movement. Although LIB shares similarities with

Surrealist automatic drawing, it can be better described as a sort of "human-generative" drawing or "enactive drawing". In this technique the approach consists in enacting the drawing by reducing the time between thought and action, similarly to the "enactivism" described in Manning and Massumi [28]. In this method, the role of real-time and the access to intuitive reasoning is crucial for the expressive gesture.

Computational Drawings While the concept of *automatism* in drawing is generally associated with an expressive activity conducted by humans, drawings made by or with machines might rather be associated to the concept of *automation*. However these two concepts are not necessarily in conflict and in some cases they can intersect. In fact, computers have been used to produce art since the early 60's by computer scientists [32] as well as by artists that treated algorithms as a new artistic medium like Vera Molnar [31] or Harold Cohen [29]. Being able to produce geometric shapes and lines using programming languages was the gateway to produce more complex agents like Cohen's AARON [29], the painting fool [9] or, more recently, Paul the drawing robot by Patrick Tresset [43] or D.O.U.G the collaborative system created by Sougwen Chung. These machines or software are considered by their creators either as tools, extensions, collaborators or as independent entities able to produce art of their own. As with automatic drawings, where non-conscious thinking is involved and the action results in almost random patterns, in computational arts, it is also possible to find approaches like generative arts or evolutionary art, involving a reduced control of the human artist, leaving the machine to operate in (almost) complete autonomy [12].

Drawings and Machine Learning Research in computational drawing has boomed with the ever increasing availability of touchscreen devices and the recent advances in Machine Learning (ML). Nevertheless, the absence of colour and texture, the abstract nature of sketches and the variability in styles are still representing some of the greatest technical challenges [46]. Some attempts have been made in the direction of drawing abstraction at the stroke level [1, 39], however, the target of these architectures is usually a figurative sketch or a photograph. Early attempts to generate drawings using computers mostly exploited perceptual properties of vision by modifying brightness and contrast of existing images, usually relying on features such as edge detection [19]. More recent generative models [1, 24, 13] have demonstrated great results by analysing drawings at a stroke level in their spatio-temporal distribution and classifying by purpose (e.g., shading or contour) [1] or by semantic grouping using deep learning techniques [24].

A milestone in sketch synthesis was reached with the release of SketchRNN [21] a Variational Auto-Encoder (VAE) that uses the "Quick,Draw!" (QD) dataset, the largest dataset of hand-drawn sketches to date. SketchRNN is made of a bidirectional RNN encoder and an autoregressive RNN decoder. At present, this architecture is the state-of-the-art in sketch synthesis. Whereas models based on raster data might be very effective for sketch analysis (e.g., CNN) or generation (e.g., GAN) by using the spatial domain, SketchRNN learns from the very process of drawing in its temporal domain at a fine-grained level.

3 Method

The motivation of this work is twofold. On the one hand, the art installation asks the question "What is human?", inviting the audience to reflect on how the line between the concepts of *human* and *artificial* is often blurred. The artistic exploration uses drawing strokes abstracted from any specific conceptual meaning. On the other hand, the aim of the study is to investigate people's ability to distinguish human-made and artificially generated content as well as exploring some of the factors involved in the differentiation. The following sections describe the process involved in the making of the art installation and the methods used in the study.

3.1 The LIB dataset

For this study, the artist initially created a set of 300 automatic drawings using the LIB technique (see section 2.4). The drawings consist only of abstract lines (no figurative subjects) and were created over a month-long period. In this text, we name this group of drawings as "LIB dataset" (fig. 1). All drawings were realised on paper (size A5: 148 X 210 mm) using a digital tablet capable of translating the artists' gestures from pencil/pen to Scalable Vector Graphics (SVG) format. This guaranteed a more accurate and natural feeling in the making of the drawing.

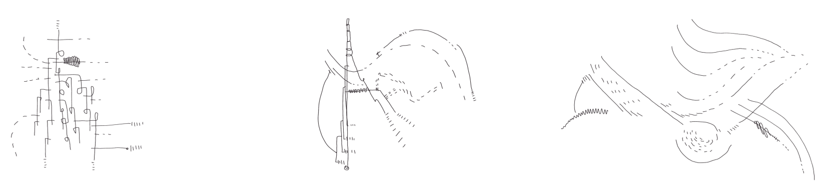

Fig. 1. Three examples of LIB drawings from the LIB dataset

3.2 Preparing the drawings' dataset

The first part of the work consisted in preparing a custom dataset for training SketchRNN. We converted each drawing from SVG format to a Numpy array where the coordinates of the drawing are stored in a specific format: ΔX, ΔY, p [21]. The ΔX and ΔY are the offset's distance of the 2D spatial coordinates in each point of the drawing on the canvas, whereas the p represents the binary state of the pen: 0 if the pen is on the paper, 1 if the pen is lifted.

Modelling the dataset The "Quick,Draw!" (QD) dataset [21] is made of 50 million drawings across 345 categories whereas our dataset (300 drawings) is not just smaller than any class of QD, it is also very different in terms of stroke distribution. While the maximum number of Δoffsets per drawing (ΔX,ΔY) that SketchRNN supports is 250, each drawing of our dataset counts a number of offsets much larger than the limit supported. Whereas both offsets Mean value and Standard Deviation per drawing in the LIB dataset are not in the same range when compared to the QD categories, the high sampling frequency (140Hz) of the tablet produced a total number of offsets much more consistent with QD.

For this reason and because of the abstract nature of the drawings, we considered the whole set of coordinates as a long, whole drawing which we split into smaller units or strokes. This strategy addresses and solves 2 existing problems: 1) the new dataset so obtained will contain a number of strokes much higher than the initial 300 drawings which allows the use of deep learning techniques; 2) The strokes obtained from the split operation will be compatible with SketchRNN standards because their length can be tailored to a maximum of 250 offset points.

Training, generating, evaluating Nine different splitting strategies were tried and nine respective datasets produced and then used to train SketchRNN. From each of the nine trained models we generated 60 strokes among which the artist selected the most similar to his drawing style and evaluated them from *"not acceptable; acceptable; good; best"* . The model that produced the highest number of *good* and *best* strokes was then used to generate a new dataset of artificially generated strokes.

The results of the process so far are two datasets of strokes: 1) *Human Strokes*, a collection of strokes obtained from the original drawings (LIB dataset) and 2)*Artificial Strokes*, a collection of generative strokes learnt from the original drawings. Both datasets contain the same number of strokes (n = 20981).

3.3 The installation

The installation was located in a room (D=1,65 x W=2,23 x H=3,27mt) and all 300 drawings of the LIB dataset were displayed on three walls. In the middle of the room, the interactive experience is accessed via touch screen and a web-based interface (fig. 2). The colour palette used for the installation was limited to black and white. One reason for this choice is to recall the style of LIB drawings. At the same time, the choice of binary colours is a play on the dichotomy between *human* and *artificial*, as commonly perceived in modern society [12]. The experience is designed so that neither the black or white colours can be linked to a specific category (*human*/*artificial*).

Fig. 2. Pictures of the interactive art installation

3.4 User experience

The interactive experience consists of 6 stages accessed through a simple touch interface. The initial stage displays the question "What is Human?", which plays as a philosophical question as well as an instruction. The following three stages consist of simple tasks, specifically three types of TTs for artistic creativity. Once the third task is completed, the results are presented as a data visualisation. Finally, the last touch directs the user to the end of the experience which restates the initial question: "What is human?". In this paper, we analyse and discuss the second task.

The Task 2 is designed as a matrix of 4x4 tiles where the artificial and human-made strokes are randomly displayed. The goal of this task is to recognise the human-made strokes among the artificially generated ones by clicking on the preferred tile. The grid always shows images from both sets but no duplicates are displayed to the user in the same section. The images are randomly taken from both *Human Strokes* and *Artificial Strokes* sets in random proportion. Therefore, each completed task has a variable probability of success determined by the maximum number of human-made strokes displayed in the matrix. The users are not aware of the number of human-made strokes available per session and each time a tile is selected, it disappears from the screen leaving a gap in the matrix. The task is completed when the user has given a number of choices equal to the maximum number of human-made strokes displayed, despite their choices being correct or wrong. The audience has no feedback on whether their selection is correct.

3.5 Materials

The interactive interface was created using P5js, an open source Javascript library based on Processing [38]. The interface is accessed via a touch screen from an 11,6" Acer Chromebook R11. The datasets used for this installation are described in a previous section as *Human Strokes* and *Artificial Strokes*. Both datasets contain n = 20981 strokes. Each of the 300 drawings from the LIB dataset was printed on an 11" photographic paper. The whole dataset is divided into three 10x10 tiles composition and displayed on 3 adjacent walls. The drawings are placed in no particular order (fig. 2).

3.6 Data collection and quality control

This study was conducted in a naturalistic setting, in our case an art gallery. In this context, the participants are the visitors of the exhibition interacting with the installation without any structured introduction or controlled test procedures as it might happen in a lab setting. The behavioural data collected during the interaction consist of the images selected and the time response per selection. No personal data was recorded and the sessions were not monitored, so it can be assumed that one user could have interacted with the installation multiple times or that multiple users interacted within the same session.

To minimise the limitations described above, we conducted additional data quality control steps. In particular, among the two hundred twenty-five (n=225) completed user experiences collected in total, we decided to exclude the first response per user, where the person might be getting familiar with the interface and the task. Then, by controlling the distribution of the data we excluded the outliers completing the task above 30 seconds and with a time response Standard Deviation for each category (*AI* and *Human*) lower than 6 seconds.

Performance score As explained in section 3.4, in each task there is a different probability of success determined by the number of human-made strokes displayed in the matrix. Therefore, all 225 completed tasks represent a binomial distribution where each task is a series of independent Bernoulli distribution. Each independent Bernoulli distribution has two outcomes: 1) human-made or 2) artificially generated. For each completed task, we calculate how the user results compares to the expected probability of success. If we adopt the null hypothesis that the users are randomly guessing, we can reject that hypothesis in all those cases where the probability is below or equal to 0.05. This value tells us to what extent the user was randomly guessing between *human* and *artificial* options. For the behavioural analysis, we analyse only the tasks where the user's probability of randomly guessing is below or equal to 0.05, in which case we assumed that the participant was actively engaging with the task. This parameter is also used as a factor to calculate a performance score on a scale between 0 and 1, where a lower probability of randomly guessing means a better performance.

3.7 Behavioural analysis

The data collected also included the response time for each selected tile. We analysed the differences between the average response time for the human-made strokes and the AI-generated ones. Then, we control how the performance relates to the response time for each category. To do so, we split the mean response time by percentile (25%, 50%, 75%) and obtain three groups: *fast* below 528ms (N=59; M=355ms; SD=115ms), *medium* between 528 and 1686ms (N=119; M=952ms; SD=338ms), *slow* above 1686ms (N=59; M=3256ms; SD=2836ms).

3.8 Strokes analysis

We undertook further analysis at the pixel level of the strokes displayed to the audience during the exhibition. Out of 3600 strokes (225 trials x 16 tiles), the audience saw in total 3418 unique strokes (1700 human-made; 1718 AI-generated), meaning that 182 strokes were shown more than once in the matrix. The table 1 shows the distribution of the strokes according to the audience's choices.

Stroke groups	number of strokes
human-made recognised	927
AI-generated mistaken for human	816
human-made not recognised	773
AI-generated not selected	902

Table 1. Distribution of strokes according to people's choices.

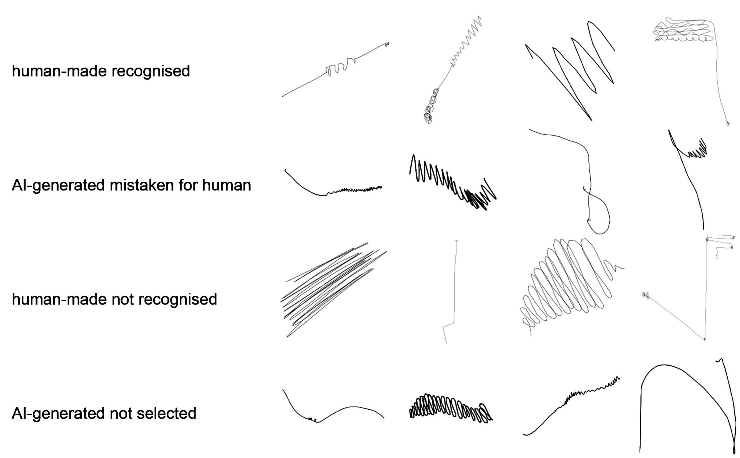

Fig. 3. Some samples from the 4 groups of strokes

For each image in the 4 stroke groups, we analyse and compare two visual features that could be involved in the audience differentiation strategy: the entropy and the symmetry. The entropy [44] is a statistical measure of randomness of the visual stimulus, calculated at the pixel level. Whereas, symmetry is a visual property particularly involved in aesthetic judgment [37] and that the human brain is highly specialised in detecting in nature [25]. We obtained the symmetry from the method proposed in Loy and Ekludun [26]. The feature extracted for each image is a polar coordinate that we converted into a Cartesian coordinate. Then, we calculated the distance from the origin (0,0) and considered this as a scalable value of symmetry.

4 Results

4.1 Turing Test of artistic creativity

Considering the 225 completed tasks, in 63.7% of the cases, the audience was able to select the human-made strokes above the chance.

4.2 Behavioural results

For each of the two classes of strokes (human-made and AI-generated) a one-way ANOVA was conducted to compare the effect of time response as Independent Variable (IV) on performance as Dependent Variable (DV) in the three groups *fast, medium, slow*. In the human-made class, there was a significant effect of time response $F\ (2,120)=3.158$, $p=0.04$. Post-hoc comparisons using the Games-Howell test indicated that the mean performance score for the *fast* group ($M=0.994$, $SD=0.007$) was significantly different than the one in the *slow* group ($M=0.987$, $SD=0.144$), $p=0.036$. However, the *medium speed* group ($M=0.990$, $SD=0.138$) did not significantly differ from the other groups.

4.3 Stroke groups

Entropy An independent t-test was conducted on the human-made ($M=15.997$; $SD=0.003$) and the AI-generated ($M=15.997$; $SD=0.002$) stroke groups. No significant difference was found for entropy $t(3158)=-0.512$, $p=0.609$. A one-way ANOVA was conducted to compare the effect of (DV) entropy on the (IV) audience choice in the four groups: a)human-made recognised; b)AI-generated mistaken for human-made; c)human-made not recognised; d)AI-generated not selected. No significant effect was found for entropy $F\ (3,3414)=0.267$, $p=0.849$.

Symmetry An independent t-test is conducted on the human-made ($M=132.51$; $SD=60.26$) and the AI-generated ($M=151.84$; $SD=57.32$) stroke groups. A highly significant difference was found for symmetry $t(1253)=-5.225$, $p<0.001$. A one-way ANOVA was conducted to compare the effect of (DV) symmetry on the (IV) audience choice in the four groups: a)human-made recognised; b)AI-generated mistaken for human-made; c)human-made not recognised; d)AI-generated not selected. We find a highly significant effect for symmetry $F\ (3,1251)=0.267$, $p<0.001$. Post-hoc comparisons using the Tukey(HSD) test indicated a highly significant difference in symmetry between the human-made strokes ($M=134.03$; $SD=57.79$) and the AI-generated mistaken for human-made ($M=150.36$; $SD=58.01$), $p=0.007$; a highly significant difference in symmetry between the human-made strokes ($M=134.03$; $SD=57.79$) and the AI-generated not selected ($M=153.12$;

SD=56.74), p=0.001. Furthermore, a highly significant difference in symmetry between the AI-generated mistaken for human-made (M=150.36; SD=58.01) and the human-made strokes not recognised as such (M=130.45; SD=63.58), p=0.002; a highly significant difference in symmetry between the the human-made strokes not recognised as such (M=130.45; SD=63.58) and the AI-generated not selected (M=153.12; SD=56.74), p<0.001.

human-made not recognised	AI-generated not selected	p<0.001
human-made recognised	AI-generated not selected	p=0.001
AI-generated mistaken for Human	human-made not recognised	p=0.002
AI-generated mistaken for Human	human-made recognised	p=0.007

Table 2. comparison of significant differences in symmetry among the stroke groups

5 Discussion

The result of our TT of artistic creativity shows that, in a significant majority of cases, people were able to recognise human-made strokes among the AI-generated ones. We can exclude the idea that the entropy of the images displayed during the exhibition is involved in the audience judgement because the difference between the two groups of strokes was not significant. In contrast, the analysis shows a significant difference in symmetry between the strokes made by a human hand and the ones artificially generated with SketchRNN. Therefore, we might suggest that symmetry is one of the features used by the audience to recognise the human-made strokes. Nevertheless, a deeper analysis of the individual groups of strokes selected by the audience indicates that this is not always valid.

For instance, there is a highly significant difference in symmetry between the strokes generated with SketchRNN but never selected by the audience, and the human-made strokes recognised by the audience (p=0.001) as well as the AI-generated ones that fooled the audience and the human-made ones which were not recognised as such (p=0.007). This still points to symmetry as a key factor in the differentiation strategy. However, the most significant difference is found between the human-made strokes not recognised by the audience and the AI-generated strokes not selected (p<0.001). Although we cannot assert that the AI-generated strokes not selected by the audience were actually recognised as artificially generated, we can still ask why, if there is such an evident difference of symmetry between these two groups, the human-made strokes were not recognised? Furthermore, the difference in symmetry between the human-made strokes recognised and not recognised is statistically not significant (p=0.944), meaning that this feature might not be determinant for the final choice. Similarly, when we look at the images artificially generated and mistaken for humans', we can see that their symmetry is significantly different from the group of human-made strokes not recognised as such (p=0.002). However, the symmetry of the

AI-generated images that fooled the audience are not significantly different from the ones that did not fool them (p=0.891).

Further analysis of the behavioural data might offer a deeper insight into this matter. Our results show that people responding faster performed significantly better than the slower ones. The average response time of the slow group was around 3.2 seconds, whereas the faster people average time response is just below 500 milliseconds, a threshold usually associated with non-conscious processes [16] and compatible with an intuitive thinking [23, 3]. Therefore, the results indicate that the differentiation might have happened before conscious thought.

One factor to take into account in our analysis is that the majority of the people interacting with the installation are art-goers. Therefore, it is likely they have a certain familiarity with drawings and this might contribute to the results of our TT of artistic creativity. However, in discussing empirical studies on drawing and perception, Pignocchi [35] explains how even people without formal training are able to recognise drawings made with dexterity or spontaneity, before activating any sort of propositional knowledge. This shows how complex visual features can be processed by the human brain even before activating conscious cognitive processes. Pignocchi creates a case for what he calls the "Motor Perception Hypothesis" (MPH) according to which the simple sight of a drawing can unconsciously inform the viewers about the artist's movements. In the same direction, more recent research in neuroaesthetics [41, 42, 47] demonstrated a strict correlation between drawings and the sensorimotor system, such that one can think of drawings as physical gestures that leave a trace behind [47]. Some of these studies show how simple exposure to static drawing strokes activates in the viewers' brain regions associated with the Mirror Neuron System [41, 42], a group of neurons that are argued to be responsible for mechanisms of social cognition, empathy and Embodied Simulation (ES) [18, 17]. Similarly to the MPH [35], the Embodied Simulation theory proposes that the viewers can non-consciously embody the artists' gestures by simply looking at their drawing strokes.

Considering our results, in particular how intuitive thinking affected the audience's ability to distinguish human-made strokes, we suggest that the properties of the lines involved in the stroke differentiation were processed by the viewers before the conscious cognition. In the ES framework, the brain of the viewer activates as if it was themselves executing the artist's gesture. Therefore, we can speculate that, at least in our case, the model was not able to learn and synthesise such properties in the generative strokes. If human-made strokes are embodied by viewers as the actions of another human, we should ask whether our generative strokes were processed on a non-conscious level as non-human, perhaps because they were not compatible with human gestures. In that case, the differences between human-made and artificially generated strokes might reside not in their mere visual features, rather in the very process of their making.

6 Conclusion

This paper presented the interactive art installation "Grammar#1" exhibited at the Albumarte gallery in Rome and a study conducted with data collected during the audience interactive experience. The art piece questions the audience about the concept of *human* and *artificial*, tapping into their primordial cognition by using abstract drawing strokes. The creative process involved in the making of the art work included the creation of 300 automatic drawings (LIB Dataset) which were used to obtain a larger dataset named in this text *Human Strokes* and to train SketchRNN to generate a second dataset called *Artificial Strokes*. The installation was designed as a TT of artistic creativity [2] where the audience were asked to recognise human-made strokes in different tasks. The data gathered from one of these tasks was used for a study conducted in naturalistic setting where we looked at the behavioural patterns of the audience. Finally, we analysed at pixel level two visual features of the strokes selected by the audience, specifically the entropy and symmetry. The results show that people were able to recognise human-made strokes above the chance in the 63.7% of the cases and we argued that the images' entropy was not involved in their judgment. Although there is a significant difference in symmetry between the human-made and AI-generated images, the discrepancies among the subgroups of selected images led us to suggest that this feature alone was not key for the distinction. Results suggests that the audience achieved better results when using intuitive judgment. We point to the theories of the "Motor Perception Hypothesis" [35] and "Embodied Simulation" [18, 17] to explain the possible dynamics involved in the evaluation. In conclusion, considering our results, the model obtained from training SketchRNN with our dataset was only partially able to learn from the artist's drawings. Further research exploring the process of human-made drawings and people's perception of drawing at a stroke level might eventually produce generative results more difficult to differentiate for the human eye. At the same time, a model able to generate strokes as if they were drawn by human hands, might contribute to the knowledge of human drawing.

Acknowledgements EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology.

References

1. Berger, I., Shamir, A., Mahler, M., Carter, E., Hodgins, J.: Style and abstraction in portrait sketching. ACM Transactions on Graphics (TOG) **32**(4), 1–12 (2013)
2. Bishop, M., Boden, M.A.: The Turing test and artistic creativity. Kybernetes (2010)
3. Bowers, K.S., Regehr, G., Balthazard, C., Parker, K.: Intuition in the context of discovery. Cognitive psychology **22**(1), 72–110 (1990)
4. Breton, A.: Manifestoes of surrealism, vol. 182. University of Michigan Press (1969)

5. Buk, A.: The mirror neuron system and embodied simulation: Clinical implications for art therapists working with trauma survivors. The Arts in Psychotherapy **36**(2), 61–74 (2009)
6. Burns, R.C., Kaufman, S.: Kinetic Family Drawings (KFD): An introduction to understanding children through kinetic drawings. Brunner/Mazel (1970)
7. Casati, R., Pignocchi, A.: Communication advantages of line drawings. Pragmalingüística pp. 000–000 (2007)
8. Chamberlain, R., Mullin, C., Scheerlinck, B., Wagemans, J.: Putting the art in artificial: Aesthetic responses to computer-generated art. Psychology of Aesthetics, Creativity, and the Arts **12**(2), 177 (2018)
9. Colton, S.: The painting fool: Stories from building an automated painter. In: Computers and creativity, pp. 3–38. Springer (2012)
10. Crowther, P.: What drawing and painting really mean: the phenomenology of image and gesture. Taylor & Francis (2017)
11. Daniele, A., Song, Y.: Artistic Assemblage. In: xCoAx 2019: Proceedings of the Seventh Conference on Computation, Communication, Aesthetics & X. pp. 240–255. xCoAx 2019 (2019)
12. Daniele, A., Song, Y.Z.: Ai+Art=Human. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. pp. 155–161 (2019)
13. Das, A., Yang, Y., Hospedales, T., Xiang, T., Song, Y.Z.: Béziersketch: A generative model for scalable vector sketches. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 632–647. Springer International Publishing, Cham (2020)
14. DeLoache, J.S.: Becoming symbol-minded. Trends in cognitive sciences **8**(2), 66–70 (2004)
15. Elgammal, A., Liu, B., Elhoseiny, M., Mazzone, M.: Can: Creative adversarial networks generating “art” by learning about styles and deviating from style norms. In: 8th International Conference on Computational Creativity, ICCC 2017. Georgia Institute of Technology (2017)
16. Fifel, K.: Readiness potential and neuronal determinism: New insights on libet experiment. Journal of Neuroscience **38**(4), 784–786 (2018)
17. Freedberg, D., Gallese, V.: Motion, emotion and empathy in esthetic experience. Trends in cognitive sciences **11**(5), 197–203 (2007)
18. Gallese, V.: Embodied simulation: From neurons to phenomenal experience. Phenomenology and the cognitive sciences **4**(1), 23–48 (2005)
19. Gooch, B., Reinhard, E., Gooch, A.: Human facial illustrations: Creation and psychophysical evaluation. ACM Transactions on Graphics (TOG) **23**(1), 27–44 (2004)
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
21. Ha, D., Eck, D.: A Neural Representation of Sketch Drawings. In: International Conference on Learning Representations (2018)
22. Ives, S.W.: The development of expressivity in drawing. British Journal of Educational Psychology **54**(2), 152–159 (1984)
23. Kahneman, D.: Thinking, fast and slow. Macmillan (2011)
24. Li, Y., Song, Y.Z., Hospedales, T.M., Gong, S.: Free-hand sketch synthesis with deformable stroke models. International Journal of Computer Vision **122**(1), 169–190 (2017)
25. Little, A.C., Jones, B.C.: Attraction independent of detection suggests special mechanisms for symmetry preferences in human face perception. Proceedings of the Royal Society B: Biological Sciences **273**(1605), 3093–3099 (2006)

26. Loy, G., Eklundh, J.O.: Detecting symmetry and symmetric constellations of features. In: European Conference on Computer Vision. pp. 508–521. Springer (2006)
27. Maclagan, D.: Line let loose: Scribbling, doodling and automatic drawing. Reaktion Books (2013)
28. Manning, E., Massumi, B.: Thought in the Act: Passages in the Ecology of Experience. U of Minnesota Press (2014)
29. McCorduck, P.: Aaron's code: meta-art, artificial intelligence, and the work of Harold Cohen. Macmillan (1991)
30. Moffat, D., Kelly, M.: An investigation into people's bias against computational creativity in music composition. In: The Third Joint Workshop on Computational Creativity. ECAI 2006, Universita di Trento (aug 2006)
31. Molnar, V.: Toward aesthetic guidelines for paintings with the aid of a computer. Leonardo pp. 185–189 (1975)
32. Nake, F.: Computer art: a personal recollection. In: Proceedings of the 5th conference on Creativity & cognition. pp. 54–62 (2005)
33. Noll, A.M.: Human or machine: A subjective comparison of Piet Mondrian's "composition with lines" (1917) and a computer-generated picture. The psychological record **16**(1), 1–10 (1966)
34. Pease, A., Colton, S.: On impact and evaluation in computational creativity: A discussion of the Turing test and an alternative proposal. In: Proceedings of the AISB symposium on AI and Philosophy. vol. 39 (2011)
35. Pignocchi, A.: How the intentions of the draftsman shape perception of a drawing. Consciousness and cognition **19**(4), 887–898 (2010)
36. Ragot, M., Martin, N., Cojean, S.: AI-generated vs. human artworks. a perception bias towards artificial intelligence? In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–10 (2020)
37. Ramachandran, V.S., Hirstein, W.: The science of art: A neurological theory of aesthetic experience. Journal of consciousness Studies **6**(6-7), 15–51 (1999)
38. Reas, C., Fry, B.: Processing: a programming handbook for visual designers and artists. Mit Press (2007)
39. Riaz Muhammad, U., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Learning deep sketch abstraction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8014–8023 (2018)
40. Sayim, B., Cavanagh, P.: What line drawings reveal about the visual brain. Frontiers in human neuroscience **5**, 118 (2011)
41. Sbriscia-Fioretti, B., Berchio, C., Freedberg, D., Gallese, V., Umiltà, M.A.: Erp modulation during observation of abstract paintings by franz kline. PLoS One **8**(10), e75241 (2013)
42. Thakral, P.P., Moo, L.R., Slotnick, S.D.: A neural mechanism for aesthetic experience. Neuroreport **23**(5), 310–313 (2012)
43. Tresset, P., Leymarie, F.F.: Portrait drawing by Paul the robot. Computers & Graphics **37**(5), 348–363 (2013)
44. Tsai, D.Y., Lee, Y., Matsuyama, E.: Information entropy measure for evaluation of image quality. Journal of digital imaging **21**(3), 338–347 (2008)
45. Turing, A.M.: Computing machinery and intelligence. In: Parsing the Turing test, pp. 23–65. Springer (2009)
46. Yu, Q., Yang, Y., Liu, F., Song, Y.Z., Xiang, T., Hospedales, T.M.: Sketch-a-net: A deep neural network that beats humans. International journal of computer vision **122**(3), 411–425 (2017)
47. Yuan, Y., Brown, S.: The neural basis of mark making: A functional MRI study of drawing. PloS one **9**(10), e108628 (2014)