Computers & Graphics 33 (2009) 484–495

Contents lists available at ScienceDirect

Computers & Graphics

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

Computational Aesthetics 2008

Categorizing art: Comparing humans and computers

Christian Wallraven [a,*], Roland Fleming [a], Douglas Cunningham [b], Jaume Rigau [c], Miquel Feixas [c], Mateu Sbert [c]

[a] *MPI for Biological Cybernetics, Tübingen, Germany*
[b] *WSI-GRIS, University of Tübingen, Germany*
[c] *University of Girona, Spain*

ARTICLE INFO

Article history:
Received 1 December 2008
Received in revised form
24 February 2009
Accepted 13 April 2009

Keywords:
Computational aesthetics
Multi-dimensional scaling
Computer vision
Human studies

ABSTRACT

The categorization of art (paintings, literature) into distinct styles such as Expressionism, or Surrealism has had a profound influence on how art is presented, marketed, analyzed, and historicized. Here, we present results from human and computational experiments with the goal of determining to which degree such categories can be explained by simple, low-level appearance information in the image. Following experimental methods from perceptual psychology on category formation, naive, non-expert participants were first asked to sort printouts of artworks from different art periods into categories. Converting these data into similarity data and running a multi-dimensional scaling (MDS) analysis, we found distinct categories which corresponded sometimes surprisingly well to canonical art periods. The result was cross-validated on two complementary sets of artworks for two different groups of participants showing the stability of art interpretation. The second focus of this paper was on determining how far computational algorithms would be able to capture human performance or would be able in general to separate different art categories. Using several state-of-the-art algorithms from computer vision, we found that whereas low-level appearance information can give some clues about category membership, human grouping strategies included also much higher-level concepts.

1. Introduction

What is art?—What constitutes a work of art?—How can we separate ‘good’ from ‘bad’ art (and do these categories make sense)? These questions have been debated in art history time and time again—perhaps culminating in the definition of playwright Tom Stoppard who said of modern art: ‘It is not hard to understand modern art. If it hangs on a wall it is a painting, and if you can walk around it, it is a sculpture.’ Nevertheless, art is one of the most important driving forces behind civilization as a whole, as well as cultural, philosophical, and scientific developments in particular. There have been many attempts at structuring art both in terms of contents as well as in terms of historical developments. Art historians commonly divide Western pictorial art into distinct stylistic classes, movements, or periods. In many cases, the attribution of an artist or artwork to a particular class or movement is uncertain, and the number of art periods, as well as their duration and definition is a matter of ongoing debate. In recent times, the clear-cut definition of a few major art movements (such as Surrealism, Expressionism, etc.) that encompass the artistic production of a time period has become problematic, as art has become focussed much more on individual artists and their aesthetics.

Nevertheless, the categorization of artworks into periods or movements has had a profound influence on how art is presented, analyzed, and appreciated. How can one achieve such a categorization—how can different works of art and different painters be attributed to a specific art period? Given a quick glance at a painting, knowledgeable experts can identify the art period (as well as the painter and the title of the work) based on several types of information which need to be integrated:

- Low-level pictorial information: technique, thickness of brush strokes, type of painting material (oil, acrylic, etc.), color composition of the scene
- Mid-level content information: specific objects or scenes that are depicted, type of painting or subject (landscape painting, portrait, etc.)
- High-level background information: knowledge about specific historical events, knowledge about artists and art periods in general.

Based on this coarse division of the large interpretation space of art, different clusters can be formed and the history of art can

[*] Corresponding author.
E-mail address: christian.wallraven@tuebingen.mpg.de (C. Wallraven).

0097-8493/$ - see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.cag.2009.04.003

be structured by assessing similarities within a large selection of pictorial art. Following Little [1], we can distinguish between periods defined by a trend within the visual arts, a broad cultural trend, an artist-defined movement, and a retrospectively applied label. These different types of art periods, however, are usually found to be defined in quite abstract, art historical terms emphasizing high-level, expert-defined concepts.

The first main question which this paper addresses is that: to what degree can art periods be determined based on low-level and mid-level information alone? The answer to this question will be therefore the degree with which *art periods are perceptually founded*. Similar to how art experts would assess the category of images by identifying similarities and dissimilarities, we asked naive participants to group several hundreds of images into categories based on the artistic style. The data which we gathered gives insights not only into perceptual and cognitive qualities of art but also allows us to investigate the perceptual validity of the canonical art periods.

Going beyond human categories, the second main question addressed in this paper was: can we also model the results computationally? In other words, can we find a computer algorithm which takes as input the same paintings given to our participants and which then clusters the paintings in a similar fashion? As today's computer vision algorithms are only beginning to be able to work on mid-level content information, this part of the study mostly focuses on how low-level, pictorial information determines (perceptual) art periods.

2. Related work

As this study is rooted both in perceptual and computational contexts, the following reviews the most relevant work in these two research areas.

2.1. Perception of art

The study of art in perceptual terms enjoys a long tradition in perception research leading from the interests of the early perceptual psychophysicists in the aesthetic experience [2] to eye-tracking studies on the experience of modern, so-called indeterminate art [3,4]. Perhaps one of the most influential accounts is the book by Arnheim [5] in which he develops a theory of art perception based on Gestalt principles. He considered art—similarly to perception—to be based on the 'Gestalt' of colors and shapes. In addition, many of the fundamental Gestalt principles also had an influence on and were influenced by art and art theory—especially for the composition of paintings. In Pinna [6] which calls for a 'visual science of art', possible interesting analysis factors for such a science are presented in terms of antinomies: realistic versus unrealistic representation, good-formation versus de-formation, beauty versus ugliness, form versus content, and real versus illusory. Even in the neurosciences, perception of art has gained an interest, most notably in the recent book by Zeki [7] exploring the connections between physiological structures of the visual system and art. As we cannot give an exhaustive overview of the many attempts to connect vision science and art, we refer the reader to two special issues of the journal 'Spatial Vision' [8,9], a short review paper by Spillmann [10], and the book by Livingstone [11].

Going beyond the visual analysis of art, what is perhaps the most general framework for describing the aesthetic experience of art was presented in Leder et al. [12]. The model they develop in this paper consists of five stages: perceptual analyzes, implicit memory integration, explicit classification, cognitive mastering and evaluation. The output of the model consists of both aesthetic pleasure (an emotive component) and aesthetic judgment (a cognitive component). These five stages together with the processing steps encompass and extend the three levels of analysis presented in the introduction. In the present study, we are more interested in the categorization, or classification of artworks rather than trying to capture the full aesthetic experience. In this respect, our experiments will only deal with a subset of the processes mentioned in [12].

The formation of categories is one of the most fundamental aspects of perception and perceptual learning [13] and has been studied extensively in cognitive and perceptual psychology, as well as in neuroscience. Recent work emphasizes especially the role of *similarity judgments* in creating categories [14,15]. That is, in order to define categorical relationships the similarity of the items is taken into account. The question of how naive observers are able to group different works of art into consistent, historically correct categories has been considered in Hasenfus et al. [16]. In this study, observers had to group baroque, neoclassical, and romantic cross-medial works of art (including paintings, architecture, and music) into groups. On average, nine groups were created based on three different strategies: emotion-based, content-based, and style-based. Agreement among participants was rather high and discrimination of the three different art periods based on this data yielded already around 60% accuracy showing that certain styles can be picked up even across different media. In a recent study [17], the question of expertise in art appreciation of contemporary art was investigated. Using a natural grouping task and correspondence analysis on a set of ten paintings, they found that experts created many more groups based on a larger variety of analysis factors than non-experts. Additionally, experts tended to create groups based on style, whereas non-experts mentioned more emotional attributes. Beyond this, only few differences between the groups in terms of analysis attributes were found.

In this paper, we are mainly interested in how naive observers might define categories based on the artistic style—that is, we tackle one specific part of the aesthetic experience in a complementary approach to that taken in Augustin and Leder [17]. Whereas they were interested in the overall aesthetic appreciation of artworks, here, we directly ask participants to consider only the artistic style when making their decisions about the categories. Additionally, our study spans a much larger space of artworks than just contemporary art in order to uncover broad dimensions along which people might consider and judge art.

2.2. Computational analysis of art

Computational analyzes of art have mostly dealt with classification problems—that is, how can we extract features and train classifiers in order to, for example, label an image as belonging to Van Gogh, or as belonging to the Expressionists. Inspiration for features sometimes comes from perceptual or physiological sources as in the study by Redies [18] which used natural image statistics—which the human visual system is highly tuned to—to analyze different works of art. Mostly, however, features are taken from the computer vision domain and are primarily restricted to low-level pictorial features (the lowest level of interpretation mentioned in the introduction). One example of this approach is Cutzu et al. [19] in which the authors proposed a framework for distinguishing paintings from photographs based on color edges, spatial variation of colors, number of unique colors, and pixel saturation. No single feature could do the task, but a combination of features was shown to yield good discrimination performance. In Marchenko et al. [20] concepts

Table 1
Art periods (and abbreviations) and representative artists chosen for our experiments.

Art period	Artist				
Gothic (Goth)	Bondone	Bosch	di Buonisegna	Van Eyck	Weyden
Renaissance (Rena)	Altdorfer	del Sarto	Botticelli	Duerer	Signorelli
Baroque (Baro)	Caravaggio	Poussin	Rembrandt	Reni	Velazquez
Rokoko (Roko)	Fragonard	Nattier	Pesne	Rigaud	Watteau
Classicism (Clas)	David	Dietrich	Ingres	Kauffmann	Mengs
Romanticism (Roma)	Friedrich	Delacroix	Gericault	Spitzweg	von Blechen
Realism (Real)	Courbet	Menzel	Millet	Repin	Thoma
Impressionism (Impr)	Cézanne	Liebermann	Manet	Monet	Sisley
Expressionism (Expr)	Macke	Marc	Nolde	Pechstein	Schiele
Surrealism (Surr)	Chirico	Duchamp	Ernst	Magritte	Tanguy
Postmodern Art (Post)	de Maria	Hesse	Malevich	Newman	Stella

based on color temperature (warm versus cold), color palette (primary, complimentary) and color contrasts (light-dark) derived from paintings are introduced. These resulted in good classification performance for modern versus medieval artwork.

In Marchenko et al. [21] the feature concepts were extended to include texture features as well as brush stroke analyzes and even an annotation with higher-level concepts. These were used to train expert systems to annotate modern and medieval artworks resulting in increased classification and annotation performance. Similar to their work, Leslie et al. [22] go beyond low-level feature analysis and propose a method which assigns visual concepts (such as the coloring or the brush stroke used) to parts of a painting. Using an ontology-based disambiguation method, their approach outperforms other techniques relying solely on classification of low-level features.

All of the mentioned approaches mainly rely on lower-level features to analyze and categorize artworks. The book by Leyton [23] proposes a mathematical framework for understanding the *structure* of paintings, i.e., the spatial organization and composition of the different elements and how they serve to create an aesthetic ‘whole’—although this is perhaps one of the most ambitious attempts at a quantitative description of art, so far no automatic algorithms exist which can perform the tasks required by the framework of Leyton [23].

In Datta et al. [24], 56 computational features were used not for classification but for modeling human aesthetic scores on a database of photographs. These features comprised color, texture, as well as shape cues, and also included higher-level scores based on the photograph’s similarity to a ‘standard database’ of images. The latter turned out to be a crucial ingredient to modeling the aesthetic scores showing how prior knowledge is a crucial ingredient in aesthetic judgments [12]. Similarly, in Rigau et al. [25], measures from information theory (based on and extended from Rigau et al. [26]) were used to group the works of Van Gogh into periods that corresponded surprisingly well to the different phases and styles he used and developed.

3. Stimuli

A total of 11 art periods were selected for investigation based on various sources. We tried to select art periods at a sufficiently broad level, which would also include larger periods of time in art history. Using this criterion, we came up with 11 art periods: Gothic, Renaissance, Baroque, Rokoko, Classicism, Romanticism, Realism, Impressionism, Expressionism, Surrealism, and Postmodernism.

For each of these art periods we identified five major artists whose work mainly can be attributed to the time period in question—these are listed in Table 1.[1] Then, for each of the artists, we chose 10 representative paintings with which we tried to cover the artistic spectrum. As we were not interested in a familiarity task, we tried to choose less well-known paintings for the more famous artists in our database. All images were taken from www.prometheus-bildarchiv.de—an online library which provides access to a variety of large, online collections of art. Fig. 1 shows example paintings from the 11 different art periods.

The 550 paintings were split into two datasets in order to cross-validate our findings. Each dataset therefore contained 5 images $\cdot$ 5 artists $\cdot$ 11 art periods $=$ 275 paintings. Experiments 1 and 3 were run on the first dataset, whereas Experiment 2 was run on the second dataset.

4. Experiments

In order to assess ‘naive’ judgments of art, we conducted three different free-sorting experiments with similar experimental design and analyzes. In all three experiments, participants had to sort the 275 images of each dataset into a number of coherent clusters based on painting style. In order to validate the naivety of our participants, we first ran a familiarity experiment for all three participant groups, which is described first in the following. The goal of Experiments 1 and 2 specifically was to compare free sorting of a large number of paintings on the two different datasets. Participants were asked to sort the 275 images into any number of clusters. The goal of Experiment 3 was to see how results would change if we constrained the free-sorting task to yield exactly 11 clusters. In addition, we collected data from a questionnaire in the end, in which we debriefed participants about the strategies they used in the free-sorting task, as well as their familiarity with the different art periods we selected.

Overall, we had 40 participants for our three experiments: two groups of 15 participants took part in the first two experiments and 10 participants in the third. We specifically selected participants who indicated to be non-experts in art, art history, or art practice.

4.1. Familiarity experiment

4.1.1. Experimental design

Before the main experiment took place, we first gathered familiarity ratings for all images from each participant. This was

[1] Our mapping of artist to art period (as our particular choice of art periods) is in some cases debatable—in these experiments, however, the focus will be more on the participants’ sorting behavior.

Fig. 1. Example images for each art period from Gothics to Postmodernism.

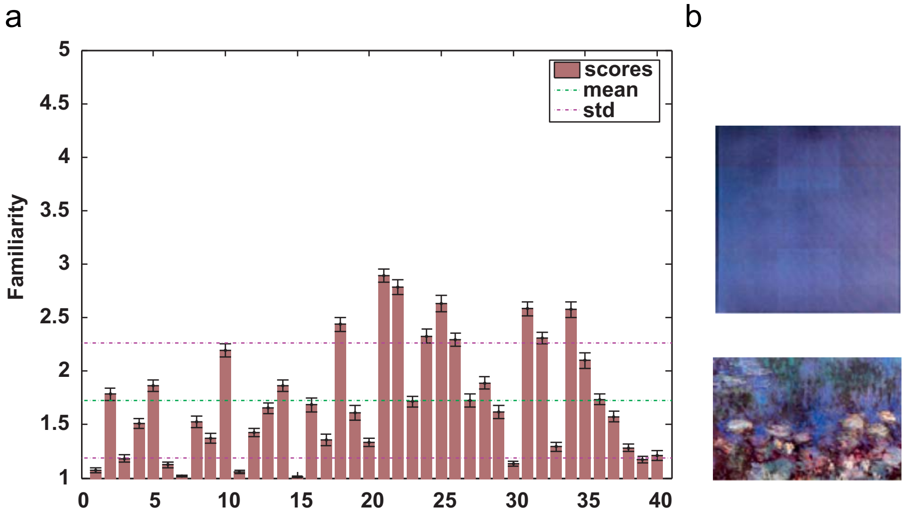

Fig. 2. (a) Familiarity ratings averaged across the two datasets for the 40 participants who took part in Experiments 1–3 (error bars show SEM). The plot also shows the mean rating value and its standard deviation. (b) The least (upper, by Newman) and most (lower, by Monet) well-known painting.

done not only to make sure that familiarity with a subset of the paintings would not interfere too much with the grouping, but also to familiarize participants with all paintings once. The data from all three experiments will be analyzed here.

For this task, participants viewed all 275 images in randomized order on a standard 21″ CRT set to 1280 × 1024 pixel resolution. Images were rescaled to a maximum width of 1024 and a maximum height of 768 while preserving their aspect ratio. They were shown centered on the screen and participants were instructed to rate the familiarity of each image on a scale of 1 (not familiar at all)–5 (extremely familiar). In order to properly anchor the scale, every number was also explained with a short sentence:

- 1: Totally unfamiliar. You are sure that you have not seen this image before, and have no idea who might have created it.
- 2: Moderately unfamiliar. You feel that you probably have not seen this particular image before, although you may have seen similar ones.
- 3: Uncertain. You are unsure whether you have seen this particular image before, but are confident that you have seen similar ones.
- 4: Moderately familiar. You think you probably have seen this image before, or have seen many similar ones and can guess who might have created it.
- 5: Extremely familiar. You are certain that you have seen this particular image before. You feel you know the image well and may know details about the image, such as the artist, location or title.

Participants could set their own pace for this task and finished in ≈ 13 min on average for all 275 images.

4.1.2. Analysis

Familiarity ratings were averaged across all paintings. We also looked for ‘outliers’, i.e. paintings which received consistently higher ratings. Familiarity ratings for all three experiments overall were very low—on average participants' familiarity with all paintings was only 1.7 (this is in between the lowest and the second-lowest category) (see Fig. 2a). The maximum average value (averaged across all 40 participants) for a single painting was 2.8, the minimum value was 1.0—these images are shown in Fig. 2b. An analysis of variance (ANOVA) across the experiments revealed a slight influence of the participant group. Participants in Experiment 2 had a slightly larger overall familiarity (1.9) with the paintings than participants in Experiment 1 (1.4). We will comment on this difference in the discussion of the questionnaire in Section 4.6.

Taken together this data from 40 participants indicates that—even though some paintings received the maximum rating of 5 from a participant—*overall* familiarity with the stimulus set was very low, such that we can exclude any influence of this factor on the data from the main sorting experiments.

4.2. Free-sorting experiments

4.2.1. Experimental design

The main experiments followed immediately after the rating task. Printouts of the paintings were shuffled and spread out on a large table. Participants were then instructed to group the printouts into clusters according to *painting style or art period*. They were explicitly asked *not* to group according to image content (still life, portrait, etc.), if possible. For the first and second experiment, we did not constrain the number of clusters with the only difference being the two datasets that were used. Comparing the data from these two experiments thus will yield insights into how valid the findings on our particular subset of images really are.

For the third experiment, we instructed participants to group all 275 images of dataset 1 into exactly 11 clusters. This was done in order to investigate whether specific grouping strategies would change if participants knew beforehand how many style-based clusters we would be expecting. For all three experiments, the number of paintings per cluster could be set freely (therefore leading to unevenly-sized clusters).

After the experiment we gathered some more information in a questionnaire—these included specific strategies for solving the task as well as subjective ratings of the difficulty of the experiment, participants' expertise in art and familiarity with our chosen art periods. Overall, participants took ≈ 80 min on average for the main task and ≈ 10 min to fill out the questionnaire.

4.3. Analysis methods

For the main experiments, the raw data for the first set of analyzes consisted of determining which image was put into which cluster for each participant. From this data we determined three measures:

- Number of clusters for this participant (for the Experiment 3, this value was, of course, always 11).
- Consistency score for each artist c_{artist}: regardless of our attribution of artists to art periods, the only 'true' labels in our two datasets are the five images which belong to one artist. In order to determine how well participants did in the grouping task, this score therefore counts the number of clusters n_{artist} across which the paintings of one artist are spread. The higher the consistency score $c_{artist} = (4 - (n_{artist} - 1))/4$ (with $0 \leq c_{artist} \leq 1$), the more paintings of one painter were put into the same cluster.
- Consistency score for each art period c_{period}: by averaging all five c_{artist} we can determine how consistently one particular art period was treated.

For the next step, we converted the data into a similarity matrix: for each image i we increased a counter in cell (i,j) if i was in the same cluster as j. Averaging all matrices across participants yields an average similarity matrix. Matrices for Experiments 1 and 3 are shown in Fig. 4. Such a matrix enables two types of analyzes: (i) multi-dimensional scaling (MDS), which projects the data into a lower-dimensional space enabling us to investigate the criteria along which similarity was judged; and (ii) clustering analysis which tries to predict distinct, consistent clusters allowing us to see which items would be grouped together.

MDS refers to a family of algorithms which operate on proximity data taken between pairs of objects [27]. The output is a configuration of objects embedded in a multidimensional space. Psychologists have used MDS to explore perceptual representations of different visually (e.g., [28]) presented object sets. The technique has also found a large following in domains such as knowledge mapping and marketing because it allows for the identification of psychological dimensions of stimulus variation (e.g., dimensions along which buyers differentiate amongst competing products) and quantification of perceptual distances between stimuli (e.g., how closely related fields of research are). In our experiments, we use metric (classical) MDS.

Clustering analysis was also applied to the output of the similarity data. Here, we used two types of clustering: k-means clustering which splits feature vectors into k clusters based on Euclidean distance between vectors, and hierarchical clustering which splits data from a similarity matrix into a set of hierarchically organized clusters.

In the following, we will first discuss the raw data analyzes for Experiments 1–3 and then discuss the findings from MDS and cluster analyzes.

4.4. Raw data analyzes

4.4.1. Experiment 1

Number of clusters: We found a large variety across participants ranging from 7 to 41 clusters. The average number of clusters was 17.3±2.3 SEM (yielding an average number of ≈ 16 images per cluster). Just looking at this number we already can tell that participants on average did not try to put each of the 55 artists into a separate cluster but rather tried to build larger groups. In order to go into more details, we need to look at the consistency scores, however.

Average consistency scores: The average consistency score over all participants was $c_{artist} = 0.61 \pm 0.02$ SEM varying from 0.45 to 0.78. Whereas this is far from perfect performance, a simulation of a group of 'random participants' who would create the same number of clusters, but with a random clustering of images, yields $c_{artist} = 0.22 \pm 0.02$ SEM.[2] This result shows that participants were significantly better than chance at creating clusters which consistently contained paintings from the same artist.

One of the disadvantages of our measure is that a strategy of randomly putting paintings into very *few* clusters would also result in a higher consistency score. Thus it could be that participants who made fewer clusters also scored higher on average only for this reason. A correlation analysis between number of clusters and consistency score, however, revealed only a moderately strong (but statistically significant) correlation of $r^2 = 0.44$. Participant 14 had a larger number of clusters than participant 15, for example, but at the same time had a significantly *higher* consistency score showing that it cannot be the number of clusters alone which determines participants' consistency. Additionally, a random participant with the fewest number of clusters observed in the experiment (7) corresponds to $c_{artist} = 0.43$. The participant in question, however, had a considerably larger consistency score of 0.78 showing again that the clustering strategy was far from random.

Consistency scores for artists and art periods: Fig. 3a plots c_{artist} broken down by art period and artist as a color-coded matrix. Here, we can see a large degree of variation: some artists are clearly treated more consistently than others. For example, Bondone, Di Buonisegna, Pechstein, Newman, and Stella were grouped most often into the same cluster, whereas Courbet and Thoma were the least consistently grouped artists.

A closer look at the clusters shows that participants often had no trouble identifying older versus Postmodern art (for example, Christian motifs versus fully abstract art) which explains this finding. Averaging across artists to yield c_{period} we find that participants were least consistent for Realism and most consistent for Postmodernism, Expressionism, and Gothic.

Similarity matrix: The full pattern of results can be seen in the similarity matrix in Fig. 4a which plots the average cluster

[2] Each random participant was averaged over 20 runs to gather more robust statistics.

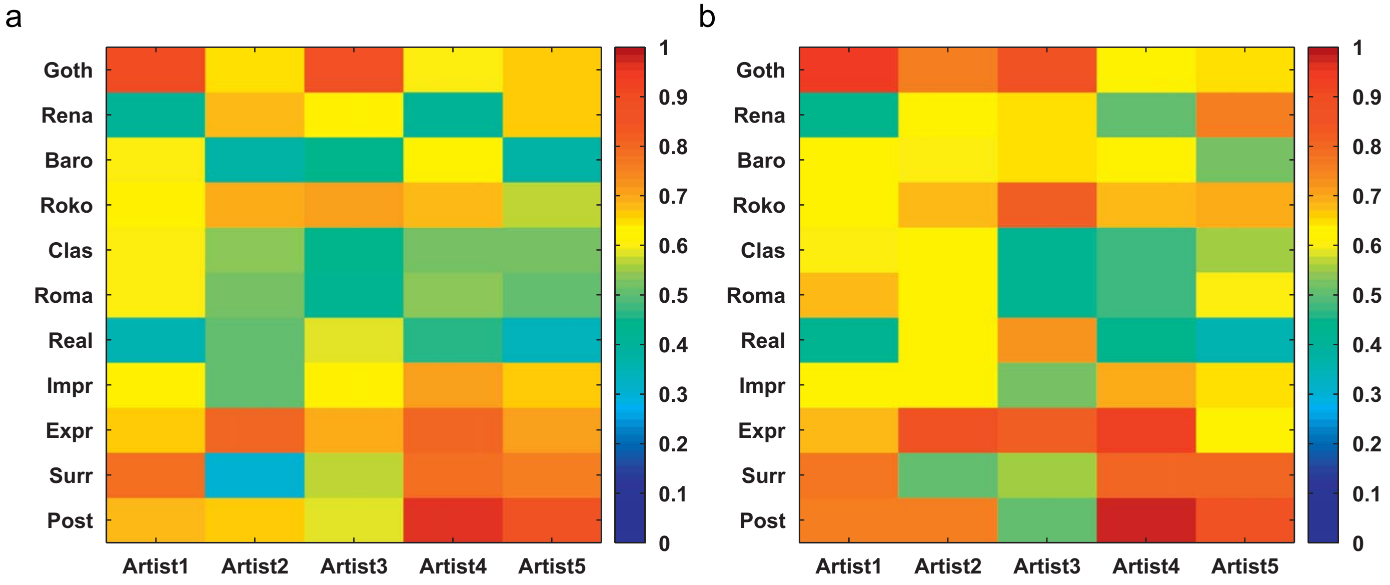

Fig. 3. Consistency matrices showing all c_{artist} for (a) Experiment 1, (b) Experiment 2. Each entry in the matrix corresponds to the entry in Table 1 such that 'Artist 1' for Gothics would be Bondone and 'Artist 3' for Expressionism would be Nolde, for example.

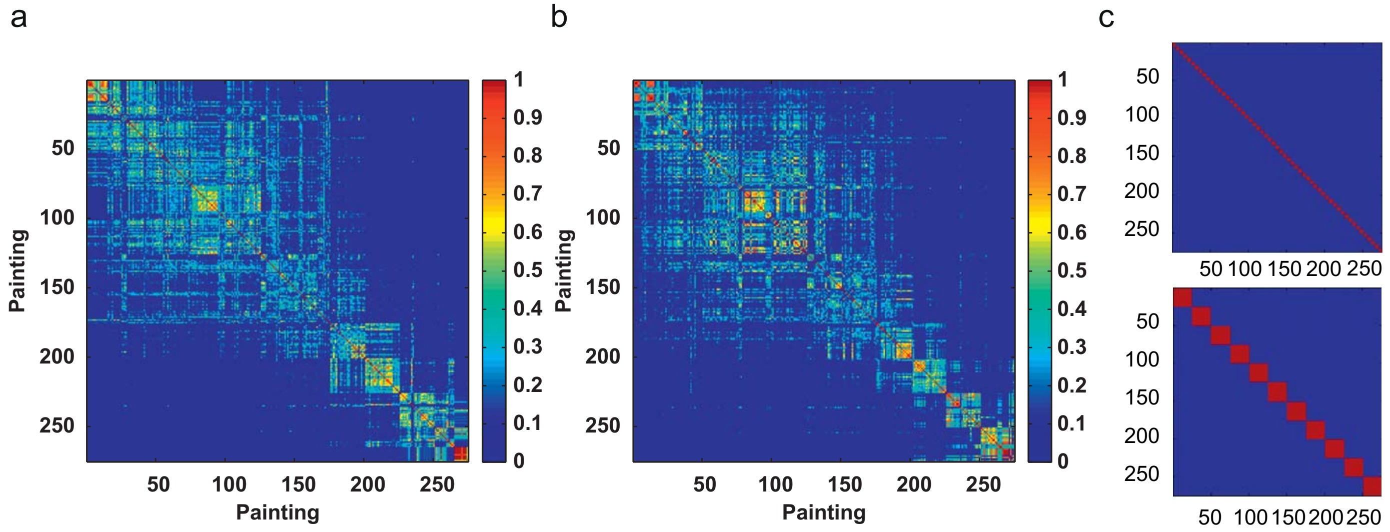

Fig. 4. Similarity matrices for (a) Experiment 1, (b) Experiment 3, (c) ideal grouping according to artists or periods. Note how (a) and (b) are much more similar to the ideal grouping according to periods. Art periods proceed from old (Gothics) to new (Postmodernism) from top to bottom and from left to right.

similarity between all 275 images. At the largest level, two large clusters emerge as square patterns—these correspond to earlier and modern art periods. A closer look shows that a clear split between the two square patterns occurs at the Impressionism period. Interestingly, this corresponds to an important transition in the development of modern art, when artists moved away from realism, and experimented with new painting techniques resulting in thicker, broader brush strokes, vivid colorings, and abstracted shading patterns. On average, these patterns thus show that participants were well aware of the historic turn from increasing realism to more abstract art[3]—a non-trivial result given that all participants were non-experts!

The consistent art periods show up as small, yellow-to-red square patterns in the matrix. Fig. 4c also shows how the pattern should look like if participants were grouping by artist or by (our notion of) art period: for the first possibility, the similarity matrix should consist of 55 5 × 5 blocks, whereas for the second, it should consist of 11 25 × 25 blocks. Comparing these ideal patterns to the experimental similarity matrix clearly shows that participants grouped by period rather than by artist: especially the Gothics period. Finally, the matrix also can give insights into the most common confusions that occur. Some examples of this include: some of Duchamp's paintings were frequently put into other categories; Van Eycks portraits were rarely in the earliest art clusters; paintings by Newman and Stella were often confused; etc.

4.4.2. Experiment 2

Number of clusters: In this experiment, we observed on average 19.3 ± 2.4 SEM clusters. A post-hoc *t*-test comparing individual results across the two experiments did not find any differences

[3] Note, that this statement only applies to the general trends in art—there are, of course, examples of realism in modern art.

($t(28)=0.62, p>0.05$) in the number of clusters created. This is an indication that the two datasets at least afforded similarly sized groupings.

Average consistency scores: Average consistency across all participants was 0.62 ± 0.03 SEM ranging from 0.44 to 0.77. A *t*-test comparing these scores across Experiments 1 and 2 again did not find any significant differences ($t(28)=0.33, p>0.05$) showing that participants seemingly relied on very similar grouping strategies in terms of attribution of artists to clusters.

Consistency scores for artists and art periods: Fig. 3b shows the data for c_{artist} broken down by art period and artist. Here, we find a slight deviation in the scores for one artist which a planned post-hoc comparison supports: the Surrealist artist Duchamp was grouped significantly more consistently in this experiment compared to the previous one. A closer look at the paintings in the two datasets reveals a particularly unfortunate split of the 10 paintings for this artist: in Experiment 1, two paintings are from very different periods in his oeuvre, whereas the paintings in Experiment 2 are more similar. None of the results for the other artists, however, reaches statistical significance. Overall, this shows that although our measure seems sensitive enough to detect potential discrepancies, participants grouped the two datasets very consistently.

Similarity matrix: As this similarity matrix is the 'raw data' for our experimental analyzes, we conducted a correlational analysis to check whether the two matrices from Experiments 1 and 2 yield similar patterns. The correlation coefficient was $r^2=0.64$ and highly significant thus showing that the degree of similarity between the two experiments was extraordinarily high. As the results for this dataset so far have been virtually identical to the results of Experiment 1, all further analyzes from above also hold true here.

4.4.3. Experiment 3

Average consistency scores: On average $c_{artist}=0.65\pm0.03$ SEM which is a little higher than that of Experiments 1 and 2. A post-hoc *t*-test comparing the results across participants, however, did not yield significant differences ($t(23)=0.3, p>0.05$). The slight increase of 0.04 on average could also simply be explained by the increase in the consistency score for random participants who all create 11 clusters: in this case $c_{artist}=0.26\pm0.02$ SEM (compared to $c_{artist}=0.22\pm0.02$ SEM for Experiment 1, for example). Thus, overall it seems that participants did not find a better, more consistent strategy when they were forced to use 11 clusters. This might seem surprising at first, but it becomes understandable if one considers that participants did not have any art expertise that might have helped them in creating more consistent clusters given prior information of 11 distinct periods/styles in the data.

Average consistency scores and similarity matrix: The correlation analyzes between Experiments 1 and 3, and between Experiments 2 and 3 yield $r^2=0.74$ and 0.59, respectively. Although slightly higher, the correlation between Experiments 1 and 3 is not significantly different from the one between Experiments 1 and 2 thus showing that the overall raw data is very similar in all experiments.

4.5. MDS and cluster analyzes

The previous results have shown that participants seemingly used very similar strategies for clustering the two datasets into categories. In the following, we will use MDS and cluster analyzes to investigate what kind of dimensions or features they might have used. All analyzes were done for all three experiments separately—not surprisingly given the similarity across experiments, results were virtually identical and thus are not reported separately here.

Table 2
Results of dimension analysis following MDS analysis of the raw similarity matrices of Experiments 1–3.

Dimension	Experiments 1–3
1	Older realistic motifs ↔ Surr, Expr
2	Perspective (flat ↔ open)
3	Post ↔ Expr
4	Landscapes ↔ portraits
5	Surr ↔ Post
6	Figures ↔ portraits

MDS analyzes: Depending on the MDS algorithm used, the recommendations differ as to find out how many dimensions can be used to approximate the similarity data well enough. For classical MDS, the point where the Eigenvalues of the MDS solution taper off after a sharp, initial drop is commonly determined. Alternatively, one can refer to how many Eigenvalues are needed to explain X% of the total variance. This 'elbow' can be set at roughly 25 dimensions which then corresponds to explaining $\approx 80\%$ of the variance in the similarity data. The MDS solution can thus be used to localize each of the paintings in a 25-dimensional space. In order to determine what some of these dimensions might correspond to, we projected the paintings onto each dimension and plotted the images occupying the extreme ends of the scale. The results of this analysis are listed in Table 2 for the first 6 dimensions (corresponding to $\geqslant 60\%$ of the variance) which had a clear perceptual interpretation.

The most important dimension separated the historically old art periods from the newer ones. The second dimension separated flat images (such as the abstract, single-color paintings from Newman, but also the perspective-free Gothic images) from images with more depth structure (such as landscapes, but also street scenes). The third dimension was used to distinguish between Postmodern and expressionist art, whereas the fourth separates landscape paintings from portraits. The fifth and sixth dimension were used to differentiate between Surrealism and Postmodernism, and between whole-figure paintings versus portraits. On each of the six dimensions, neighboring paintings very often shared the same art period/artist.

Whereas some dimensions are clearly genre-based, others separate whole art periods, which corresponds well to the consistent cluster structure seen in the similarity matrix in Fig. 4. Participants thus used a mixture of historical and content-based clues to group the paintings. The reason for the success of content-based information—even though we explicitly asked participants not to group based on genre—lies in the fact that many art periods do, indeed, have a clear content preference: For the period of Rokoko, for example, many painters were appointed to court and were expected to create often highly stylized and glorifying paintings of the absolutist monarchs—thus creating many portraits and full-figure paintings which are typical for this period.

As the output of MDS in our case consists of a 25-dimensional vector for each of the 275 paintings, we can also derive classifiers for distinguishing art periods from this data. Using half of the 275 vectors as training set (labeled with the 11 art periods) and the other half as the testing set, we trained a simple classifier that fits multi-variate densities to each of the 11 training subsets. On average, art periods are predicted correctly in 66.3% of all cases—considering the difficulty of the task and participant's lack of expertise a surprisingly good performance. The results broken down by art period are shown in Fig. 5. As chance performance would be at around 9%, all classification results are significantly

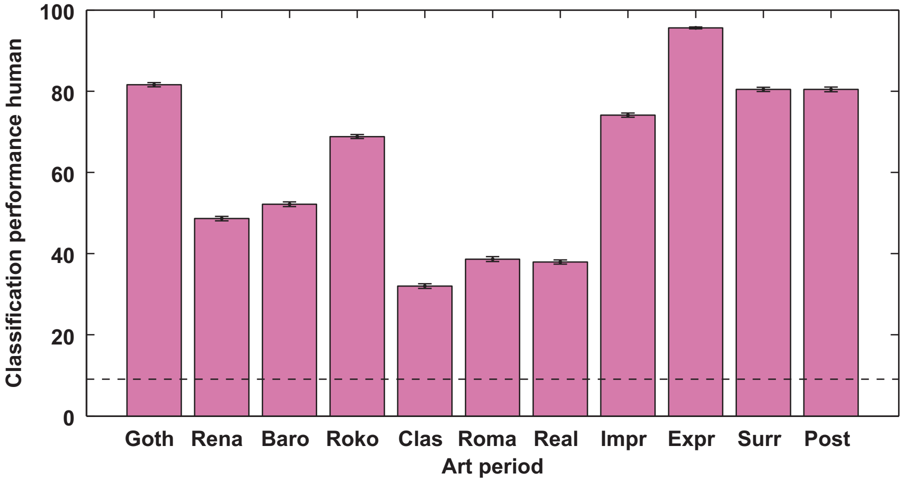

Fig. 5. Classification rate for the 11 art periods as predicted by the participant data. The plot also shows chance performance. Error bars depict SEM.

above chance. The remaining pattern resembles that of the prior analyzes: most notably, the Classicism, Romanticism, and Realism periods fare worst, whereas Gothic, Impressionism, Expressionism, Surrealism, and Postmodernism can be grouped well. The reason for the slightly inferior classification performance for Impressionism lies in the fact that participants usually split the artists in this period into distinct, far apart clusters—that is, the 'break' in Fig. 4a and b occurs *in the middle* of the Impressionist period.

Clustering analyzes: Whereas the dimensionality analysis from MDS has revealed possible image properties that participants use to classify the paintings, clustering analysis allows us to look at which paintings on average will be grouped together based on the similarity matrix. First, we grouped all images into two clusters to find the most important distinction within the data. Second, we created a grouping based on 11 clusters to check what a perceptual grouping of the same size as our art period grouping would yield.

For all three experiments, the two clusters split the 275 paintings into a larger group containing paintings up to the Impressionist period and a smaller group starting with the late Impressionists. This very stable result confirms both the overall pattern in the similarity matrix as well as our classification results obtained above and testifies to the ability of even non-experts to create theoretically meaningful structure in our sample of paintings. The results for the 11 clusters are summarized in Table 3. Again, results for all three experiments are very similar. Additionally, we find period-based clusters, content-based clusters, as well as clusters based on painting style (Cluster 7). The use of all three levels of analysis from low-level to high-level features can thus also be seen in the clustering data.

4.6. Questionnaires

After the three experiments, participants were asked to fill out a questionnaire in which they were debriefed about the difficulty of the experiment, their strategies for creating the clusters, as well as the degree of knowledge about the 11 art periods. All numeric ratings were done on a scale of 1 (not difficult/not knowledgeable)–5 (very difficult/very knowledgeable).

Overall, participants rated the experiment as moderately difficult (3.4 ± 0.16 SEM)—mostly due to the amount of images that had to be sorted. As was the case for the pre-experiment, the overall ratings for knowledge about the art periods was low at 2.2 ± 0.2 SEM. There were some differences across the art periods, however: an ANOVA across the participants resulted in a highly significant effect of art period ($F(10,390) = 6.8$; $p < 0.001$). Using a post-hoc Scheff criterion, for example, we found that both the Postmodern and the Rokoko period were significantly less well-known than the remaining art periods. Impressionism, Expressionism, Surrealism ranked highest in terms of participants' knowledge. Note that these results are only a weak predictor of either grouping consistency or classification performance. This could be either due to misestimating one's own knowledge about an art period or to insufficient knowledge about potential features that could differentiate the periods or define a style. These statements of course need to be taken with caution as overall familiarity was low after all.

Table 3
Results of the hierarchical cluster analysis for the human data.

Cluster	Experiments 1–3
1	Goth, Rena
2	Portraits (Goth,Rena,Baro,Roko,Clas)
3	Landscapes (Roma,Real,Clas)
4	Paintings with people (Rena,Baro,Clas,Roma)
5	Impr landscapes
6	Outliers
7	Fine brushstroke paintings
8	Impr, Expr
9	Expr (mostly Cubist)
10	Surr
11	Post

It is interesting that here we could not find any difference between the three experimental groups—participants' judgment about their overall knowledge about the art periods did not differ significantly. A closer look at the experimental ratings in Experiments 1 and 2 revealed that mainly two participants were responsible for the higher rating values in the second experimental group. Additionally, tests with the similarity matrices calculated without these two participants did not influence the clustering results or any of the consistency measures.

Finally, we analyzed the reports about potential strategies and features that participants used to create the clusters. For these, we first identified possible categories that were mentioned consistently across responses—the percentage of participants who used particular response categories are shown in Table 4. There were

Table 4
Clustering strategies mentioned in the questionnaires grouped by analysis levels.

Prior knowledge about art periods	40.5
Tried to find similar paintings by artist	7.5
Content	38.1
Overall style impression	35.7
Realistic versus abstract	33.3
Brush stroke	38.1
Color	21.4
Perspective	11.9
Lighting	9.5
Expressivity	4.8

The number specifies the percentage of participants who mentioned this specific strategy.

three groups of responses which fall into the three different levels of analysis mentioned in the introduction: the first group comprises prior, high-level knowledge about art periods and artists ('I identified Baroque paintings by sumptuous dresses'; 'I knew that some of the paintings were similar to Dali's'), the second group contains features based on image content ('I was looking for abstract versus representational images'; 'I looked at the way the faces were painted'), whereas the third group contains lower-level features of the painting style itself ('I grouped based on the thickness of the brush stroke as well as the degree of perspective in the paintings.'). As is evident from the response frequencies, all three groups were used during the clustering—this was true not only on average but also for some participants who reported to have employed multiple analysis levels.

In summary, these results confirm the previous MDS and clustering analyzes: participants have used a number of higher-level cues to form categories of similar paintings. Beyond the way the image is painted (from the choice of colors and the quality of the brush strokes to how much perspective it contains), clusters were created based on content-based criteria (from judging how specific objects are painted to the degree of abstraction), as well as on prior knowledge about art periods (from trying to recognize some well-known artists to the fact that surrealist paintings share certain aesthetic principles or recognizing cubist artworks). Comparing our results to that of Augustin and Leder [17], we can find a number of commonalities in the features that people mention explicitly. Not surprisingly, since we asked participants to group by *style*, emotional attributes which featured in their study cannot be found in this list.

5. Computational experiment

For the computational experiment, we were interested in finding out to which degree low-level, pictorial cues would be able to explain the groupings in the previous, perceptual experiments. To this end, we implemented several image processing algorithms that created a feature vector from each painting which are described below. We then simulated a series of 'computer participants' in order to compare the data of human grouping and computer grouping using the same analyzes.

5.1. Algorithms

Raw image data (*IMA*): One of the most straightforward low-level measure is to take the pixel values of a downsampled ($\approx 30\times 30$ pixels) version of the image. This is of course very far from being perceptually plausible as the brain does not assess similarity by subtracting two images, but it provides a good baseline for the further measures.

Color features (*RGB/CIE*): As color is one of the most prominent features in art, color histograms might provide a better correlation with the human data. In order to test this hypothesis, color histograms were extracted from both RGB-versions and CIELAB-versions of the paintings. Histograms always had 20 bins for each of the three color channels. CIELab is a perceptually plausible color space—the *L*-dimension corresponds to the luminance, the *a*-dimension to green–red transitions, and the *b*-dimension to blue–yellow transitions. The color opponency scheme is derived from both perceptual and physiological experiments on how the human visual system processes color (among other things, color opponency explains the after-images appearing after fixating a colored surface for a long time).

Fourier analysis (*FOU*): The human visual system is highly tuned towards the statistics of the environment as measured by fourier statistics (see [29]). The slope of the amplitude spectrum (as a function of frequency) of natural images, for example, was found to be close to 2 ($a=1/f^2$) corresponding to the scale invariance or self-similarity usually found in nature. Perhaps the amplitude spectra of paintings could be used to tell different artists, or art periods apart given that fourier analysis is sensitive to spatial frequencies (for example, thicker brush strokes in later art periods, very fine brush strokes in the Realism period). For our experiments, we determined the amplitude spectra by binning across 100 frequency bins in the two-dimensional fourier spectrum yielding a 100-dimensional vector.

Gist (*GIS*): Several studies in scene perception have shown that humans are able to understand the general context of novel scenes even when presentation time is very short (<100 ms) [30]. This overall meaning of a scene is often referred to as 'gist' and most commonly refers to low-level, global features such as color, spatial frequencies and spatial organization. In computational studies [31] it was found that a very simple frequency analysis (based on the output of filters tuned to different orientations and scales) of the spectrum of an image can already be enough to classify a scene as indoors, outdoors, open, closed—in short, the frequency spectrum allows to determine the *spatial content* of an image. In recent scene categorization experiments [31], this approach was also shown to have perceptual relevance as it was able to account for several experimental results on scene categorization. For the following experiments, all parameters of the implementation based on Oliva and Torralba [31] were set to the default values which yield a total of 320 values for a grayscale-image.

Information theory measures (*ENT/AES/REG*): Our final set of measures is based on the discussion of aesthetics in terms of information theory conducted in Rigau et al. [26,25]. These measures analyze the behavior of a set of informational measures based on the entropy of the palette, the compressibility of the image, and an information channel capturing the basic structure of the painting.

Entropy (*ENT*): The palette is considered as the range of colors selected by the artist with an associated probability distribution which is obtained from the normalization of the color histogram. Given an image of N pixels, the palette C can be represented by two random variables X_{rgb} and X_ℓ (and their corresponding probability distributions) which are defined using two different color representations. These color representations are sRGB color space and Y_{709} luminance, with an alphabet of 256^3 and 256 intensity values, respectively. The palette entropy $H(C)$ expresses the average uncertainty (or information content) of a pixel and $N\cdot H(C)$ measures the information content of an image (in bits). For our experiments, we used both $H(X_{rgb})$ and $H(X_l)$ as $H(C)$ yielding a two-dimensional vector.

Aesthetic measures (*AES*): The relative redundancy M_B expresses the reduction of pixel uncertainty due to the choice of a palette with a given color probability distribution instead of a uniform distribution and, in a certain way, can be seen as a measure of order [25]. The Kolmogorov complexity $K(x)$ of an image x is the length of the shortest program to compute x on an appropriate universal computer [32]. Due to the non-computability of K, JPEG compression is used to estimate $K(x)$. From a Kolmogorov complexity perspective, the *order* in an image can be measured by the difference between the image size (obtained using a constant length code for each color) and its Kolmogorov complexity. The ratio M_K between the order and the initial image size expresses the degree of order of the image without any a priori knowledge on the palette. In addition, if the initial image size is substituted by the palette entropy times the number of pixels, the order can be given by $N \cdot H(C) - K$ and expresses the reduction of uncertainty because of the compression achieved by Kolmogorov complexity. The ratio M_Z between this order and $N \cdot H(C)$ quantifies the degree of order created from a given palette [25]. In the experiments, we used all three normalized measures M_B, M_K, M_Z as a three-dimensional vector.

Regional measures (*REG*): The creative process described by Bense [33] can be understood as the realization of an information channel between the palette and the set of regions of the image [25,26]. From this channel, we use a BSP-partitioning algorithm which produces quasi-homogeneous regions extracting all the information of a painting and revealing its structure. This algorithm is based on the maximization of mutual information and yields a number of regions R for a given ratio of information extracted. This algorithm reveals the image's composition (macro-aesthetic description) clearly after relatively few partitions. Conversely, the details or forms in the painting appear when we reach a refined mesh (micro-aesthetic description). In our experiments, we used a 13-dimensional vector, which contains the number of regions created at values of 0.05–0.65 (in 0.05 steps) of mutual information.

5.2. Analysis and results

For the subsequent analyzes, all feature vectors were first stored for each image. We then simulated a 'computer participant' who would create the same number of clusters as in the human experiments. This was done by subjecting all 275 feature vectors to a simple k-means algorithm (with k set to the values of the participants). The output of this algorithm was then further processed into a similarity matrix similarly to Experiments 1–3. This approach enables us to use the exact same analysis tools in order facilitate comparison of computational and human data on categorization of artworks.

Consistency scores: Fig. 6 shows the results for c_{artist} scores for both human and computer data. First of all, among the computational measures, AES, GIS, and RGB fare best—albeit by a small margin. Post-hoc t-tests between the measures reveal the difference between FOU and GIS/AES to be significant ($t(28) = 2.2/2.3$; $p < 0.05$), whereas none of the other comparisons reaches significance. As is evident from Fig. 6, none of the measures are able to reach human performance which is almost twice as good. Given the low dimensionality of the measures derived from information theoretic considerations, however, their performance (especially the combined aesthetic measures of complexity) is surprisingly good in this context.

Similarity matrices: As one might expect already from the previous analysis, none of the similarity matrices created by the computational measures looks similar to the human data. In order to see which of the computational measures would correlate best with the human data, we conducted correlation analyzes between the similarity matrices. The comparisons with the human data yielded the following (significant, but low) r^2 values: IMA: 0.18, RGB: 0.15, CIE: 0.17, FOU: 0.15, GIS: 0.2, ENT: 0.1, AES: 0.1, REG: 0.1. Given that the correlation between Experiments 1 and 2 on the human data is $r^2 = 0.64$, we can already conclude that none of the measures comes close in fully explaining the human grouping results.

MDS analysis: Following MDS, we first tried to identify dimensions along which the different computational measures would analyze the data. Using the variance criterion, the following number of dimensions were be necessary to explain 80% of the variance in the similarity data: IMA(11), RGB(9), CIE(12), FOU(10), GIS(14), ENT(10), AES(10), REG(10). The overall number of dimensions needed is therefore in all cases much lower than that of the human data.

Overall, our analysis revealed that the first 6 dimensions of most computational measures (as far as they were interpretable) extract low-level properties of the artworks such as overall brightness, or large color differences. Fourier analysis—as could be expected—also is sensitive to texture. The information theoretic measures are sensitive to the complexity of images—both in terms of color palette (ENT) as well as in terms of 'painting style' (REG, AES). GIS, however, differs from the other measures as it specifically focuses on scale properties of the images. More specifically, dimension 4 corresponds to a higher-level dimension as it separates images with flat appearance from those with more depth structure—nevertheless, the separation is not as clear as for dimension 2 of the human data. Similarly, even though dimension 5 of the GIS measure distinguishes fine-grained from coarse-grained paintings, it does not do this consistently along art periods. Nevertheless, this result seems to be the reason for the highest correlation with the human data as seen previously. Apart from this exception, however, none of the other computational measures seems to correlate with the human data.

We then used the same classification scheme outline above to classify the dimensionality-reduced new feature vectors into art periods. The overall classification performance across all

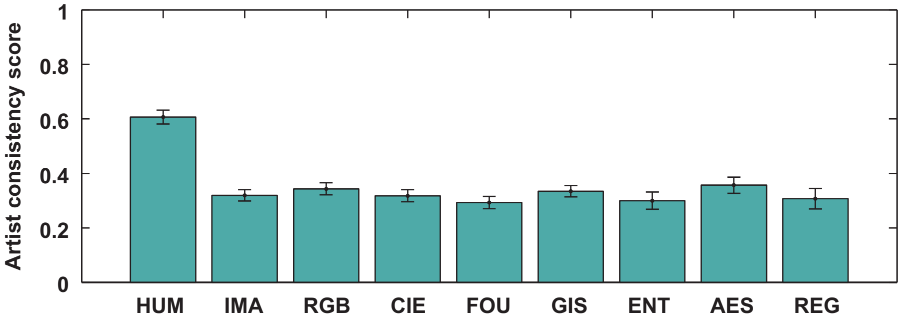

Fig. 6. Comparison of mean c_{artist} for human participants and simulated computer participants.

Table 5
Classification performance for computational features broken down by art period and overall performance.

Period	IMA	RGB	CIE	FOU	GIS	ENT	AES	REG
Goth	x				31		x	24
Rena	x	x	41			41		
Baro		30		30	x		33	x
Roko	21	x		28	28		33	22
Clas	42	26		x		x	27	x
Roma	x				x	41	44	25
Real			x			29		
Impr	32	30	21	x	x			
Expr	x	25	x	33	24	x		x
Surr	x	x	21	x				
Post	x	23	45	x	39	28		21
Perf. (%)	18.7 ± 1.1	16.5 ± 1.0	17.8 ± 1.2	14.8 ± 0.8	18.2 ± 1.1	18.5 ± 0.9	19.0 ± 0.9	15.8 ± 1.1

art periods is shown in the bottom row of Table 5. Whereas all classifiers performed significantly above chance, overall performance is rather low.[4] Additionally, the difference among classifiers is small—as would be expected from the artist consistency scores—with the best features being AES, IMA, ENT, and GIS. Again, it is surprising how comparatively well the low-dimensional information theory measures perform here.

Additionally, Table 5 lists results for individual art periods based on statistical tests whether classification of a specific periods was above chance (an 'x' lists performance above chance, values are only given in case performance is better than 20%). One can see that, for example, the IMA classifier can predict class membership for nine art periods above chance, whereas the AES classifier can only predict five periods including three strong classification results for Romanticism, Baroque, and Rokoko. In general, it seems that all classifiers have problems with the Gothic and the Surrealist periods. It is interesting to note that the first five classifiers all have above-chance performance for the newer art periods of Impressionism, Expressionism, Surrealism, and Postmodernism, whereas the information theory measures fare much better for the early art periods. This is an indication that different types of appearance-based, low-level information are suitable for classifying art into periods. Finally, at least with our choice of paintings, some periods seem 'more predictable'—albeit at rather low levels of performance.

6. Conclusion and outlook

Our study has shown that non-experts were able to reliably group unfamiliar paintings of many artists into meaningful categories. Dimension, clustering, and questionnaire analyzes have shown that this was done based on both low-level, more perceptual (brush stroke, perspective) and mid-level, more cognitive information (genre, motif) resulting in historically correct categories. Overall, however, higher-level information was used to a much greater extent in this task. The most salient finding includes a clear category break between pre-Impressionist and post-Impressionist art which corresponds well to the beginnings of modern art in art historical terms. This result was validated on a second dataset containing different paintings by the same artist showing that grouping strategies did not depend on our particular choice of paintings. In addition, few differences were observed when restricting the number of categories versus free grouping showing that the underlying processes in both types of tasks were highly similar. Additionally, computational classifiers created from the participant data are able to categorize art periods with a performance of about 66%—a truly impressive result for non-experts. As this is the first study to address categorization of art on such a broad scale spanning multiple art periods, our results therefore show that the judgments of non-expert observers can lead to surprisingly 'canonical' categorizations—a result which clearly points to a *common aesthetic* among our participants.

In order to address the issue of expertise more explicitly, we are planning comparisons with different populations ranging from 'true' non-experts (children) to full experts (students of art) on our stimulus set. This will give insights into the development of cognitive skills on the various levels of judgment. Additionally, as we asked participants to group based on the style of the image, it would be interesting also to group by different criteria such as early/late art, genre (still life, portrait) in order to get grouping results for these properties as well. Finally, in order to uncover the *truly perceptual* components of this task, we plan to use rapid categorization tasks as in Thorpe et al. [30]. These will allow us to look at perceptual processes in isolation as the speeded task prevents cognitive influences.

Several computational measures sensitive to color, texture, and spatial composition were implemented in order to seek low-level correlates with the human data. Both in terms of dimension and clustering analyzes results, none of the computational measures—with the notable exception of the GIS feature—correlated with the human data. In our opinion, this emphasizes the higher-level processing that even non-expert viewers make when viewing and interpreting works of art. Additionally, no single feature resulted in a good classification performance, although we found that the information theoretic measures were surprisingly efficient given their low dimensionality.

One of the interesting applications of the information theory measures, however, is to *characterize* the different art periods in terms of, for example, their spatial complexity [26]. The average behavior of the REG measure is shown in Fig. 7 and shows that, for example, the spatial complexity of images has more or less increased over time—with the notable exception of the Surrealism period with its propensity for larger, evenly colored areas. It remains to be seen how these results will generalize to larger datasets.

Our failure to find good correlates for human data (resulting also in good overall classification performance) could have several reasons:

- our measures left out some important low-level correlate of human judgments,

[4] A more 'direct' classification scheme based on the *original* feature vectors, performs only marginally better—on average, performance increases by 3.5%.

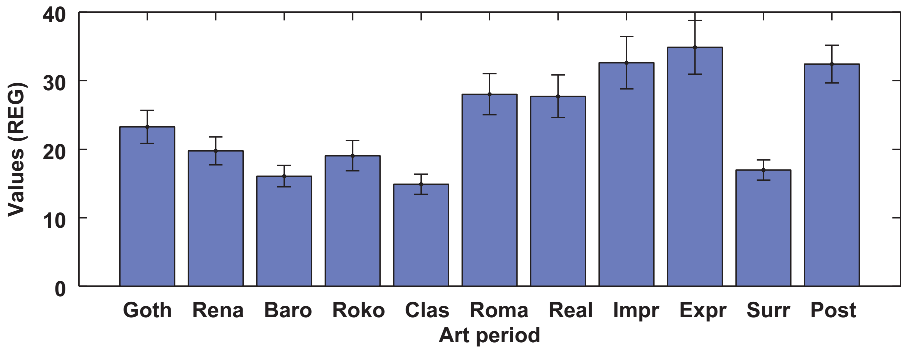

Fig. 7. Average number of regions for the REG measure.

- no single measure can correlate with human data, multiple measures need to be taken into account,
- humans also use higher-level properties of paintings or semantic knowledge to do the task.

Although of course a host of algorithms measuring low-level image properties exist, the fact that neither texture, color-based, scale-sensitive or complexity measures correlate at any dimension casts doubt on whether another measure will do much better. The second point—while perhaps needed to boost computational classification performance to better levels—also seems unlikely given that the dimensions for the human data clearly correspond to mid- and higher-level image properties. Additionally, even higher dimensions in the computational data never capture such properties explicitly. In our opinion, participants therefore clearly used mid- and higher-level properties of the paintings which are currently beyond what our computational measures—or any combination thereof—can extract. Nevertheless, we want to stress that our computational studies were only the beginning of a larger set of experiments, for which we also plan to implement more complex types of measures. Given the success of the Gist method and the information theoretic measures, we will test more sophisticated features based on automatic analysis of the depth structure [34] of paintings which might yield further insights.

Taking a step back, what does our study mean in the context of the burgeoning field of computational aesthetics? On the one hand, we have shown that computational analyzes (such as clustering, MDS) provide powerful tools to uncover specific perceptual and cognitive dimensions of aesthetics which help to make aesthetics quantifiable. On the other hand, algorithms and approaches from computer vision—while being useful for characterizing low-level aspects of images very efficiently—still have to improve considerably before being able to fully explain the complex processes underlying aesthetic judgments.

References

[1] Little S. Isms: understanding art. Universe Publishing; 2004.
[2] Fechner G. Vorschule der aesthetik (Preschool of aesthetics). Olms; 1876.
[3] Wallraven C, Kaulard K, Kürner C, Pepperell R, Bülthoff HH. Psychophysics for perception of (in)determinate art. In: APGV 2007. New York, USA: ACM Press; 2007. p. 115–22.
[4] Wallraven C, Kaulard K, Kürner C, Pepperell R, Bülthoff HH. In the eye of the beholder: perception of indeterminate art, in: Computational aesthetics 2007; 2007. p. 121–8.
[5] Arnheim R. Art and visual perception: a psychology of the creative eye: the new version. Berkeley, CA: University of California Press; 1974.
[6] Pinna B. Art as a scientific object: toward a visual science of art. Spatial Vision 2007;20(16):493–508.
[7] Zeki S. Inner vision: an exploration of art and the brain. Oxford: Oxford University Press; 1999.
[8] Spatial vision—art and perception: towards a visual science of art, Part 1; 2006.
[9] Spatial vision—art and perception: towards a visual science of art, Part 2; 2007.
[10] Spillmann L. Artists and vision scientists can learn a lot from each other—but do they? Gestalt Theory—An International Multidisciplinary Journal 2007; 29.
[11] Livingstone M. Vision and art. The biology of seeing. NY: Harry N. Abrams; 2002.
[12] Leder H, Belke B, Oeberst A, Augustin D. A model of aesthetic appreciation and aesthetic judgments. British Journal of Psychology 2004;95(Pt 4):489–508.
[13] Harnad S. Cognition is categorization, paper presented at UQ̂M Summer Institute in Cognitive Sciences on Categorisation 2003; 2005.
[14] Sloutsky V. The role of similarity in the development of categorization. Trends in Cognitive Sciences 2003;7:246–51.
[15] Cooke T, Jäkel F, Wallraven C, Bülthoff H. Multimodal similarity and categorization of novel, three-dimensional objects. Neuropsychologia 2007; 45:484–95.
[16] Hasenfus N, Martindale C, Birnbaum D. Psychological reality of cross-media artistic styles. Journal of Experimental Psychology—Human Perception and Performance 1983;9:841–63.
[17] Augustin D, Leder H. Art expertise: a study of concepts and conceptual spaces. Psychology Science 2006;48:135–56.
[18] Redies C. A universal model of esthetic perception based on the sensory coding of natural stimuli. Spatial Vision 2007;21:97–117.
[19] Cutzu F, Hammoud R, Leykin A. Distinguishing paintings from photographs. Computer Vision Image Understanding 2005;100(3):249–73 ISSN 1077-3142.
[20] Marchenko Y, Tat-Seng C, Irina A. Analysis and retrieval of paintings using artistic color concepts. In: IEEE international conference on multimedia and expo. ICME 2005; 2005. p. 1246–9.
[21] Marchenko Y, Chua T-S, Jain R. Ontology-based annotation of paintings using transductive inference framework. In: 13th international multimedia modeling conference; 2007. p. 13–23.
[22] Leslie L, Chua T-S, Ramesh J. Annotation of paintings with high-level semantic concepts using transductive inference and ontology-based concept disambiguation. In: MULTIMEDIA '07; 2007. p. 443–52.
[23] Leyton M. The structure of paintings. Berlin: Springer; 2007.
[24] Datta R, Joshi D, Li J, Wang JZ. Studying aesthetics in photographic images using a computational approach. In: Proceedings of ECCV; 2006. p. 288–301.
[25] Rigau J, Feixas M, Sbert M. Informational dialogue with Van Gogh's paintings. In: Computational aesthetics 2008; 2008. p. 115–22.
[26] Rigau J, Feixas M, Sbert M. Informational aesthetics measures. Computer Graphics and Applications 2008;28(2):24–34.
[27] Borg I, Groenen P. Modern multidimensional scaling. 2nd ed. Berlin: Springer; 2005.
[28] Shepard R, Cermak G. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. Cognitive Psychology 1973;4:351–77.
[29] Simoncelli E, Olshausen B. Natural image statistics and neural representation. Annual Review of Neuroscience 2001;24:1193–216.
[30] Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–2.
[31] Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition. Progress in Brain Research: Visual Perception 2006;155:23–36.
[32] Li M, Vitányi PMB. An introduction to Kolmogorov complexity and its applications. In: Graduate texts in computer science. Berlin: Springer; 1997. ISBN 0-387-94868-6.
[33] Bense M. Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie (Introduction to the information-theoretical aesthetics. Foundation and application in the text theory). Rowohlt Taschenbuch Verlag GmbH; 1969.
[34] Hoiem D, Stein A, Efros A, Hebert M. Recovering occlusion boundaries from a single image. In: Proceedings ICCV; 2007.