Distinguishing between Abstract Art by Artists vs. Children and Animals: Comparison between Human and Machine Perception

LIOR SHAMIR, Lawrence Technological University
JENNY NISSEL and ELLEN WINNER, Boston College

Abstract expressionism is a school of art characterized by nonrepresentational paintings where color, composition, and brush strokes are used to express emotion. These works are often misunderstood by the public who see them as requiring no skill and as images that even a child could have created. However, a recent series of studies has shown that ordinary adults untrained in art or art history, as well as young children, can differentiate paintings by abstract expressionists and superficially similar works by preschool children and even animals (monkeys, apes, elephants). Adults perform this distinction with an accuracy rate of ∼64%, significantly higher than chance. Here we ask whether machine perception can do as well. Using the same paintings, we show that in ∼68% of the cases the computer algorithm can discriminate between abstract paintings and the work of children and animals. We also applied a method that computes the correlation between the degree of artisticity deduced from human perception of the paintings and the visual content of the images, and we show significant correlation between perceived artisticity and visual content. The image content descriptor that was the strongest predictor of correct identification was the fractality of the painting. We also show that the computer algorithm predicts the perceived intentionality of the paintings by humans. These results confirm perceptible differences between works by abstract expressionists and superficially similar ones by the untrained and show that people see more than they think they see when looking at abstract expressionism.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; I.5.0 [**Pattern Recognition**]: General; I.2.0 [**Artificial Intelligence**]: General—*Cognitive simulation*

General Terms: Human Factors

Additional Key Words and Phrases: Art, abstract expressionism

ACM Reference Format:

Lior Shamir, Jenny Nissel, and Ellen Winner. 2016. Distinguishing between abstract art by artists vs. children and animals: Comparison between human and machine perception. ACM Trans. Appl. Percept. 13, 3, Article 17 (May 2016), 17 pages.
DOI: http://dx.doi.org/10.1145/2912125

17

Authors' addresses: L. Shamir, Computer Science department, Lawrence Technological University, 21000 W Ten Mile Rd., Southfield, MI 48075; email: lshamir@mtu.edu; J. Nissel, Department of Psychology, Boston College, McGuinn 300140 Commonwealth Ave, Chestnut Hill, MA 02467; email: Jenny.Nissel@gmail.com; E. Winner, Department of Psychology, Boston College, McGuinn 343, 140 Commonwealth Ave, Chestnut Hill, MA 02467; email: ellen.winner@bc.edu.

DOI: http://dx.doi.org/10.1145/2912125

1. INTRODUCTION

Since the first half of the 20th century, abstract expressionism has been a primary school of art, and members of the movement, such as Jackson Pollock, Mark Rothko, and Mark Tobey, have been some of the world's most renowned modern painters. Abstract expressionist paintings are entirely non-representational, and understanding abstract expressionism requires familiarity with what the artists are trying to accomplish [Varnedoe 2006]. People who are not familiar with abstract expressionism are sometimes shocked by abstract art, wondering whether these seemingly random, messy, or apparently simple patterns require any skill to produce and wondering secretly whether anyone—even a child—might be able to create the same canvas. However, a series of studies has demonstrated that adults lacking any specialized knowledge of art history are capable of distinguishing works by abstract expressionists from superficially similar (in medium, color, and pattern) works by preschool children and animals (monkeys, apes, and elephants).

Hawley-Dolan and Winner [2011] showed people paintings by famous abstract expressionists (e.g., Hans Hofmann, Cy Twombly, Franz Kline, and Sam Francis) paired with paintings by preschool children or animals (gorillas, chimpanzees, monkeys, elephants). There were 30 pairs, with paintings in each pair matched as closely as possible on at least two of the following characteristics: color, line quality, brushstroke, medium, and composition. Participants were asked which painting they liked more and which they thought was a better work of art. Whether the pairs were presented with correct attribution labels (artist vs. child, monkey, or elephant), reversed labels, or no labels, people selected the artist works over the child/animal works at a rate significantly above chance. Whether or not these participants might have commented that "my kid could have done that," their responses show us that they are able to distinguish between abstract works by artists and works by children and animals.

This finding was replicated in another study in which the 60 images were presented singly and in random order, without labels. The task was to decide, in each case, whether the work was by an artist rather than a child or animal [Snapper et al. 2015]. Participants' d-prime scores differed significantly from zero. Thus, successful discrimination of artist works from works by children and animals does not depend on viewing these as "matched" pairs (which could, in principle, make the task easier by accentuating the differences). However, a large variability was found across images, with correct classifications ranging from 12% to 94% for the artist images and from 29% to 91% for the child/animal images. Clearly, some artist images are easy to classify correctly, while others are confusable with works by children and animals.

In a first attempt to discover what differentiates these two classes of paintings, Snapper et al. [2015] presented participants unfamiliar with art history with the same 60 images, without mentioning that some were by artists and others by children and animals, and asked participants to rate each image on 7-point scales for degree of intentionality, structure, negative space, metaphorical meaning, communication, and inspiration. Only one rating scale differentiated paintings by artists vs. children or animals, and that was how intentional the painting was perceived to be: Paintings by artists were rated significantly higher on intentionality than were the paintings by children and animals. As mentioned, there was considerable variability of accuracy in identifying paintings as by artists. Those paintings by artists that were clearly recognizable as by artists were also rated as more intentional and more structured than those paintings by artists that were often confused with works by children and animals. Similarly, child/animal works that were clearly recognizable as by children or animals were seen as less intentional and less structured than those often confused with works by artists.

Here we use computational methods to analyze abstract art and approach the question of which perceptible features or sets of features distinguish works by professional artists from those by children and animals. A number of studies have demonstrated that automated classification of painting styles is possible [Keren 2002; Li and Wang 2004; Li et al. 2014; Taylor et al. 2002, 2007; Khan et al. 2014; Cetinic and Grgic 2013]. Applications include the association of paintings with schools of art [Shamir et al. 2010], association of paintings with relevant key words [Mattison 2004; Tsai 2007], identification of drawing methods [Kammerer et al. 2007; Roussopoulos et al. 2010], recognition of painters [Taylor et al. 1999, 2007; Li et al. 2012; Kronner and Lattner 1998; Lyu et al. 2004], and quantification of similarities among artistic styles [Shamir 2012; Kim and Kim 2013]. Further studies have also shown that computers can analyze art in an unsupervised fashion [Shamir et al. 2010; Shamir and Tarakhovsky 2012] and automatically associate similar artistic styles without any prior knowledge or training.

Here we subjected the images used by Snapper et al. [2015] to the automated algorithm used in Shamir et al. [2010] in order to compare machine perception to human perception in terms of respective ability to distinguish works by famous abstract expressionists from superficially similar works by preschoolers, monkeys, apes, and elephants (Study 1). We also applied a method of detecting correlations between visual content and numerical values to test whether there was a correlation between those paintings most easily classified by humans and those most successfully classified by the computer algorithm, and, if so, to identify the numerical image content descriptors that predicted successful identification of works by artists.

In two additional experiments we tested for consistency in what non-experts consider art and show that computer analysis can predict the artisticity of an abstract painting as perceived by an individual person. Instead of using the creator of each painting as the ground truth as done in Experiment 1, we train the machine-learning system for each participant based on the paintings she believed were the work of artists and paintings she perceived as the work of children and animals.

2. MATERIALS

Sixty digitized paintings were used in Experiments 1 and 2, with 30 painted by known abstract expressionist artists and 30 by preschool-aged children and animals. Each work by a child or animal had been initially selected because it was superficially similar in color and composition to one of the works by the artists. That made it possible to pair the paintings such that each pair contained one professional artist painting and one child painting similar to the other in medium, color, and composition [Hawley-Dolan and Winner 2011]. These pairs were presented to participants in random order with no time limitation. In Hawley-Dolan and Winner's study, participants were asked to decide which member of each pair was the better work of art. Works by artists were chosen at a rate significantly above chance, even when incorrect labels were provided (e.g., pairing works by an artist with the label "child").

Snapper et al. [2015] showed these images unpaired and without labels to a cohort of 103 participants who identified themselves as unfamiliar with abstract expressionism. Forty-nine of the participants were male, and 54 females, and all participants received payment for their work. The age of the participants ranged from 19 to 76 years (M = 34). Each participant annotated the 60 images by answering the question "Please choose whether you think it was made by an artist or by a child or animal." The annotations were recorded by an on-line survey through Qualtrics.

The participants also had to answer an attention check question about the last painting (after it was no longer visible) to make sure they were concentrating on the task and were looking at the paintings. The task took an average of 8 minutes. Participants were able to distinguish artist works from those by children and animals at a mean rate of ~64% [Snapper et al. 2015], which was significantly above chance.

In Experiment 3 the individual votes of 71 participants were used to construct 71 different datasets separated into paintings that the participant indicated were created by an artist and those that

Fig. 1. Sample images used. Artist images are on the left (top: Hans Hofmann, *Laburnum*, credit: ARS; bottom: Sam Francis, *Untitled*, 1989, credit: ARS); non-artist images are on the right (top: Ronan Scott, preschool student, credit: Ronan Scott; bottom: Ging-Gaow, Elephant, credit: Asian Elephant Art & Conservation Project (AEACP).

the participant indicated were created by a child or an animal. Instead of separating the paintings into paintings actually created by artists and paintings created by children/animals as was done in Experiment 1, the paintings were separated into paintings that the participant believed were created by an artist and paintings the participant believed were painted by a child or an animal. Since each participant had different separation of paintings, the experiment was performed with 71 different datasets.

In Experiment 4 we used data produced by 21 participants who rated each painting on a scale of 1 through 7 for the intentionality of the work, with higher scores meaning that the work appeared to be more intentionally created. The purpose of the experiment was to test how well the machine-learning algorithm can predict how intentional a certain piece of art would seem to a certain non-expert art viewer.

3. COMPUTER ANALYSIS METHOD

We subjected these 60 images to computer analysis. The size of all paintings was normalized to 500K pixels such that the original aspect ratio was preserved. Figure 1 shows a few examples of images used, two by abstract expressionists (left column), one by a child (right column), and one by an elephant (right column).

The image analysis method used here is based on the Wndchrm feature set [Shamir et al. 2008a, 2009b, 2009c, 2010, 2013b; Shamir 2008], which includes 4,025 numerical image content descriptors. That feature set has proved its efficacy in automatic analysis of visual art [Shamir et al. 2010; Shamir

2012, 2015; Shamir and Tarakhovsky 2012], and its breath and comprehensive nature allows it to measure very many different aspects of the visual content. In summary, these image features include the following:

- High-contrast features such as object statistics and Gabor filters [Gabor 1946] measure the size and spatial distribution of high-contrast objects detected in the image.
- Edge statistics measure the direction and magnitude of edges that separate between objects or other parts of the image using the using the Prewitt edge detection operator.
- Texture features [Haralick et al. 1973], and Tamura [Tamura et al. 1978] reflect the distribution of textures in the paintings. The Tamura features refer to the coarseness, contrast, and directionality of the textures in an image. The Haralick texture measure uses the co-occurrence matrix of an image that stores information about pixels with equal gray values located in the same pixel neighborhoods. That information is then used to compute the energy, entropy, and correlation of the textures [Haralick et al. 1973].
- Statistical distribution of the pixel values such as the first four moments and multi-scale histograms are sensitive to the distribution of pixel brightness in the image. In their most basic form, multi-scale histogram features can differentiate between paintings that are more uniform in brightness and paintings that have are made of dark and bright areas.
- Fractal features [Mandelbrot 1982] measure repeating patterns that can exhibit themselves at different scales. We analyzed the fractality of the paintings by using the multi-resolution box-counting fractal features described in Wu et al. [1992].
- Two-dimensional Chebyshev statistics [Gradshtein and Ryzhik 1994] represent an image through the coefficients of its polynomial approximation. The histogram of the coefficient values reflects the approximation of the changes in the brightness of the pixels.
- Zernike polynomials [Teague 1980] reflect pixel brightness changes around the unit disk through the absolute values of the complex Zernike coefficients.
- Chebyshev-Fourier statistics [Orlov et al. 2008] are computed similarly to the Chebyshev statistics but after applying the Fourier transform.
- Radon features [Lim 1990], measure the projection of pixel intensities at different angles from the center of the image to identify a correlation of the pixels to a specific angle [Lim 1990].
- Statistical distribution of colors [Shamir 2006] measure the prevalence of each color in the image based on a fuzzy logic model that follows the human perception of colors [Shamir 2006].

A detailed description of these image features can with performance evaluation be found in Shamir et al. [2008a], Shamir [2008], Orlov et al. [2008], and Shamir et al. [2009b, 2009c, 2010, 2013b].

The image content descriptors are computed not just from the raw images but also from image transforms and multi-order image transforms, which are transforms of the raw pixels as well as transforms of these transforms [Orlov et al. 2008; Shamir et al. 2009a, 2008a]. That is, global image features are computed from the raw pixels, from transforms of the raw pixels, and from transforms of these transforms. Image transforms include the Fourier, Chebyshev, Wavelet, edge-magnitude transform, hue transform, and different combinations of compound multi-order transforms as described in Shamir [2008] and Shamir et al. [2008a, 2009b, 2010]. This novel approach of using multi-order image transforms has demonstrated its efficacy for analyzing complex morphology [Shamir et al. 2008a, 2009a, 2009b, 2010]. For more details and performance evaluation of the Wndchrm feature set and multi-order transforms, see [Shamir 2008; Shamir et al. 2008a, 2009b, 2009c, 2010, 2013b, 2014; Shamir and Tarakhovsky, 2012]. The source code of the method is publicly available [Shamir et al. 2013a].

3.1 Automatic Classification Method

Wndchrm computes a very large set of numerical image content descriptors, not all of which are equally informative. To remove non-informative numerical image content descriptors, the Fisher discriminant score for each feature is computed using the training samples according to Equation (1):

$$W_f = \frac{\sum_{c=1}^{N}\left(\overline{T_f} - \overline{T_{f,c}}\right)^2}{\sum_{c=1}^{N}\sigma_{f,c}^2} \cdot \frac{N}{N-1}, \tag{1}$$

where W_f is the Fisher discriminant score, N is the total number of classes, $\overline{T_f}$ is the mean of the values of feature f in the entire dataset, $\overline{T_{f,c}}$ is the mean of the values of feature f in the class c, and ${\sigma_{\mathrm{f,c}}}^2$ is the variance of feature f among all samples of class c.

After the Fisher discriminant scores are computed, the 70% of the features with the lowest Fisher discriminant score are removed from the analysis [Shamir et al. 2010; Shamir and Tarakhovsky 2012; Shamir 2015].

After removing 70% of the features, the remaining feature set is still large and contains 1,207 features. It can be reasonably assumed that not all features in such a large feature set are equally informative. Therefore, the classification is performed using a Weighted Nearest Neighbor scheme such that the Fisher discriminant scores computed using the training set are used as weights. That is, the Fisher discriminant scores are used to filter some of the non-informative numerical image content descriptors and are then used to assign different weights to the features so the impact of each feature on the classification result is proportional to its Fisher discriminant score. This method was shown to differentiate paintings by artists and schools of art [Shamir et al. 2010; Shamir and Tarakhovsky 2012] and identify differences between artistic styles [Shamir 2012; Shamir 2015].

3.2 Automatic Correlation with Painting "Artisticity"

In Experiment 1 we asked whether the computer algorithm could rival humans in the ability to distinguish works by abstract expressionists vs. works by children and animals. The human annotation of the paintings showed that some paintings were classified by over 90% of the participants as paintings made by professional painters, while other paintings were voted "professional" by just 12% [Snapper et al. 2015], showing that the degree of "artisticity" varied strikingly across paintings. In Experiment 2 we test whether the human perception of the "artisticity" of the paintings correlated with machine perception.

We used the 60 images from Snapper et al. [2015] described above. We first assigned a score to each painting reflecting its perceived "artisticity" described in Equation (2):

$$A = \frac{p}{N}, \tag{2}$$

where A is the perceived artisticity score of the painting, p is the number of people who identified the painting as the work of a professional painter, and N is the total number of people who viewed the painting. That score provides additional information to the simple binary classification of the paintings into work of artists and work of non-artists and allows us to test for the existence of a correlation between the visual content as analyzed by the computer and the perceived artisticity of the paintings determined based on the answers of the participants.

That correlation is computed by first using a computer method to predict the artisticity of each painting. After all predicted artisticity values are computed, the Pearson correlation between the predicted artisticity and the artisticity determined based on the answers of the human participants can be deduced.

To compute the predicted artisticity of a painting, we first assigned weights to each feature using the training samples such that the weight is the Pearson correlation between the feature f with the numerical value v [Shamir 2011] as described in Equation (3):

$$W_f = \left| \frac{1}{N} \sum_{i=1}^{N} \left(\frac{f_i - \bar{f}}{\sigma_f} \right) \cdot \left(\frac{v_i - \bar{v}}{\sigma_v} \right) \right|, \tag{3}$$

where W_f is the weight assigned to feature f, f_i is the value of feature f in painting i, $\bar{f}$ is the mean value of feature f, v_i is the artisticity of painting i, and N is the number of paintings in the training set.

When the values of a certain numerical image content descriptor f have a strong correlation with the artisticity of the paintings, W_f is higher and gets closer to 1, while when the correlation is weak, W_f is lower and closer to 0. Because the weight W_f is the absolute value of the Pearson correlation, the same will be true also for inverse correlation between the values of the content descriptor and the artisticity of the paintings in the training set. Therefore, a numerical image content descriptor that provides information about the artisticity of the painting will be assigned a higher weight, while less informative image features will be assigned a lower weight. That allows assigning weights and selecting the most informative features when the data are not divided into distinct classes as in Shamir et al. [2010], but instead each image is assigned with a numerical value, which is the artisticity of the paintings as determined by human perception.

The 40 features that had the strongest correlation with the artisticity of the paintings were then used for predicting the artisticity value v_i of a certain painting i by the following Equation (4):

$$v_i = \frac{\sum_{j=1}^{K} d(i, j, W) \cdot v_j}{\sum_{j=1}^{K} d(i, j, W)}, \tag{4}$$

where *d(i, j, W)* is the weighted Euclidean distance between the feature vectors of image i and j, W is the vector of the weights of all features computed using Equation (3), and K is the number of closest neighbors to the feature vector I found in the training set.

4. RESULTS

4.1 Experiment 1: Automatic Classification of Paintings by Artists vs. Children/Animals

In the first experiment we applied the classification method described in Section 3.1 to test whether the computer method could distinguish automatically between paintings made by professional artists vs. those by children and animals. The 60 images of the dataset described in Section 2 consisted of two classes of 30 images painted by professional artists and 30 paintings by children or animals. Twenty-five images from each of the two classes were used for training, and 5 images per class were used for testing. The experiment was performed 15 times, with randomly allocated images assigned to training vs. test sets for each run.

The average classification accuracy of all 15 runs was $\sim 0.68 \pm 0.0296$, showing that the computer method was able to classify between paintings made by professional abstract expressionist artists and those made by children or animals ($P < 0.0017$). The mean accuracy of machine and human are surprisingly close. Table II specifies the results of the individual classification of each painting.

Appendix Figure 1 shows the accumulated Fisher discriminant scores of the different groups of numerical image content descriptors extracted from the different image transforms. As the figure shows, many different types of numerical image content descriptors were associated with discrimination between the two classes of works, but the most informative descriptors were the fractal features. Other informative features were the Gabor textures, Zernike features, and the Chebyshev statistics.

Following Shamir et al. [2010], we also tested the classification accuracy when the pattern recognition is performed with Support Vector Machine (SVM) instead of the pattern recognition system described in Section 3.1. SVM^{light} was used for the classification, with linear kernel, radial basis function (RBF) kernel ($\gamma = 10$), and polynomial kernel ($d = 4$). The classification accuracies with linear, RBF, and polynomial kernel were $\sim$61.4%, $\sim$63.1%, and $\sim$63.5%, respectively. Although the classification accuracy is lower, these results still demonstrate computer classifications paintings by professional painters vs. children or animals at a rate significantly above chance ($P < 0.026$).

4.2 Experiment 2: Automatic Prediction of Painting Degree of Artisticity

In the second experiment we tested whether computed artisticity correlates with human-perceived artisticity of the painting. As described in Section 3.2, in a previous study by Snapper et al. [2015] each painting had been annotated by multiple participants. The percentage of times each image was annotated as the work of an artist is provided in the Appendix, with 1.0 indicating that all participants had classified that painting as by an artist, 0.5 indicating that half had done so, and 0 indicating that none had done so. The percentage of participants who classified each painting as the work of an artist can be used as a measure of the "artisticity" of that painting, as defined by Equation (2). In the experiment, we tested whether the method described in Section 3.2 can compute the human perceived artisticity of the paintings

The method described in Section 3.2 was applied to that dataset, with 45 images used for training, 15 for testing, and three nearest neighbors ($K = 3$ in Equation (4)). The experiment was repeated 10 times, with randomly allocated images assigned to training and test sets for each run. The artisticity predicted by the computer and the artisticity determined by the manual annotation as described by Equation (2) provided a series of pairs, such that each test painting was assigned two values–the machine perception artisticity and the human perception artisticity. The Pearson correlation between these values shows the correlation between the human and machine perception.

The results show that the correlation between the predicted artisticity deduced by the computational method described in Section 3.2 and the artisticity determined based on the human annotation as defined by Equation (2) was $\sim$0.3 ($P < 0.024$). Table II specifies the computed artisticity for each individual painting, and Figure 2 shows the human perceived artisticity and the computed artisticity of the different paintings.

Similarly to the Fisher discriminant scores used in Experiment 1, the Pearson correlation that was used in Experiment 2 to select and weight the features can be used to identify the most informative features that had the strongest impact on the analysis. The analysis of the most informative numerical image content descriptors is available in the Appendix.

The analysis shows that of the 40 most informative numerical image content descriptors, 25 features are fractal features. No other group of features exhibited dominance, though there was some presence of texture features such as two Haralick texture descriptors and three Tamura texture descriptors. Appendix Figure 2 shows the number of features among the 40 features with the highest Pearson correlation with the artisticity of the paintings.

4.3 Experiment 3: Automatic Classification of Paintings Perceived by Humans as by Artists vs. by Children/Animals

In Experiment 3, for each participant we separated the paintings into two image classes: paintings that the human participant believed were the work of artists, and paintings that that were believed by the participant to be the work of a child or animal. The machine-learning method was the same method

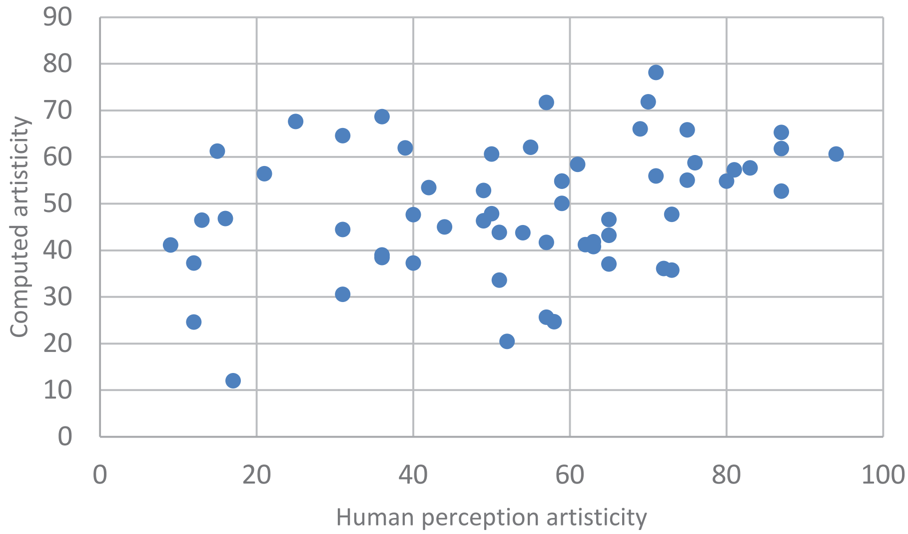


Fig. 2. Human perception artisticity and machine perception artisticity of the different paintings.

used in Experiment 1, but the paintings were separated by the way the paintings were perceived by each participant, and not by their creator, as was done in Experiment 1.

The average accuracy with which the algorithm predicted each of the 71 human participants' accuracy was $\sim$0.657 $\pm$ 0.042. This demonstrates a consistency between algorithms and humans in what is classified as a work by an adult artist vs. work by children or animals.

Appendix Figure 3 shows the average Fisher discriminant scores of some of the features used for the discrimination between paintings *perceived* as the work of adult artists vs. those *perceived* as the work of children or animals, regardless of the actual creator of each piece. When we compare this figure to Appendix Figure 1 (which shows the features that the algorithm uses to classify the paintings), we see that fractals are informative in both. However, a comparison of the two figures also shows that humans attempting to identify whether the painting is the work of a professional artist or a child or animal use the fractality of the paintings less than the artists themselves as they paint.

4.4 Experiment 4: Automatic Prediction of the Human Perceived Intentionality of Paintings

In Experiment 4 we used the method described in Section 3.2 and used in Experiment 2, with the data collected from 21 participants who rated each painting for its degree of intentionality on a scale from 1 (not at all intentional) to 7 (very intentional), as described in Section 2. The method was trained and tested using the data of each participant separately with a leave-one-out cross-validation strategy, therefore repeating the experiment 21 times, each time with the annotations of a different participant.

The mean difference between the human perceived intentionality and the intentionality computed by the algorithm was small: $\sim$0.759 $\pm$ 0.081. Thus, we demonstrate that our algorithm can be used to predict the level of intentionality perceived in a work of art by non-expert humans. That is, given a set paintings annotated by a certain non-expert viewer for their intentionality, the algorithm can predict the intentionality that the viewer will perceive in a new painting.

5. DISCUSSION

Numerous cognitive and cognitive neuroscience experiments have shown that the perception of visual art is not driven by what the human eye can sense but by the perceptual processes activated in the brain when an observer is exposed to art [Ramachandran and Herstein 1999], leading to the conclusion that artists paint with their brain, not with their eyes [Zeki and Nash 1999]. Here we use computational approaches to study the human perception of abstract art and show that the human perception of artisticity of abstract art paintings correlates with the artisticity computed by a machine-learning algorithm.

In Experiment 1 we showed that a computer can differentiate paintings by abstract expressionist painters from those by children and animals with an accuracy rate of ~68%, favorably comparable to the performance of untrained human observers on this same task. While the classification accuracy is far from 100%, it should be noted that machine vision is in most cases inferior to human vision. Indeed, even relatively simple vision tasks such as object or face recognition are considered challenging problems for computing machines and are not performed with 100% accuracy.

The second experiment examined the artisticity of the abstract expressionist paintings (defined by how often they were correctly classified by human participants) and revealed a significant correlation between the artisticity as judged by human observers and as judged by the computer algorithm. Most of the features that correlate with the artisticity of the paintings are fractal features, suggesting that human recognition of the distinction between the two classes of paintings studied here may be linked to fractality, and that fractality may underlie the human perception of greater intentionality in works by artists compared to works by the untrained.

The third and fourth experiments showed, respectively, consistency between the algorithm and human participants in classifying a work as by an artist rather than a child or animal and in rating the work of art in terms of how intentional it appears to be.

The link between fractality and abstract expressionism has been discussed in the context of the work of Jackson Pollock, and Taylor et al. [1999, 2007] showed that fractal analysis of Pollock's drip paintings provided evidence that fractality can be used to authenticate Jackson Pollock paintings [Taylor et al. 1999, 2007]. However, the authentication using fractal analysis was criticized by Jones-Smith and Mathur [2006], who showed that fairly simple paintings could reproduce the fractality observed in Jackson Pollock's drip paintings, hence challenging the claim that the fractals were sufficiently unique to fully distinguish authentic from forged Pollock paintings. Fractal analysis has, however, shown that fractals can be found in Pollock's works and that his fractality changed gradually over the years [Taylor et al. 2002, 2006]. It is notable that no Pollock paintings were included in the image set used here, and thus our findings point to fractality outside of Pollock's works.

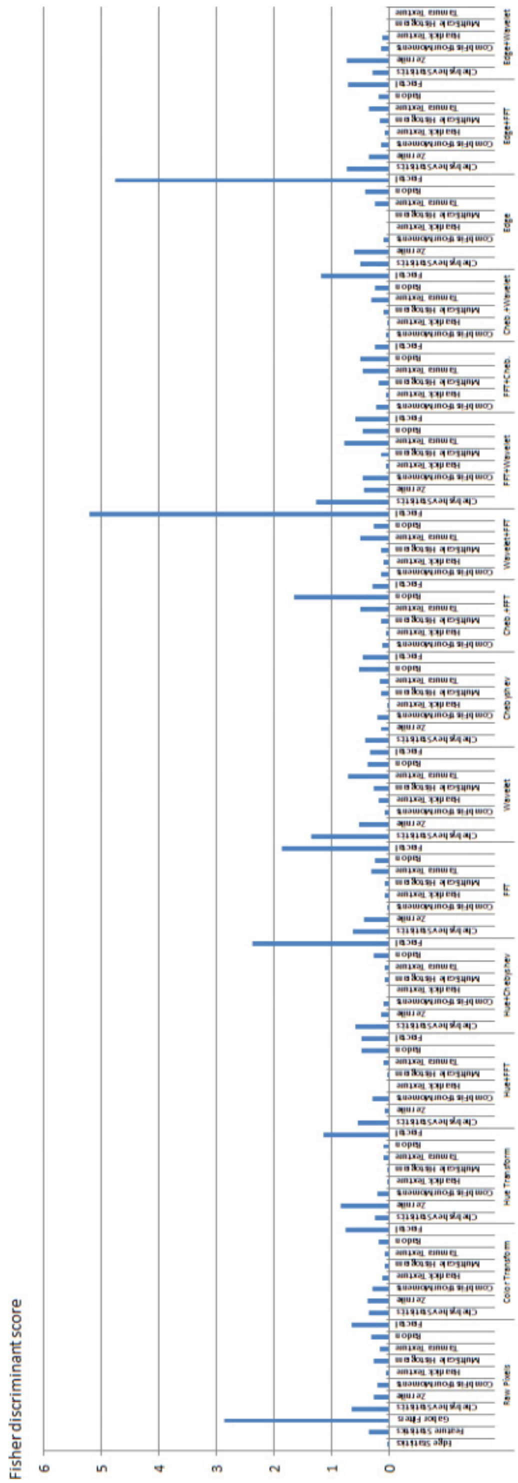
The results presented in this article show that computers can distinguish between abstract paintings by artists vs. children and animals and that computers show similar error patterns to humans. This study is the first to our knowledge to uncover perceptible features that distinguish abstract expressionist paintings from superficially similar and often charming works by children and animals, leading to the conclusion that fractality is the most important feature guiding the computer in the discrimination. We suggest that humans may be using this feature as well when they make this discrimination. Untutored observers see more in abstract expressionism than they think they see. Our computer algorithm confirms that there are systematic and perceptible differences distinguishing abstract expressionist paintings from what appear to the untrained eye to be strikingly similar works by children and animals.

APPENDIX

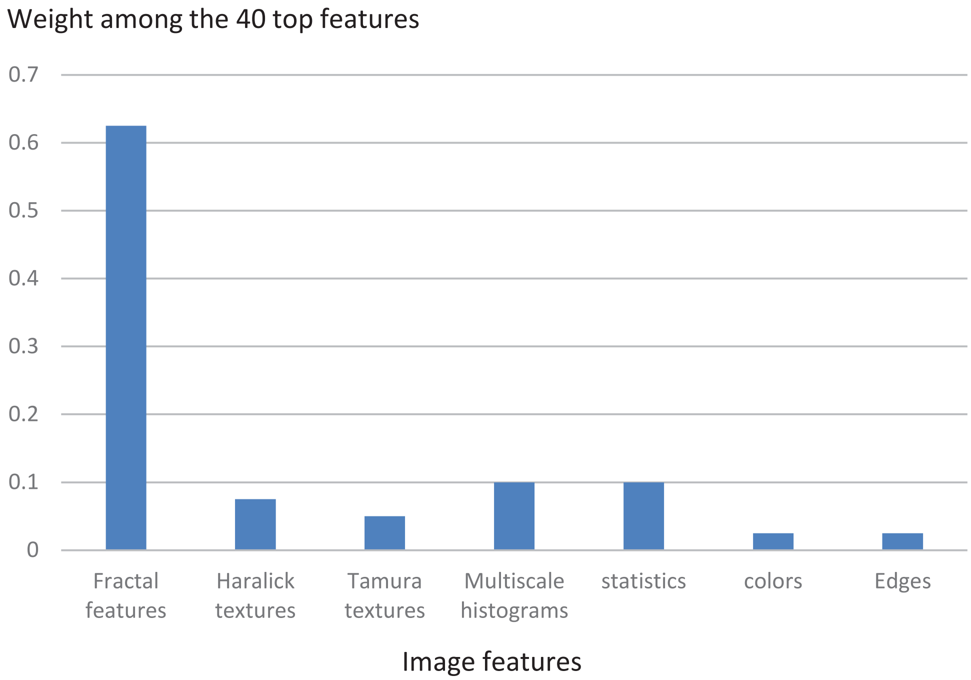
The 40 numerical image content descriptors with the strongest correlation to the artisticity of the paintings as averaged over the 15 runs are displayed in Table I.

Appendix Table I. The Numerical Image Content Descriptors That Have the Highest Correlation with the Human Perceived Artisticity of the Paintings

Feature rank	Feature name	Correlation with artisticity
1	Fractal 12 (Fourier Wavelet)	0.353488
2	Edge Magnitude Histogram bin 7	0.327330
3	Fractal 11 (Fourier Wavelet)	0.322587
4	Fractal 12 (Wavelet Fourier)	0.321418
5	Fractal 12 (Chebyshev Wavelet)	0.320123
6	min (Chebyshev Fourier)	0.312625
7	Fractal 11 (Chebyshev Fourier)	0.309647
8	Fractal 10 (Fourier Chebyshev)	0.306816
9	Haralick Texture 24	0.306446
10	Fractal 12 (Wavelet)	0.298350
11	max (Wavelet Fourier)	0.295474
12	Fractal 13 (Fourier)	0.293891
13	Haralick Texture 20	0.293311
14	Fractal 13 (Edge Transform)	0.289046
15	Fractal 13 (Chebyshev)	0.288709
16	Tamura Texture 3 (Fourier)	0.287689
17	MultiScale Histogram bin 7	0.287142
18	Tamura Texture 3 (Wavelet Fourier)	0.286787
19	CombFirstFourMoments 28 (Chebyshev Wavelet)	0.285091
20	Fractal 13 (Edge Fourier Transform)	0.284705
21	Tamura Texture 3 (Fourier Wavelet)	0.283715
22	Fractal 9 (Fourier Wavelet)	0.283203
23	Fractal 11 (Wavelet Fourier)	0.282731
24	Fractal 11 (Fourier Chebyshev)	0.282698
25	Fractal 11 (Fourier)	0.280507
26	Fractal 10 (Chebyshev Fourier)	0.280247
27	Fractal 13 (Fourier Chebyshev)	0.278682
28	MultiScale Histogram bin 14	0.277341
29	MultiScale Histogram bin 22 (Wavelet)	0.276581
30	Fractal 7 (Fourier Wavelet)	0.275853
31	Fractal 10 (Fourier Wavelet)	0.275797
32	Fractal 6 (Fourier Wavelet)	0.275228
33	Fractal 8 (Fourier Wavelet)	0.275083
34	stddev (Wavelet Fourier)	0.274876
35	MultiScale Histogram bin 7 (Wavelet)	0.272794
36	color histogram bin 1	0.272679
37	Fractal 15 (Chebyshev Wavelet)	0.272217
38	Fractal 13 (Chebyshev Wavelet)	0.272150
39	Fractal 13	0.272143
40	Fractal 5 (Fourier Wavelet)	0.271670



Appendix Fig. 1. The average Fisher discriminant scores of the different numerical image content descriptors computed from the 15 runs. The figure shows that the fractal features are the most informative features.



Appendix Fig. 2. The number of image features showing the highest Pearson correlation with the human perceptions of the artisticity of the paintings. A brief description of these features is provided in Section 3.

The paintings used in the study and the correctness scores from Snapper et al. [2015] are provided in Tables II and III.

Appendix Table II. Paintings by Known Artists and Scores. The Computed Artisticity Is Determined by a Machine-Learning System Trained Using the Perceived Artisticity. The Computer Classification Is Determined By a Machine-Learning System Trained by the Painter Ground Truth (Professional Painter or Child/Animal)

% Identified as Artist	Computed artisticity	Computer classification	Artist name	Title of Work
94	60.63	Child/animal	Charles Seliger	Forest Echoes, 1961
87	61.81	Child/animal	Joan Mitchell	Hemlock, 1956
87	65.31	Artist	Clyfford Still	1945-R, 1945
87	52.72	Artist	Theodoros Stamos	Documenta II, 1959
83	57.69	Artist	Mark Tobey	New World Stage, circa. 1960
81	57.25	Artist	Ralph Rosenborg	Untitled (Floral Study), 1976
80	54.81	Child/animal	Sam Francis	Untitled, 1989
76	58.75	Artist	Kenzo Okada	Points No.19, 1954
75	65.81	Artist	Mark Rothko	Number 18, 1948
75	55.06	Child/animal	Hans Hofmann	The Climb, 1960
73	47.75	Child/animal	James Brooks	Boon, 1957
73	35.75	Artist	Mark Rothko	Untitled, 1948
72	36.12	Child/animal	Morris Louis	Addition V, 1959
71	55.94	Artist	Ralph Rosenborg	Autumn Landscape, 1974

(Continued)

Appendix Table II. Continued

% Identified as Artist	Computed artisticity	Computer classification	Artist name	Title of Work
65	37.06	Child/animal	Elaine de Kooning	On the Way to San Remo, 1967
63	40.81	Artist	Karel Appel	Untitled, 1960
63	41.81	Child/animal	Sam Francis	Tokyo Blue, 1961
62	41.19	Artist	Franz Kline	Untitled, 1958
61	58.43	Artist	Hans Hofmann	Fiat Lux, 1963
59	50.06	Artist	Gillian Ayres	Distillation, 1957
57	25.68	Artist	Sam Francis	Untitled
57	41.69	Artist	Sam Feinstein	Untitled
55	62.12	Artist	Helen Frankenthaler	Before the Caves, 1958
50	60.62	Artist	Philip Guston	For M, 1955
50	47.88	Artist	Hans Hofmann	Astral Nebula, 1961
44	45.00	Artist	Hans Hofmann	Laburnum, 1954
42	53.44	Artist	Cy Twombly	Nine Discourses on Commodus Part V, 1963
36	39.00	Artist	Joan Mitchell	Untitled, 1967
36	68.69	Child/animal	Hélène Hurot	D'après Sam Francis, 2007
12	37.25	Child/animal	Joan Mitchell	Pastel, 1990

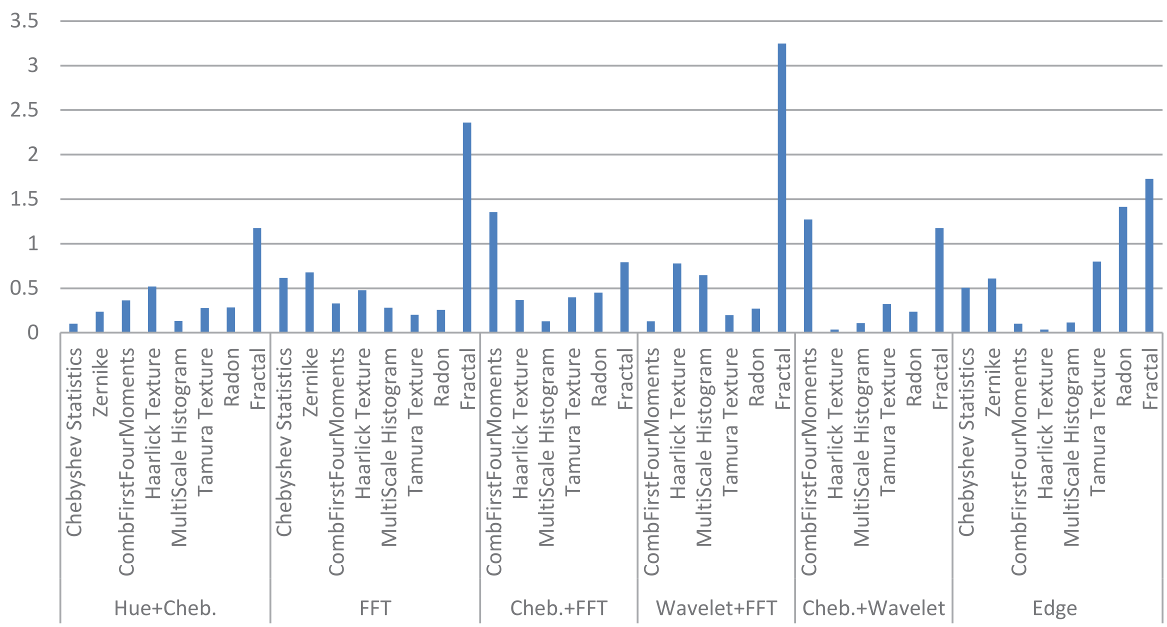
Appendix Table III. Paintings by Animal/Children and Their Scores

% Identified as Artist	Computed artisticity	Computer classification	Species, if Known
9	41.18	Child/animal	Animal
12	24.65	Child/animal	Animal, Elephant
13	46.44	Child/animal	Child, Kindergarten
15	61.25	Artist	Animal, Elephant
16	46.81	Child/animal	Animal, Elephant
17	12.00	Child/animal	Animal, Elephant
21	56.43	Artist	Animal, Monkey
25	67.61	Child/animal	Animal, Elephant
31	30.56	Child/animal	Animal, Monkey
31	44.50	Child/animal	Child, Age 2
31	64.56	Artist	Animal, Elephant
36	38.43	Child/animal	Child, Pre-K
39	61.94	Child/animal	Animal, Chimpanzee
40	47.62	Artist	Child, Kindergarten
40	37.31	Child/animal	Child, Age 4
49	52.81	Artist	Animal, Elephant
49	46.37	Artist	Animal, Orangutan
51	43.87	Child/animal	Animal, Monkey
51	33.63	Child/animal	Animal, Elephant
52	20.46	Child/animal	Animal, Chimpanzee
54	43.75	Child/animal	Child, Age 4
57	71.75	Child/animal	Animal, Chimpanzee
58	24.7	Child/animal	Animal, Elephant

(Continued)

Appendix Table III. Continued

% Identified as Artist	Computed artisticity	Computer classification	Species, if Known
59	54.11	Artist	Animal, Elephant
59	54.81	Child/animal	Animal, Gorilla
65	46.62	Artist	Animal, Gorilla
65	43.25	Child/animal	Child
69	66.06	Artist	Animal, Elephant
70	71.87	Child/animal	Child, Preschool
71	78.18	Child/animal	Animal



Appendix Fig. 3. The average Fisher discriminant scores of the numerical image content descriptors computed from the separation between the paintings perceived as the work of art and paintings perceived as the work of children or animals. Again, the figure shows the importance of fractal features.

REFERENCES

E. Cetinic and S. Grgic. 2013. Automated painter recognition based on image feature extraction. *ELMAR, 55th International Symposium*, 19–22.

D. Gabor. 1946. Theory of communication. *J. IEEE*, 93, 429.

R. M. Haralick, K. Shanmugam, and I. Dinstein. 1973. Textural features for image classification. *IEEE Trans Syst. Man Cybernet.* 6, 269.

I. Gradshtein and I. Ryzhik. 1994. *Table of Integrals, Series and Products*, Vol. 5. Academic Press, San Diego, CA 1054.

A. Hawley-Dolan and E. Winner. 2011. Seeing the mind behind the art: People can distinguish abstract expressionist paintings from highly similar paintings by children, chimps, monkeys and elephants. *Psychol. Sci.* 22, 435–441.

K. Jones-Smith and H. Mathur. 2006. Fractal analysis: Revisiting Pollock's drip paintings. *Nature*, 444, 9–10.

F. S. Khan, S. Beigpour, J. Van De Weijer, and M. Felsberg. 2014. Painting-91: A large scale database for computational painting categorization. *Mach. Vision Appl.* 25, 6, 1385–1397.

P. Kammerer, M. Lettner, E. Zolda, and R. Sablatnig. 2007. Identification of drawing tools by classification of textural and boundary features of strokes. *Pattern Recogn. Lett.* 28, 710–718.

D. Keren. 2002. Painter identification using local features and naive Bayes. In *Proceedings of the 16th International Conference on Pattern Recognition*.

M. Kim and J. Kim. 2013. Quantitative analysis of artists' characteristic styles through biologically-motivated image processing techniques: Uncovering a mentor to Johannes Vermeer. In *HCI International 2013-Posters' Extended Abstracts*. Springer, Berlin, 258–262.

S. Kroner and A. Lattner. 1998. Authentication of free hand drawings by pattern recognition methods. In *Proceedings of the 14th International Conference on Pattern Recognition*. 462–464.

J. Li and J. Z. Wang. 2004. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. Image Process.* 13, 3, 340–53.

J. Li, L. Yao, E. Hendriks, and J. Z. Wang. 2012. Rhythmic brushstrokes distinguish Van Gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 6, 1159–1176.

J. S. Lim. 1990. *Two-Dimensional Signal and Image Processing*. Prentice Hall, New Haven, CT.

S. Lyu, D. Rockmore, and H. Farid. 2004. A digital technique for art authentication. In *Proc. Natl. Acad. Sci.* U.S.A. 101, 17006–17010.

B. B. Mandelbrot. 1982. *The Fractal Geometry of Nature*. Freeman, San Francisco, CA.

D. Mattison. 2004. Looking for good art image retrieval. *Searcher*, 12, 9, 8–19.

K. Minseo and K. Jeounghoon. 2013. Quantitative analysis of artists' characteristic styles through biologically-motivated image processing techniques: Uncovering a mentor to Johannes Vermeer. *Commun. Comput. Inform. Sci.* 374, 258–262.

J. Nissel, A. Hawley-Dolan, and E. Winner. 2015. Can young children distinguish abstract expressionist art from superficially similar works by preschoolers and animals? *J. Cogn. Dev.*, In Press.

N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, and I. Goldberg. 2008. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recogn. Lett.* 29, 1684.

V. S. Ramachandran and W. Herstein. 1999. The science of art: A neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–51.

P. Roussopoulos, D. Arabadjis, M. Exarhos, and M. Panagopoulos. 2010. Image and pattern analysis for the determination of the method of drawing celebrated thera wall-paintings circa 1650 B.C. *ACM J. Comput. Cultural Heritage*, 3, 2.

C. Tsai. 2007. A review of image retrieval methods for digital cultural heritage resources. *Online Inform. Rev.* 31, 2, 185–198.

L. Shamir. 2008. Evaluation of face datasets as tools for assessing the performance of face recognition methods. *Int. J. Comput. Vis.* 79, 3, 225–230.

L. Shamir, N. Orlov, D. M. Eckley, T. Macura, J. Johnston, and I. Goldberg. 2008a. Wndchrm—An open source utility for biological image analysis. *BMC Source Code Biol. Med.* 3, 13.

L. Shamir, N. Orlov, and I. Goldberg. 2009a. Evaluation of the informativeness of multi-order image transforms. In *Proceedings of the International Conference on Image Processing Computer Vision and Pattern Recognition.* 37–42.

L. Shamir, S. M. Ling, W. W. Scott, A. Bos, N. Orlov, T. Macura, D. M. Eckley, and I. G. Goldberg. 2009b. Knee X-ray image analysis method for automated detection of osteoarthritis. *IEEE Trans. Biomed. Eng.* 56, 407–415.

L. Shamir, S. Ling, W. Scott, M. Hochberg, L. Ferrucci, and I. Goldberg. 2009c. Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarthrit. Cartilage* 17, 1307–1312.

L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg. 2010. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Trans. Appl. Percept.* 7, 2, 8.

L. Shamir. 2012. Computer analysis reveals similarities between the artistic styles of van gogh and pollock. *Leonardo* 45, 2, 149–154.

L. Shamir and J. Tarakhovsky. 2012. Computer analysis of art. *ACM J. Comput. Cultural Heritage*, 5, 2, 7.

L. Shamir. 2011. A computer analysis method for correlating knee x-rays with continuous indicators. *Int. J. Comput.-Assist. Radiol. Surg.* 6, 5, 699–704.

L. Shamir, N. Orlov, D. M. Eckley, T. Macura, J. Johnston, and I. Goldberg. 2013a. WND-CHARM: Multi-purpose image classifier. *The Astrophysics Source Code Library*, ascl:1312.002.

L. Shamir, A. Holincheck, and J. Wallin. 2013b. Automatic quantitative morphological analysis of interacting galaxies. *Astron. Comput.* 2, 67–73.

L. Shamir, C. Yerby, R. Simpson, A. Von Benda-Beckmann, P. Tyack, F. Samarra, P. Miller, and J. Wallin. 2014. Classification of large acoustic datasets using machine learning and crowdsourcing - application to whale calls. *J. Acoust. Soc. Am.* 135, 2, 953–962.

L. Shamir. 2015. What makes a Pollock Pollock: A machine vision approach. *Int. J. Art Technol.* 8, 1, 1–10.

L. Snapper, C. Oranc, A. Hawley-Dolan, J. Nissel, and E. Winner. 2015. Your kid could not have done that: Even untutored observers can discern intentionality and structure in abstract expressionist art. *Cognition*, 137, 154–165.

H. Tamura, S. Mori, and T. Yamavaki. 1978. Textural features corresponding to visual perception. *IEEE Trans Syst. Man Cybernet.* 8, 460.

R. P. Taylor, A. P. Micolich, and D. Jonas. 1999. Fractal analysis of Pollock's drip paintings. *Nature*, 399, 422.

R. P. Taylor, et al. 2007. Authenticating pollock paintings using fractal geometry. *Pattern Recogn. Lett.* 28, 695–702.

R. P. Taylor, A. P. Micolich, and D. Jonas. 2002. The construction of Jackson Pollock's fractal drip paintings. *Leonardo*, 35, 2, 203–207.

R. P. Taylor, A. P. Micolich, and D. Jonas. 2006. Fractal analysis: Revisiting Pollock's drip paintings. *Nature*, 444, E10–E11.

M. Teague. 1980. Image analysis via the general theory of moments. *J. Opt. Soc. Am.* 70, 920.

K. Varnedoe. 2006. *Pictures of Nothing: Abstract Art Since Pollock*. Princeton University Press, Princeton, NJ.

C. M. Wu, Y. C. Chen, and K. S. Hsieh. 1992. Texture features for classification of ultrasonic liver images. *IEEE Trans. Med. Imag.* 11, 2, 141–152.

S. Zeki and J. Nash. 1999. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press, Oxford, 415.

Received June 2015; revised November 2015; accepted March 2016