



Changes of Statistical Properties During the Creation of Graphic Artworks

Christoph Redies*, Anselm Brachmann** and Gregor Uwe Hayn-Leichsenring

Experimental Aesthetics Group, Institute of Anatomy I, University of Jena School of Medicine, Jena University Hospital, Jena, Germany

Received 15 July 2013; accepted 13 October 2013

Abstract

During the creation of graphic artworks, we studied the evolution of higher-order statistical image properties (complexity, self-similarity, anisotropy of oriented luminance gradients, the slope of log–log plots of radially averaged Fourier power, and the fractal dimension). First, we analyzed two series of lithographs by Pablo Picasso, which represent transformations of highly aesthetic artworks. Second, one of the authors generated a dataset of 20 grayscale drawings using the computer as a drawing tool. The dataset comprised also the unfinished state images that were saved throughout the production process. The final states of the drawings were compared to versions of the same drawings, in which the constituent pictorial elements were shuffled, thereby diminishing the overall compositional intent of the artist. Results show that self-similarity was a property closely associated with artistic merit in the different types of images analyzed. In a psychological experiment, 20 non-expert participants evaluated the original abstract drawings as more harmonious and ordered but less interesting than the shuffled versions. Our study demonstrates that statistical image properties can be studied during the creation of artworks, if artistic and analytical processes are closely coordinated in a computer-based approach, which offers the possibility to produce appropriate control stimuli.

Keywords

Experimental aesthetics, complexity, self-similarity, Fourier analysis, fractal dimension

Why do you think I date everything? Because it is not sufficient to know an artist's works — it is necessary to know when he did them, why, how, under what circumstances. . . Some day there will be undoubtedly a science — it may

* To whom correspondence should be addressed. E-mail: christoph.redies@med.uni-jena.de

** Present address: Institute of Informatics, Freie Universität Berlin, Takustraße 9, D-14195 Berlin, Germany. E-mail: brachman@mi.fu-berlin.de

be called the science of man — which will seek to learn more about man in general through the study of creative man. I often think about such a science, and I want to leave to posterity a documentation that will be as complete as possible. That's why I put a date on everything I do.

Pablo Picasso, 1943 (quoted from Green, 2005)

1. Introduction

The aim of experimental aesthetics is to understand which stimulus properties elicit brain responses that correlate with the perception of beauty. In the computational approach to this question, researchers are investigating the physical image properties that characterize visually pleasing stimuli with computer vision algorithms. For example, evidence from several groups suggests that a large subset of Western and East Asian visual artworks possesses a nearly scale-invariant (fractal-like) Fourier spectrum (Alvarez-Ramirez *et al.*, 2008; Graham and Field, 2007, 2008; Graham and Redies, 2010; Redies *et al.*, 2007a, b). Artworks share this property with images of complex natural scenes (Burton and Moorhead, 1987; Field, 1987; Geisler, 2008; Ruderman and Bialek, 1994). This similarity had led to the hypothesis that some artists tend to create artworks with properties that resemble those of natural scenes (Graham and Redies, 2010; Redies, 2007). Interestingly, the human visual system is adapted to process complex natural scenes with an efficient code (Olshausen and Field, 1996; Parraga *et al.*, 2000; Simoncelli and Olshausen, 2001; Wiltschut and Hamker, 2009).

Other statistical properties that have been studied in the field of experimental aesthetics are complexity, self-similarity and anisotropy. Berlyne (1974) proposed that the hedonic value of visual stimuli follows a u-shaped curve as the complexity of a pattern increases (see also Forsythe *et al.*, 2011). Supporting this notion, Taylor and colleagues showed that humans prefer intermediate values for the fractal dimension, a measure closely related to complexity, in natural and artificial images (for a review, see Taylor *et al.*, 2011). Our own results confirm that large subsets of visual artworks have complexity values in an intermediate range (Redies *et al.*, 2012). We also found that subsets of artworks and images of natural scenes are self-similar, i.e., details of the images have a distribution of oriented luminance gradients similar to the distribution in the entire image (Amirshahi *et al.*, 2012; Braun *et al.*, 2013; Redies *et al.*, 2012). Finally, colored artworks are, in general, highly isotropic, i.e., each image contains luminance gradients of similar strength across all orientations (Braun *et al.*, 2013; Koch *et al.*, 2010; Melmer *et al.*, 2013; Redies *et al.*, 2012).

Most studies in experimental aesthetics have used images that represent the final product of artistic creativity, i.e., completed artworks. To the best of our knowledge, there are no investigations of the question how higher-order image properties evolve during the creation of artworks. In the present study, we therefore asked how the statistical measures changed in datasets of images that comprise not only the final artworks but also state images that were saved throughout the process of their creation. State proofs are available for the printed works of many artists, but they usually cover only the final stages of the creation process, when the artist wants to assess the visual impact of the almost completed artwork. Additionally, there are movies that show how artists paint (e.g., of Pablo Picasso, Jackson Pollock or Gerhard Richter), but these movies tend to focus on the artist's movements and actions; the developmental stages of the artworks are usually not captured in a quality high enough for a comparative analysis of statistical image properties.

A notable exception are two series of lithographs by Pablo Picasso that represent variations on pictorial themes (Gauss, 2000; Mourlot, 1970; Stolzenburg, 1997). In the first series, Picasso started with a realistic representation of two female nudes and modified this composition in 17 steps until the drawing had transformed into a cubist representation of the same subjects (*Les Deux Femmes Nues*; Fig. 1). In the second series, which comprised 11 state prints, Picasso initially created a realistic representation of a bull, which he then reduced to a cubist drawing and finally to a simple line drawing (*Le Taureau*; Fig. 2). The two series of lithographs represented artworks that might be considered unfinished only at the very beginning of each series. All the other state prints represent transformations along fully artistic compositions (Stolzenburg, 1997). We would expect that image properties that reflect the high aesthetic value of these lithographs, remain relatively constant throughout each of the series, whereas image properties that are associated with particular artistic styles (realism *versus* cubism *versus* reduced line drawing) may vary to a greater extent. The analysis of the Picasso state series may thus help us to distinguish between image properties that are associated with artistic style, and properties that are more closely linked to artistic merit independent of style.

Because the two Picasso series represent transformations of highly aesthetic artworks, they tell us less about the more common process of artistic creation, which typically proceeds from a preliminary draft, to intermediary (unfinished) states of the artwork, to a complete and fully artistic composition. In view of the lack of appropriate study material, one of the authors (C.R.) created a dataset of 20 abstract grayscale paintings (here termed C.R. drawings), starting with a few pictorial elements that were painted with a brush on rice paper. He then added more pictorial elements to the initial drawings by using the computer as a drawing tool. Along the way of production, state images were saved and stored in digital format. By drawing digitally onto the com-

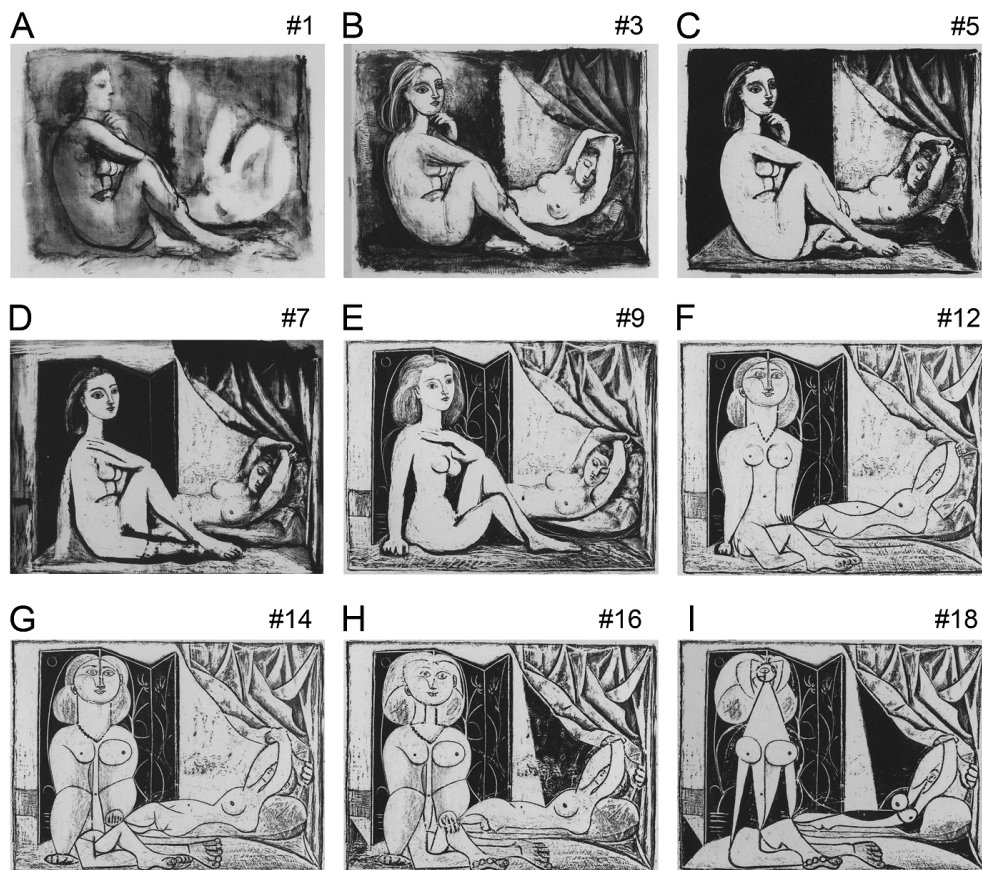


Figure 1. State series of lithographs entitled *Les Deux Femmes Nues* by Pablo Picasso, 1945/46 (No. 16 in the Mourlot [1970] oeuvre catalogue). The number of each state according to Mourlot (1970) is indicated above each panel (I, final state). © Succession Picasso/VG Bild-Kunst, Bonn 2013.

puter screen, the state images and the final drawings were produced under the same conditions and were of a similarly high digital quality. The aim of this part of the study was to ask how the higher-order statistical properties changed as the 20 drawings were generated. We hypothesized that any property, which reflects artistic merit, would approach values typical of similar types of artworks, whereas other properties that relate less to artistic merit, may change in any direction or remain constant during the creation process. Specifically, we argue that properties of the C.R. drawings, which potentially relate to artistic merit, should approach levels that are measured in a dataset of graphic art of Western provenance.

Finally, we assumed that properties, which potentially characterize artworks, should be more pronounced in images of artistic merit than in comparable images with less artistic merit. The latter type of image was produced from the C.R. drawings by randomly shuffling the pictorial elements in each image, thereby producing images, in which the artistic intent of a global composition was destroyed, at least in part. A complete destruction of artistic merit might not be possible with this approach because the individual pictorial elements

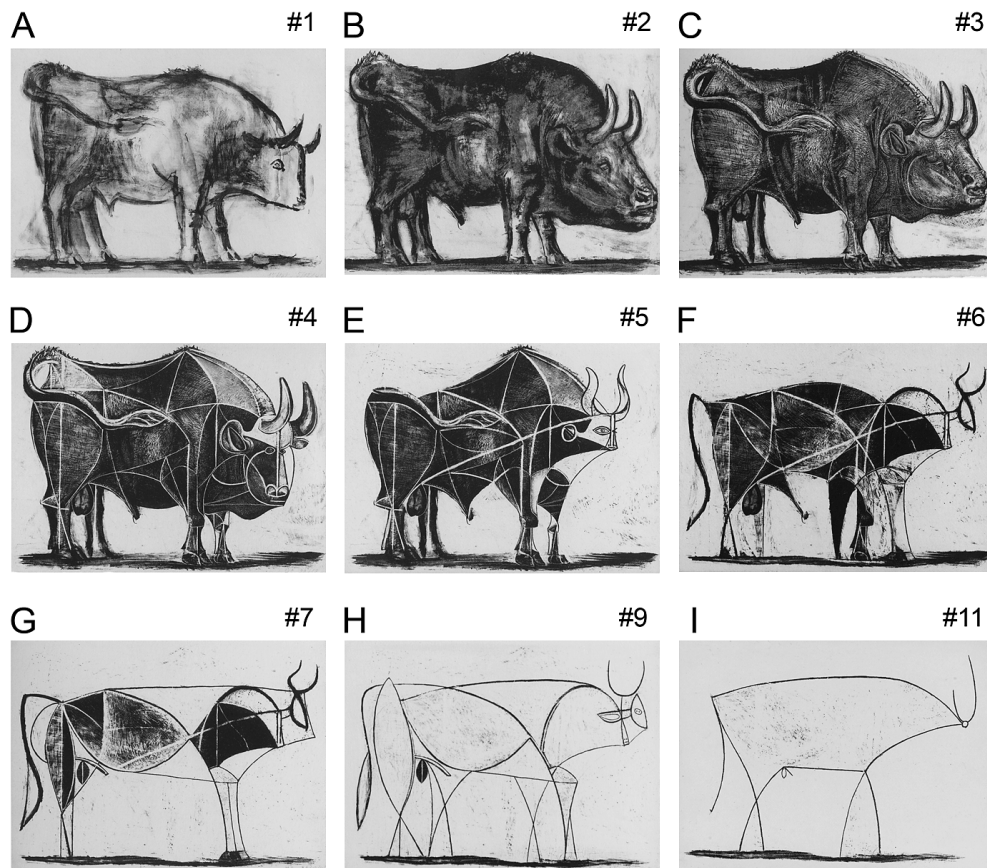


Figure 2. State series of lithographs entitled *Le Taureau* by Pablo Picasso, 1945/46 (No. 17 in the Mourlot [1970] oeuvre catalogue). The number of each state according to Mourlot (1970) is indicated above each panel (I, final state). © Succession Picasso/VG Bild-Kunst, Bonn 2013.

may also carry aesthetic value by themselves, irrespective of their global arrangement in the intended artistic composition. To study more closely in what subjective aspects of artistic merit the original and the shuffled versions of the C.R. drawings differed, the two types of images were rated by participants in a psychological experiment according to different indicators of subjective image quality. Participants were asked how aesthetic, harmonious, orderly, interesting and complex the original and the shuffled versions were (for a justification of choosing these terms, see Section 2.3). The subjective ratings were then correlated with each other and with the statistical image features.

In summary, we designed the present study to identify candidate properties that might characterize subsets of images with artistic merit, by investigating the process of artistic creation. Candidate properties can be expected (i) to remain relatively stable in the Picasso series of lithographs, (ii) to reach levels comparable to those of other graphic artworks during the creation of the C.R. drawings, and (iii) to be higher for the original versions than for shuffled versions of the C.R. drawings.

2. Material and Methods

2.1. Image Data

2.1.1. Picasso's Series of Lithographs

Two series of lithographs from Pablo Picasso were scanned from a book on his graphic works (Gauss, 2000). The series were *Les Deux Femmes Nues* [No. 16 in the Mourlot (1970) oeuvre catalogue; Fig. 1] and *Le Taureau* (No. 17; Fig. 2). [A1]¹

2.1.2. Twenty Abstract Drawings

One of the authors (C.R.) created twenty monochrome drawings, which consisted of 52–127 isolated pictorial elements (mean: 92 ± 18 SD elements) on a background of white paper. The generation of each drawing was started by painting relatively large elements with a soft brush in black ink on white Japanese rice paper mounted on cardboard. Examples are shown in Fig. 3 (left column) for four of the drawings. [A2]

Additional pictorial elements were added by C.R. at separate levels with the Corel Painter Essentials program (Version 3, Corel, Ottawa, Canada) using a variety of settings for simulated artistic material (brush, pencil, crayon, etc.), stroke width, gray levels and so on. The final result was a drawing, in which each pictorial element of the composition could be moved independently of the other elements (see Section 2.1.4).

The drawings were produced in parallel during multiple short periods between December 2007 and December 2008. In composing the drawings, C.R. did not consciously follow any explicit rules and the resulting abstract patterns were not intended to contain any figurative or contextual meaning. The only aim was to create visually pleasing drawings that follow some subjective (non-explicit) rules of harmonious visual composition. The author reported that the subjective degree of freedom for placing elements was higher early during the production process and decreased more and more towards the end of their production. The author finished the production when, on purely subjective grounds, he felt that a final and satisfying composition was reached. During the production, 6–17 state images (mean: 9.9 ± 2.5 SD) were saved for each drawing. No objective measurements were carried out on any the drawings until all drawings were completed. Although it was planned to include the drawings in future experimental studies in general, the experimental design of the present study had not been conceived at the time when the drawings were produced.

¹ The coding in square brackets refers to the description of additional technical details in the Supplementary Appendix published online.

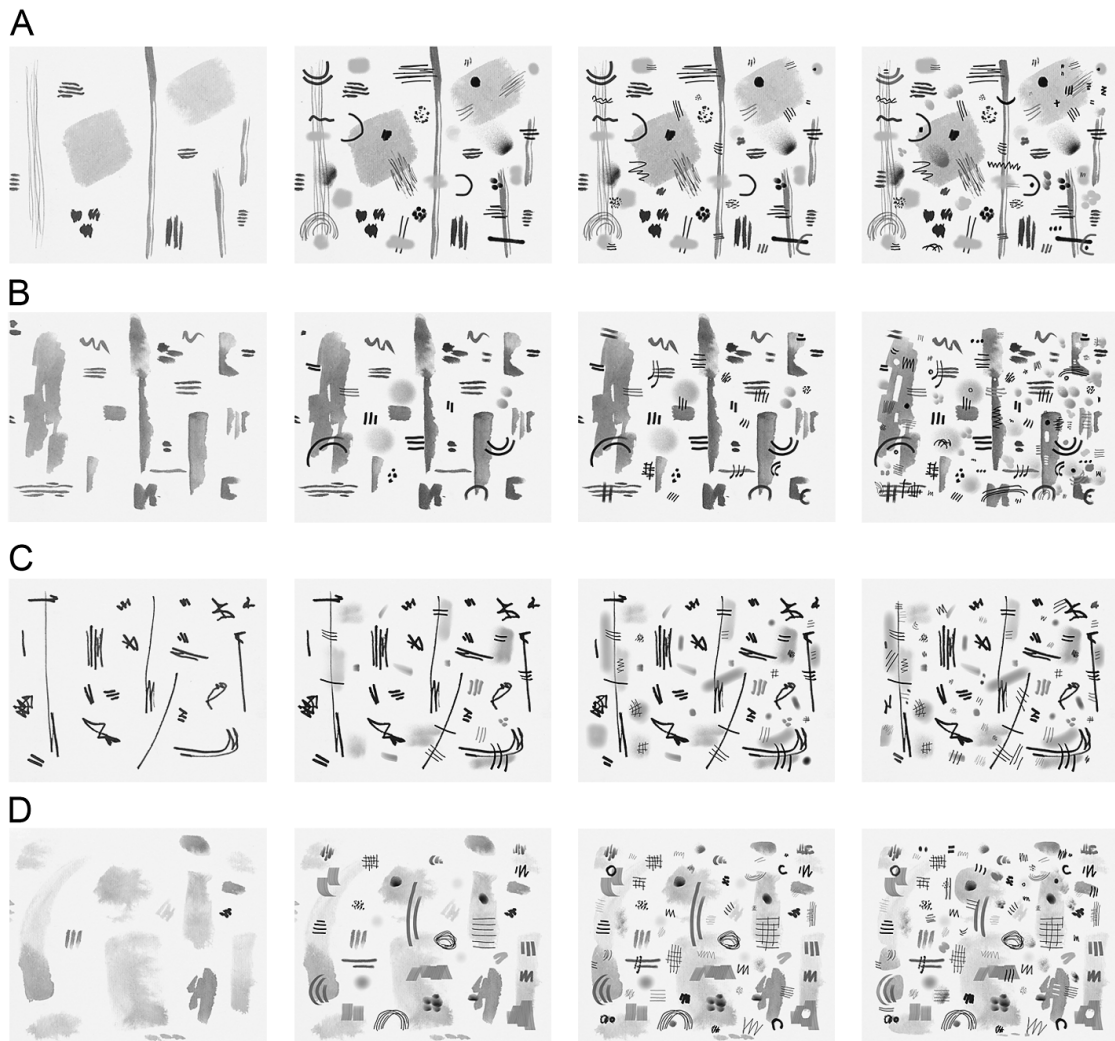


Figure 3. Examples of the state images of four of the C.R. drawings. The initial ink drawings are shown in the left hand column, representative state images in the two middle columns and the final version of each drawing in the right hand column.

2.1.3. Dataset of Monochrome Graphic Art

For comparison, we analyzed an image dataset of 200 graphic artworks that was studied before (Redies *et al.*, 2007b). The dataset comprised artworks of Western provenance from the 15th to the 20th century and covered a large variety of artistic styles, techniques and subject matters. The images were obtained by scanning reproductions from high-quality art books. [A3]

2.1.4. Shuffled Versions of the C.R. Drawings

To assess whether any of the statistical measures were specific for the final composition of the pictorial elements in the C.R. drawings, or were related to the pictorial elements themselves, irrespective of their positions relative to each other, we generated shuffled versions of the drawings. In the shuffled versions, the pictorial elements were placed at randomized positions by a

computer. However, the randomization process was restricted in order to avoid extensive occlusion of elements by superposition, especially of the large elements on top of smaller elements added later on. For each of the 20 original C.R. drawings, 10 shuffled versions were generated (200 images in total). [A4]

2.2. Image Analysis

2.2.1. PHOG Analysis

Three statistical measures (self-similarity, complexity and anisotropy) were calculated with a metric that was derived from histograms of oriented luminance gradients (HOG features; Dalal and Triggs, 2005), as described previously (Amirshahi *et al.*, 2012; Braun *et al.*, 2013; Redies *et al.*, 2012). In the image, the oriented luminance gradients correspond to edges or lines with different orientations (for an example, see Supplementary Fig. S1A–C). A more detailed description on how these measures were calculated is provided in the Supplementary Appendix published online [A5]. For a full mathematical description, see the Appendix in Braun *et al.* (2013). To illustrate the statistical measures, Supplementary Table S1 lists the values that were obtained for the three exemplary photographs shown in Supplementary Fig. S1H–J.

To measure self-similarity, HOG features were obtained at consecutive levels of an image pyramid (PHOG; Bosch *et al.*, 2007). An example of an image pyramid and its HOG values is shown in Fig. S1. The histograms at different levels of the pyramid were then compared with the ground level histogram (Fig. S1G). The self-similarity values calculated in the present study have a range between zero and one. Values close to one are obtained if the HOG features at a given level of the pyramid are similar to those of the entire image at the ground level, i.e., if the details of an image consist of an array of oriented gradients that resembles the composition of gradients in the entire image. Lower values are obtained if the details of an image have gradient compositions that differ to a larger degree from the entire image. [A5]

Complexity was defined as the sum of the strengths of the gradients in the entire image across all orientations (Redies *et al.*, 2012). Anisotropy is the variance of the luminance gradient strengths across the 16 orientation bins at level 3 of the pyramid (Redies *et al.*, 2012). The anisotropy measure indicates how much the strength of the oriented gradients differs across orientations. High values are obtained if one or a few orientations (e.g., vertical and horizontal orientations) are more prominent in the image than other orientations (e.g., oblique orientations). A value close to zero implies that the gradients in an image are about equally strong for all orientations.

2.2.2. Fourier Analysis

Fourier analysis was performed in order to measure the slope in log–log plots of the radially averaged power spectrum, as done previously in studies on

natural scenes (Field, 1987; Ruderman and Bialek, 1994) and monochrome paintings (Graham and Field, 2007; Redies *et al.*, 2007a, b). The slope value can be seen as an indicator of the relative strength of low spatial frequencies (coarse detail) and high spatial frequencies (fine detail) in an image. Images of natural scenes and monochrome paintings share mean slopes of around -2 (for a review, see Graham and Redies, 2010). A value of -2 indicates that the power spectrum remains constant as one zooms in and out of the image, i.e., it is fractal-like (scale-invariant). Values higher and lower than -2 (i.e., a shallower and steeper slope, respectively) imply that high spatial frequencies and low spatial frequencies, respectively, are more prominent in the power spectra, compared to the spectrum with a slope of -2 . [A6]

2.2.3. Analysis of the Fractal Dimension

The fractal dimension of each image was estimated with the box-counting method (Mureika and Taylor, 2013). While simple forms like a dot, a line or a square have a fractal dimension equal to their Euclidean dimension, more sophisticated patterns like a curve may have a fractal dimension somewhere between 1 and 2. The fractal dimension can be seen as an indicator for the complexity of a pattern: A high fractal dimension indicates high complexity, while a low fractal dimension is measured in patterns of low complexity (Mureika and Taylor, 2013). [A7]

2.2.4. Distance Measure for State Series

For the C.R. drawings, we quantified how different each state image was to the final version of the drawing. To obtain a measure for this distance, the absolute difference between the pixel values of the final version and of the state image was calculated on a pixel-by-pixel basis. The average pixel value of the resulting difference image was then calculated and used as a measure of the distance of the state image to the final drawing. As the state images become more similar to the final drawing, the distance measure approaches zero.

2.2.5. Statistical Analysis

Because some of the sampled data did not follow a Gaussian distribution, we used non-parametric statistical tests. To assess whether any of the image properties changed during the creation of the graphic drawings, the Spearman coefficient (ρ) was calculated in a two-tailed test in order to correlate the measured values with the distance of the state image to the final state (Section 2.2.4). The analysis was carried out for all images together and also for each image separately. The results for the different categories of images (original *versus* shuffled versions of the C.R. drawings, and graphic artworks) were compared with the Kruskal–Wallis one-way analysis of variance test using Dunn’s multiple comparison post-test. Results were considered significant for p -values smaller than 0.05.

2.3. Rating Experiment

The twenty participants who took part in the experiment were students or graduates of medical or life science studies and non-experts in arts (age: 18 to 30 years old, mean: 23.3 years, nine males). They all reported normal or corrected-to-normal vision and gave their written consent prior to the experiment. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki and was approved by the ethics committee of Jena University Hospital. The stimuli were the 20 original drawings and one sample each of the shuffled versions of the drawings. [A8]

The experiment consisted of five different phases. In each phase, participants were asked to evaluate the 20 original and the 20 shuffled images in different subjective categories that relate to perceived image quality (*aesthetic, harmonious, ordered, interesting and complex*). *Aesthetic* was chosen as a category because it is widely used; however, it is relatively unspecific and may carry different connotations. We excluded the even more widely used term *beautiful* (Augustin *et al.*, 2012; Jacobsen and Höfel, 2002) due to its lack of specificity. The categories *harmonious* and *ordered* were chosen because they refer more clearly to the compositional quality of images and thereby correlated with the artistic intentions to create visually pleasing drawings (see above). *Interesting* represents a term that carries positive valences and refers to the emotional arousal (hedonic value, see Berlyne, 1974), which an image can elicit in the observer; *interesting* and *pleasing* represent different dimensions in artworks (Cupchik and Gebotys, 1990). *Complex* is a feature that has been previously reported to contribute to aesthetic perception by several investigators (Augustin *et al.*, 2012; Berlyne, 1974; Forsythe *et al.*, 2011; Jacobsen and Höfel, 2002). Also, it seems of interest to compare the subjective evaluation for *complex* with the complexity measure used in the present study. [A9]

3. Results

3.1. State Series by Picasso

Results for the two state series by Picasso are shown in Fig. 4. For both series, complexity increases initially until the fourth state is reached (Fig. 4A). This increase corresponds to more and/or stronger luminance gradients in the prints. In agreement with the subjective visual impression gathered from the state prints in the *Le Taureau* series (Fig. 2), complexity subsequently decreases steadily until the final version is reached, which consists of a simple line drawing. In contrast, the state prints of the *Les Deux Femmes Nues* series become more complex with the emergence of the final cubist drawing that contains oriented gradients of high contrast.

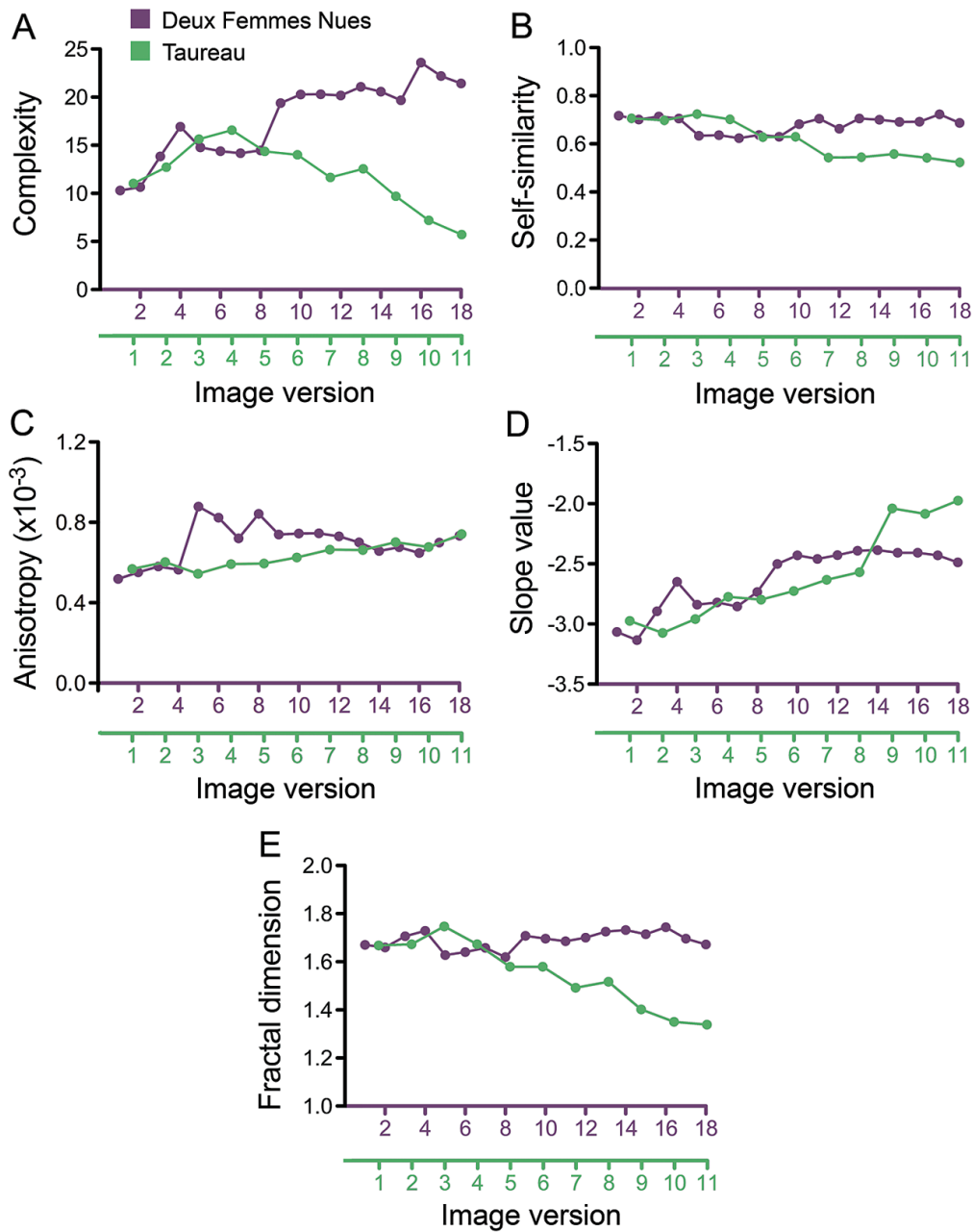


Figure 4. Results for the two state series of lithographs by Pablo Picasso (dark grey/purple, *Les Deux Femmes Nues*; light grey/green, *Le Taureau*; see Figs 1 and 2). The statistical measures are plotted as a function of the image version, i.e., the state number according to Mourlot (1970). Because the number of states differs between the two series, two separate x axes appear below each plot. This figure is published in color in the online version.

Self-similarity values remain relatively stable, especially for the *Les Deux Femmes Nues* series (mean: 0.68 ± 0.03 SD; Fig. 4B). Initially, this series and the *Le Taureau* series start off with a similar value of around 0.71. However, in the *Le Taureau* series, values gradually decline to a final value of 0.52 as the final and simple line drawing is reached. Anisotropy values increase from

0.58×10^{-3} to 0.75×10^{-3} in the *Le Taureau* series and vary between 0.52×10^{-3} and 0.88×10^{-3} in the *Les Deux Femmes Nues* series (Fig. 4C). The slope values of the 1d Fourier spectra increase from -3.0 to -2.5 and -2.0 for the *Les Deux Femmes Nues* series and the *Le Taureau* series, respectively (Fig. 4D). The fractal dimension remains relatively constant for the *Les Deux Femmes Nues* series (mean: 1.69 ± 0.04 SD) and decreases for the *Le Taureau* series from an initial value of 1.70 to a final value of 1.37 (Fig. 4E).

3.2. Abstract Drawings

3.2.1. Statistical Properties

Figure 3 shows the initial ink drawings (left columns), representative state images (two middle columns) and the final versions of four of the 20 original abstract drawings that were created by one of the authors (C.R. drawings). The final versions of the other 16 drawings are displayed in Supplementary Fig. S2. Figure 5 illustrates an example of the final drawings at a higher magnification (Fig. 5A), and an image composed of the same 85 pictorial elements that were approximately sorted by size (Fig. 5B; after Wehrli, 2004). Note that some of the elements consist of multiple dots or lines that are of similar size and form and are spaced close to each other. One of the images, in which the same elements are placed at shuffled positions, is displayed in Fig. 5C.

Supplementary Fig. 3 illustrates the results of the different measures for the series of state images of the C.R. drawings. Each colored line represents results for one drawing. Results are plotted as a function of the distance of each state image to the final version (for our definition of the distance, see Section 2.2.4). For each measure, the box plots in Fig. 6 compare the results for the 20 final (original) C.R. drawings, the 200 shuffled versions (10 images per drawing), and the dataset of 200 graphic artworks of Western provenance.

As expected, the complexity of all C.R. drawings increases during their creation (Supplementary Fig. S3A), as more pictorial elements are added and the drawings reach their final state (Fig. 3). In a joint analysis of the data points for all 20 drawings, there is a significant correlation between complexity and the distance to the final version ($p < 0.001$; Spearman ρ : -0.49). When analyzed on an individual basis, complexity tends to increase for all drawings as each one approaches its final state (Spearman ρ range for the 20 drawings: -0.72 to -0.98 ; $p < 0.05$ to $p < 0.001$). The mean complexities of the original C.R. drawings (9.84 ± 2.26 SD) and the shuffled versions (8.63 ± 1.90 SD) do not differ significantly between each other, but they are lower ($p < 0.001$) than the mean value for the dataset of graphic artworks (18.98 ± 7.61 SD; Fig. 6A).

Self-similarity tends to increase in the creation process in the entire dataset of C.R. drawings ($p < 0.001$; Spearman ρ : -0.38), but only for 13 out of the 20 drawings (Spearman ρ range for the 13 drawings: -0.61 to -1.00 ; $p < 0.05$ to $p < 0.001$; Supplementary Fig. S3B). The self-similarity of the

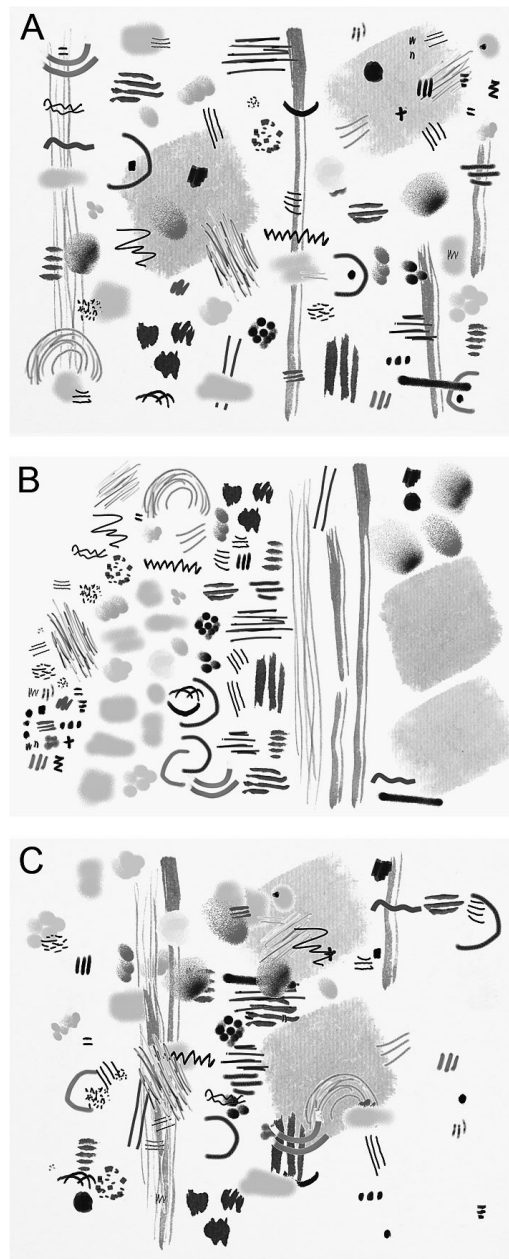


Figure 5. Representative example of the final state of a C.R. drawing (No. 2; A), a version of the same drawing with the pictorial elements roughly ordered according to their size and shape (B; after Wehrli, 2004), and a version with shuffled positions of the pictorial elements (C).

original drawings (0.69 ± 0.05) is higher than that of the shuffled versions (0.57 ± 0.05 ; $p < 0.001$; Fig. 6B) on average and also for each of the individual drawings (data not shown). Finally, the mean self-similarity of the original C.R. drawings is close to that of graphic artworks (0.73 ± 0.10 SD; Fig. 6B).

In the entire dataset, anisotropy decreases during the generation of the drawings ($p < 0.001$; Spearman ρ : 0.50) to reach average values of 0.602×10^{-3} ($\pm 0.057 \times 10^{-3}$ SD) in the final versions. This decrease was observed in 18 out of the 20 series (Spearman ρ range for the 18 drawings: 0.60 to 1.00;

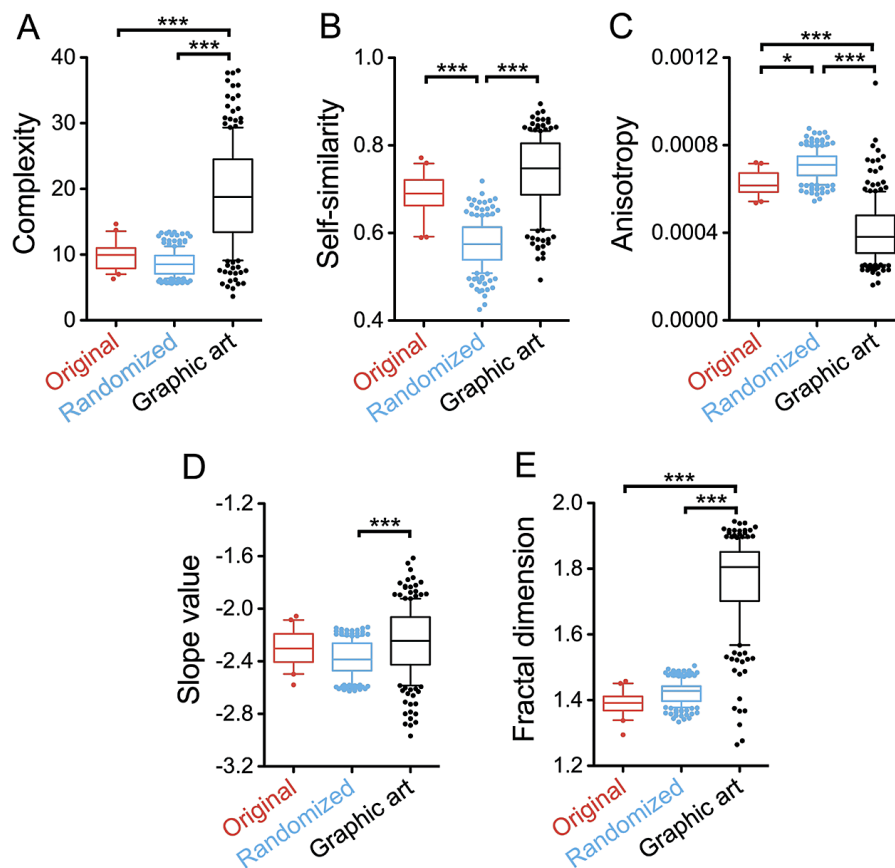


Figure 6. Box plots of the image properties of the 20 original C.R. drawings, the 200 shuffled versions and a set of 200 graphic artworks of Western provenance. Results for complexity (A), self-similarity (B), anisotropy (C), the slope of log–log plots of radially averaged Fourier power (D), and the fractal dimension (E) are shown. The whiskers represent the 10–90 percentiles. The bars above the box plots highlight significant differences (*, $p < 0.05$; ***, $p < 0.001$). This figure is published in color in the online version.

$p < 0.05$ to $p < 0.001$; Supplementary Fig. S3C). The original drawings are less anisotropic ($p < 0.05$) than the shuffled versions ($0.708 \times 10^{-3} \pm 0.065 \times 10^{-3}$ SD) but more anisotropic ($p < 0.001$) than the graphic artworks ($0.405 \times 10^{-3} \pm 0.137 \times 10^{-3}$ SD; Fig. 6C).

The slope of log–log plots of radially averaged Fourier power assumes less negative values in the entire dataset as the C.R. drawings approach the final version (Supplementary Fig. S3D; $p < 0.001$; Spearman ρ : -0.51). This increase was observed for 18 out of the 20 drawings (Spearman ρ range for the 18 drawings: -0.65 to -1.00 ; $p < 0.05$ to $p < 0.001$; Fig. S3D). The mean slope of the original drawings (-2.29 ± 0.15 SD) is about the same as that of the shuffled versions (-2.35 ± 0.14 SD) and the graphic artworks (-2.25 ± 0.26 SD). The slopes values for the shuffled versions are more negative than those of the graphic artworks ($p < 0.001$; Fig. 6D).

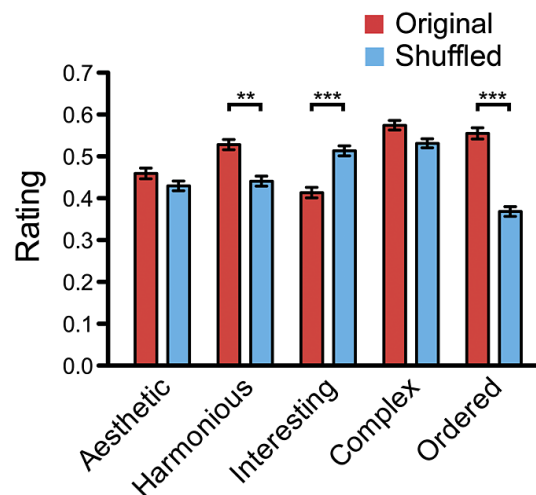


Figure 7. Results of the subjective ratings. Twenty participants provided ratings on how *aesthetic*, *harmonious*, *interesting*, *complex* and *ordered* they found the original 20 C.R. drawings (dark grey/red columns) and 20 of the shuffled versions (light grey/blue columns). The whiskers represent ± 1 SE. The bars above each pair of columns highlight significant differences (**, $p = 0.002$; ***, $p < 0.001$). This figure is published in color in the online version.

Finally, a significant increase in the entire dataset was found for the fractal dimension ($p < 0.001$; Spearman ρ : -0.60), also for 19 drawings analyzed individually (Spearman ρ range: -0.78 to -0.99 ; $p < 0.05$ to $p < 0.001$; Supplementary Fig. S3E). The fractal dimension is similar for the original C.R. drawings (1.39 ± 0.04 SD) and the shuffled versions (1.41 ± 0.04 SD). Both values are lower ($p < 0.001$) than those for graphic artworks (1.76 ± 0.13 SD; Fig. 6E).

3.2.2. Psychological Experiment

The psychological experiment revealed significant overall differences between the ratings for the original and shuffled versions of the C.R. drawings (Fig. 7). The results of the subjective tests were entered into a one-way analysis of variance (ANOVA) considering image category (originals and shuffled [randomized] versions of the images) as within-subject factor. The analysis revealed a significant effect of evaluation on *harmonious*, $F(1, 28) = 11.47$, $p = 0.002$, reflecting increased perceived harmony in the original images (mean: 0.53 ± 0.18 SE) compared to the shuffled versions (mean: 0.44 ± 0.18 SE), as well as for the evaluation on *interesting*, $F(1, 28) = 17.31$, $p < 0.001$, reflecting decreased perceived interestingness in the original images (mean: 0.43 ± 0.14 SE) over the shuffled versions (mean: 0.51 ± 0.14 SE), and for the evaluation on *ordered*, $F(1, 28) = 77.02$, $p < 0.001$, reflecting increased perceived order in original images (mean: 0.56 ± 0.16 SE) compared to the shuffled versions (mean: 0.37 ± 0.16 SE).

Furthermore, we analyzed whether there are correlations (i) within the subjective ratings, (ii) within the statistical properties, and (iii) between the

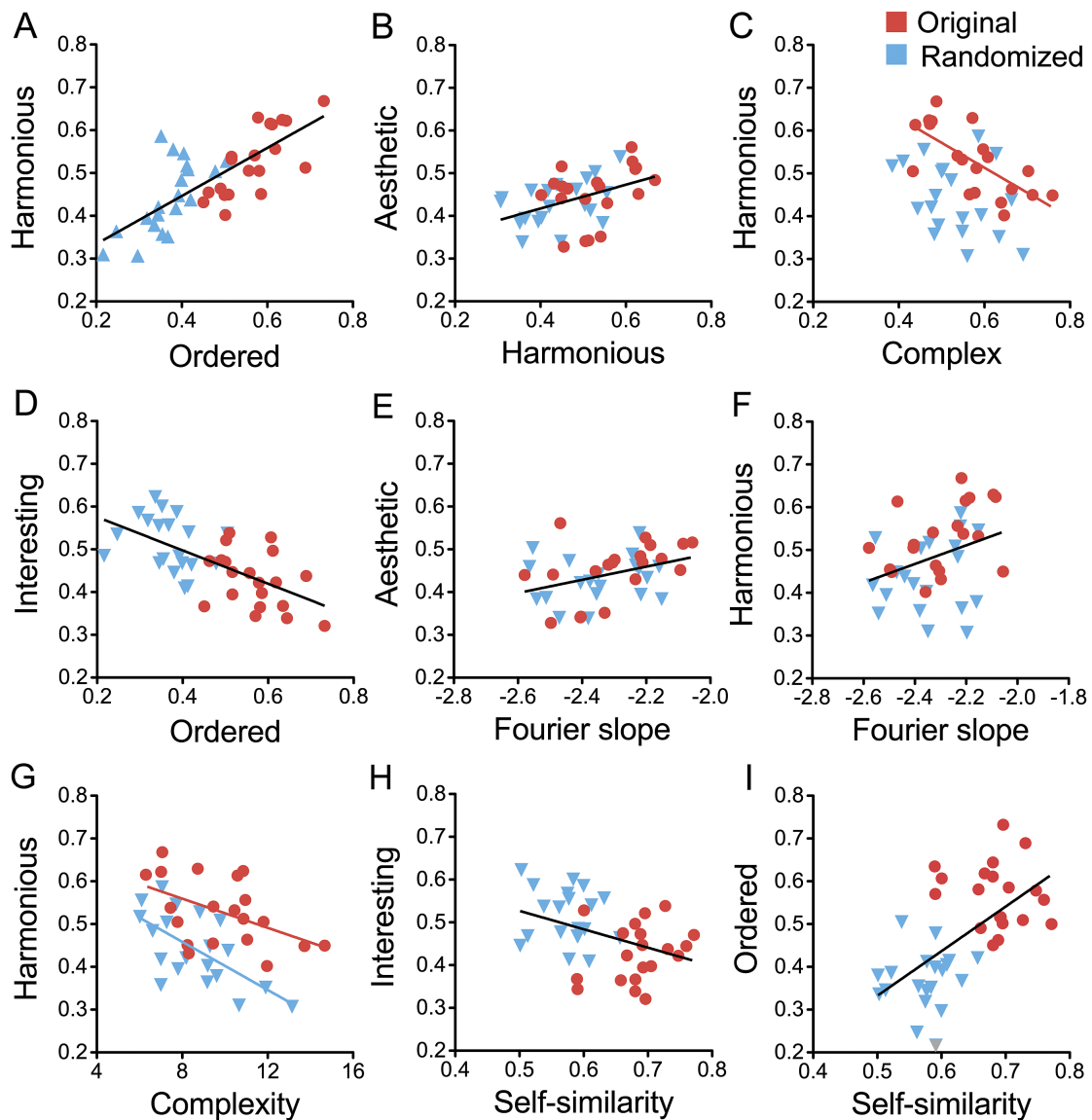


Figure 8. Scatted plots of the correlations within the subjective features (A–D) and between the subjective features and the statistical image properties (E–I) in the C.R. drawings. Each symbol represents results for one image (red circles, original versions; blue triangles, shuffled versions). The lines represent results for linear regression analysis of significant correlations (red lines, original versions; blue lines, shuffled versions; black lines, both versions; compare to Supplementary Table S2). This figure is published in color in the online version.

subjective ratings and the statistical properties. Spearman ρ coefficients for all significant correlations are listed in Supplementary Table S2. Figure 8 shows plots for some of the correlated features.

(i) Positive correlations for the original, shuffled or both versions of the C.R. drawings were found for evaluations on *harmonious/ordered* (Fig. 8A), *aesthetic/harmonious* (Fig. 8B) and *aesthetic/ordered* as well as for *interesting/complex*. Inverse correlations exist for *harmonious/complex* (Fig. 8C), *interesting/ordered* (Fig. 8D), *interesting/harmonious* and *ordered/complex*

(Table S2). Thus, the properties *ordered*, *harmonious* and *aesthetic* and the properties *interesting* and *complex* represent opposing perceptual aspects of the images (Cupchik and Gebotys, 1990).

(ii) With respect to the statistical properties, measured complexity correlates positively with self-similarity; both properties correlate inversely with anisotropy. As expected, we found a strongly positive correlation between complexity and the fractal dimension (Table S2).

(iii) Over all (original and shuffled) drawings, evaluations on *aesthetic* and *harmonious* show a positive correlation with Fourier slope (Fig. 8E, F). For evaluation on *harmonious*, we found an inverse correlation with complexity (Fig. 8G) and the fractal dimension. There is a tendency for *harmonious* to weakly correlate with self-similarity (Spearman ρ : 0.26), but this correlation does not reach statistical significance ($p = 0.10$) in our experiment. Ratings on *interesting* are correlated positively with complexity and inversely with self-similarity (Fig. 8H) and anisotropy. Ratings on *complex* are correlated positively with complexity and self-similarity and inversely with anisotropy. Finally, we found evaluations on *ordered* to correlate positively with self-similarity (Fig. 8I) and inversely with anisotropy and the fractal dimension (Table S2).

4. Discussion

In two state series of monochrome lithographs by Pablo Picasso (*Les Deux Femmes Nues* and *Le Taureau*), we measured higher-order statistical image properties. The lithographs in the series can be regarded as transformations of fully artistic compositions, with the possible exception of the initial 2–3 states (Stolzenburg, 1997). The same statistical image properties have been determined previously in visually pleasing images, including artworks (see Introduction). In the present study, the properties were also studied in 20 state series of abstract drawings created by one of the authors (C.R.). Unlike the Picasso lithograph series, the state series of the C.R. drawings correspond to a more common type of the artistic creation process, in which the drawing process proceeds from a rough preliminary outline of the drawing to several intermediary and unfinished states, to finally reach the artwork's completion, i.e., the aesthetic value can be expected to increase during the creation process. The final states of the C.R. drawings were compared to a datasets of graphic artworks from various Western artists and to versions of the same drawings, in which the constituent pictorial elements were arranged in a shuffled fashion. The measured properties developed in different directions during the creation of the drawings, and also varied between the different categories of images.

In the following sections, we will discuss whether the statistical properties can be linked to aspects of artistic merit in the different types of images. As

outlined in the Introduction, we assumed that (i) the perceived artistic merit of the Picasso lithographs remains high and relatively stable during their transformation. Furthermore, we assumed that the artistic merit of the original C.R. drawings (ii) increases during their creation to reach levels typical for other graphic artworks, and (iii) is higher in the original drawings than in their shuffled counterparts. A shuffled distribution of the constituent elements should diminish the visual quality of the artworks because the overall compositional intent of the abstract artworks depends, at least in part, on the global composition (or spatial arrangement) of the pictorial elements in each drawing. Finally, it seems reasonable to assume that, in general, artists optimize the global composition as creation proceeds. However, in the exceptional case of the two Picasso series, artistic merit seems to remain high throughout the series (see above).

4.1. Complexity and Fractal Dimension

Some of the results for complexity were anticipated. For example, the complexity in both series of lithographs by Picasso changes considerably and in opposite directions during their transformations (Fig. 4A, E). Moreover, the complexity of the C.R. drawings was expected to increase during their creation because more and more pictorial elements were added to the drawings (Fig. 3 and Fig. S3A, E). Similarly, the lack of a difference in complexity between the original and the shuffled versions of the C.R. drawings (Fig. 6A) can be explained by the fact that, for each drawing, the two versions consist of the same pictorial elements with about the same degree of overlap. The fractal dimension shows the same general tendencies as the complexity measure (Fig. 6E) and the two measures are strongly correlated (Table S2), confirming that the fractal dimension is a measure closely related to image complexity (Mureika and Taylor, 2013). Also, it comes as no surprise that patterns of higher measured physical complexity are perceived as more *complex* (Table S2).

The range of values for complexity and the fractal dimension of the graphic artworks dataset is relatively large when compared to that of the C.R. drawings (Fig. 6A, E). In the present study, different degrees of complexity were associated with particular artistic styles. For example, relatively high complexity was observed for the cubist representations in the Picasso series while lower complexity was obtained for simple line drawings. A wide range of complexity values was also found for colored paintings of Western provenance (Redies *et al.*, 2012). Together, these results suggest that artworks can take on rather different complexity values, which may depend on artistic style. Complexity values that are even higher than for artworks were measured for some images of natural patterns, such photographs of lichen growth patterns, vegetation and branches (Redies *et al.*, 2012). This result supports the notion that complex-

ity is not maximized in artworks but assumes of wide range of intermediate values, as suggested before (see Introduction).

4.2. Self-Similarity

Self-similarity seems more closely associated with the images of artistic merit than any of the other measures investigated in the present study. Specifically, in the state series of Picasso lithographs, self-similarity is relatively stable (see assumption (i) above), although it gradually decreases as the *Le Taureau* series approaches a drawing consisting of a few lines only. During the creation of the C.R. drawings, self-similarity tends to increase (Fig. S3B) and finally approaches mean values about as high as that for the dataset of 200 monochrome graphic artworks of Western provenance (ii; Fig. 6B). Moreover, self-similarity is higher for the original versions of the C.R. drawings than for the shuffled versions (iii; Fig. 6B), for each of the 20 drawings. These results are in agreement with findings from other studies that self-similarity is generally high in artworks (Amirshahi *et al.*, 2013; Braun *et al.*, 2013; Redies *et al.*, 2012).

However, the tendency of the evaluation of *harmonious* to correlate with self-similarity did not reach statistical significance in the present study. Also, there is no correlation between *aesthetic* and self-similarity in the C.R. drawings. The original versions are evaluated as being more *harmonious* than the shuffled versions. A positive correlation was found only for *ordered* (Fig. 8I) and *complex* and an inverse correlation for *interesting* (Fig. 8H; Table S2). Therefore, our results do not allow the conclusion that images, which are more self-similar, are perceived as more aesthetic or harmonious. Very high self-similarity values are characteristic for some natural patterns, such as images of clouds, lichen growth patterns or branches (Redies *et al.*, 2012). Such images may be visually pleasing in general, but they would probably not be considered more aesthetic or harmonious than artworks by most observers. In conclusion, as suggested before (Redies *et al.*, 2007a), relatively high self-similarity may be viewed as associated with images with artistic intent, but a causal or linear relationship cannot be inferred from the data of the present study.

4.3. Slope of Log–Log Plots of Fourier Power

The slope of log–log plots of radially averaged Fourier power increases during the creation of the C.R. drawings. This increase can be explained by the addition of smaller pictorial elements, which contribute more power to high frequencies, as the drawing approach their final state. The drawings with slope values closer to -2.0 tend to be evaluated as more *aesthetic* (Fig. 8E) and more *harmonious* (Fig. 8F). Moreover, the slope values of the shuffled versions of the C.R. drawings are lower than those of the graphic artworks (Fig. 6D). This

correlation is of interest because a nearly scale-invariant Fourier spectrum with slope values close to -2 has been found in large subsets of monochrome graphic art of Western and Eastern provenance (Graham and Field, 2007, 2008; Redies *et al.*, 2007a, b). Note that the PHOG-based measure of self-similarity (see above) also measures scale-invariant (self-similar) image properties, but directly in the original image. It is noteworthy that images with a scale-invariant Fourier spectrum can also be generated on a computer and they are not necessarily aesthetic (Lee and Mumford, 1999). Consequently, a scale-invariant Fourier spectrum is not a sufficient condition for high-quality artworks. Deviation from scale-invariance in synthetic images was reported to induce visual discomfort in human observers (Fernandez and Wilkins, 2008; O'Hare and Hibbard, 2011).

4.4. *Anisotropy*

A previous study showed that anisotropy is relatively low in artworks compared to other image categories (Braun *et al.*, 2013; Melmer *et al.*, 2013; Redies *et al.*, 2012). The reasons for this finding are unclear at present. Low anisotropy implies that the strength of the oriented gradients is uniformly distributed across all orientations. In the present study, anisotropy of the original versions of the C.R. drawings is lower than for the shuffled versions (Fig. 6C) and also decreases as the drawings approach their final state (Fig. S3C). However, in the Picasso series, anisotropy is more variable and increases slightly from the first to the last version of both series. We conclude that anisotropy is not consistently associated with images of artistic merit in the present study.

4.5. *Subjective Ratings*

The field of experimental aesthetics notoriously lacks clear definitions of subjective terms, such as beautiful and aesthetic (Augustin *et al.*, 2012). These terms are also subject to controversial disputes in the field of philosophical aesthetics. Attempts were therefore made to clarify some of the interrelations between the terms that refer to the subjective quality of artworks (Augustin *et al.*, 2012; Cupchik and Gebotys, 1990; Jacobsen *et al.*, 2004). Following this general aim for clarification of subjective terms, we asked participants to evaluate the original and the shuffled versions of the C.R. drawings according to how *aesthetic*, *harmonious*, *interesting*, *complex* and *ordered* they were. Because the original and shuffled versions consisted of exactly the same pictorial elements and the drawings were abstract, any differences in the judgments on the two types of images must be attributed to differences in the relative placement of the pictorial elements with respect to each other (image composition) and not to differences in image content that might potentially confound judgments on image composition.

Results from the subjective rating experiment indicated that three out of five subjective attributes differed significantly between the original and shuffled versions of the C.R. drawings (Fig. 7). On average, the participants evaluated the original versions of the abstract drawings as more *harmonious* and *ordered* but, at the same time, they found the shuffled versions more *interesting*. This evaluation was in agreement with the intent of author C.R. to create drawings that follow subjective rules of a harmonious composition (see Methods). Surprisingly, no significant differences were obtained for how *aesthetic* the two versions were evaluated. Possibly, this result can be explained by the different connotations of the term *aesthetic* (Jacobsen *et al.*, 2004), which individual observers may associate with opposing concepts such as harmony or interestingness. Opposite tendencies for evaluation on *aesthetic* might eventually cancel each other when averaging the results over larger groups of persons. Indeed, in our study, a correlation between *aesthetic* and *harmonious* was found for five participants and between *aesthetic* and *interesting* for seven other participants. Only one participant tended to evaluate the more *aesthetic* drawings as both more *harmonic* and more *interesting*.

There were also correlations between the subjective ratings and the statistical image properties, some of which have already been mentioned above. The C.R. drawings of higher measured complexity, which tend to be more self-similar, are evaluated as more *interesting* but less *harmonious* (Fig. 8G, Table S2), confirming the independent nature of the two terms (Berlyne, 1974; Cupchik and Gebotys, 1990). Nevertheless, the subjective ratings depended differently on the two measures (Table S2). Self-similarity is positively correlated with the evaluation of *ordered* (Fig. 8I) and inversely with *interesting* (Fig. 8H) while measured complexity and the fractal dimension is positively correlated with *interesting* and negatively with *harmonious* (Fig. 8G). In agreement with this finding, *interesting* and *harmonious* are negatively correlated (Table S2). In summary, the present results suggest multiple and partially opposing dependencies of the subjective attributes on different objective statistical properties of the C.R. drawings, as described previously for other datasets of images (Augustin *et al.*, 2012; Jacobsen and Höfel, 2002).

4.6. General Discussion and Conclusion

We studied statistical image properties during the transformation of representational drawings in two state series of lithographs by Pablo Picasso and during the creation of self-made abstract drawings. Our results suggest that, among the statistical properties investigated in the present study, self-similarity is the statistical measure that is most closely associated with artistic merit in the datasets of images analyzed.

Results from the subjective rating experiment illustrate the complex facets of aesthetic judgments. They indicate that the subjective evaluations of the

C.R. drawings correlate with specific statistical image properties. It has been proposed that some of the variances in aesthetic judgments between individuals may be due to differences in how important the image properties are for each observer (Augustin *et al.*, 2012). Remarkably, the term *aesthetic* does not discriminate the original and shuffled versions of the C.R. drawings on average, possibly because the participants differ in the concepts, with which they associate this term. Moreover, not only the perception of artworks differs, but artists themselves may use different strategies when creating their artworks. It should therefore be stressed that the present findings are restricted to a special type of abstract drawings that were created by a single person. Before reaching general conclusions on artistic production, it will be necessary to extend this type of study to artworks by other artists. The present work may serve as an example on how this type of investigation can be approached.

Aesthetic judgments are not only complex and biased by personal preferences but they also involve cognitive processes, such as judgments of familiarity and cognitive mastering, which depends on previous exposure to art and knowledge about art (Leder *et al.*, 2004). Nevertheless, we here illustrate that the contribution of perceptual mechanisms to the evaluation of visual image quality in artworks can be studied in isolation in principle, if well-controlled visual stimuli are used. Therefore, we used original and shuffled versions of abstract stimuli that differ only in their statistical properties, not in image content. In this respect, our work extends previous work by Jacobsen and co-workers who generated a series of simple geometric patterns with well-defined properties to study aesthetic perception (Jacobsen and Höfel, 2002; Jacobsen *et al.*, 2006). Going beyond this previous study, we used more complex visual patterns that resemble artistic drawings. By shuffling the pictorial elements in each drawing, we deliberately abolished the overall compositional intent of the artist. This approach allowed us to correlate subjective evaluations of images of higher and lower artistic intent with statistical image properties. The generation of this type of stimuli is subject to technical limitations (e.g., to restrictions in shuffling the pictorial elements) and will require close collaboration between scientists and artists. Artist-scientists are predestined to play a role in such studies, but if they are involved, the artistic process should be clearly separated from the image analysis, as was the case in the present study, in order to enforce impartial observations.

Acknowledgements

We thank Sylvia Hänßgen and Claudia Menzel for help in the psychological experiment, members of the Experimental Aesthetics Group and the Computer Vision Group (head: Dr Joachim Denzler) for discussion and critical feedback. Three reviewers provided highly constructive criticism on an earlier version of the manuscript.

References

- Alvarez-Ramirez, J., Ibarra-Valdez, C., Rodriguez, E. and Dagdug, L. (2008). 1/f-noise structures in Pollock's drip paintings, *Physica A* **387**, 281–295.
- Amirshahi, S. A., Koch, M., Denzler, J. and Redies, C. (2012). PHOG analysis of self-similarity in esthetic images, in: *Proc. SPIE (Human Vision and Electronic Imaging XVII)* 8291, 82911J.
- Amirshahi, S. A., Redies, C. and Denzler, J. (2013). How self-similar are artworks at different levels of spatial resolution?, in: *International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging 2013*, pp. 93–100. Association of Computing Machinery, New York, USA.
- Augustin, M. D., Wagemans, J. and Carbon, C.-C. (2012). All is beautiful? Generality vs. specificity of word usage in visual aesthetics, *Acta Psychol.* **139**, 187–201.
- Berlyne, D. E. (1974). The new experimental aesthetics, in: *Studies in the New Experimental Aesthetics*, D. E. Berlyne (Ed.), pp. 1–25. Hemisphere Publishing, Washington, DC, USA.
- Bosch, A., Tisserman, A. and Munoz, X. (2007). Representing shape with a spatial pyramid kernel, in: *Proc. 6th ACM Int. Conf. Image Video Retrieval*, pp. 401–408.
- Braun, J., Amirshahi, S. A., Denzler, J. and Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture, *Front. Psychol.* **4**, 808.
- Burton, G. J. and Moorhead, I. R. (1987). Color and spatial structure in natural scenes, *Appl. Phys.* **26**, 157–170.
- Canny, J. (1986). A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, 679–698.
- Cupchik, G. C. and Gebotys, R. J. (1990). Interest and pleasure as dimensions of aesthetic response, *Emp. Stud. Arts* **8**, 1–14.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection, *Int. Conf. Comp. Vis. Pattern Recogn.* **2**, 886–893.
- Fernandez, D. and Wilkins, A. J. (2008). Uncomfortable images in art and nature, *Perception* **37**, 1098–1113.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **4**, 2379–2394.
- Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J. and Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art, *Br. J. Psychol.* **102**, 49–70.
- Gauss, U. (2000). *Pablo Picasso. Die Lithographie*. Hatje Cantz, Ostfildern, Germany.
- Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes, *Annu. Rev. Psychol.* **59**, 167–192.
- Graham, D. and Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception, *Vis. Res.* **50**, 1503–1509.
- Graham, D. J. and Field, D. J. (2007). Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities, *Spat. Vis.* **21**, 149–164.
- Graham, D. J. and Field, D. J. (2008). Variations in intensity statistics for representational and abstract art, and for art from the Eastern and Western hemispheres, *Perception* **37**, 1341–1352.
- Green, C. (2005). *Picasso: Architecture and Vertigo*. Yale University Press, New Haven, CT, USA.

- Jacobsen, T., Buchta, K., Kohler, M. and Schroger, E. (2004). The primacy of beauty in judging the aesthetics of objects, *Psychol. Rep.* **94**, 1253–1260.
- Jacobsen, T. and Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: analyses of individual judgments, *Percept. Mot. Skills* **95**, 755–766.
- Jacobsen, T., Schubotz, R. I., Höfel, L. and Cramon, D. Y. (2006). Brain correlates of aesthetic judgment of beauty, *Neuroimage* **29**, 276–285.
- Koch, M., Denzler, J. and Redies, C. (2010). 1/f² Characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs, *PLoS One* **5**, e12268.
- Leder, H., Belke, B., Oeberst, A. and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments, *Br. J. Psychol.* **95**, 489–508.
- Lee, A. B. and Mumford, D. (1999). An occlusion model generating scale-invariant images, in: *Proc. IEEE Workshop Statist. Comput. Theor. Vis.*, pp. 1–20. Fort Collins, CO, USA.
- Melmer, T., Amirshahi, S. A., Koch, M., Denzler, J. and Redies, C. (2013). From regular text to artistic writing and artworks: Fourier statistics of images with low and high aesthetic appeal, *Front. Hum. Neurosci.* **7**, 106.
- Mourlot, F. (1970). *Picasso Lithographs*. Boston Book and Art Publisher, Boston, MA, USA.
- Mureika, J. R. and Taylor, R. P. (2013). The abstract expressionists and les automatistes: A shared multi-fractal depth? *Sign. Proc.* **93**, 573–578.
- O’Hare, L. and Hibbard, P. B. (2011). Spatial frequency and visual discomfort, *Vis. Res.* **51**, 1767–1777.
- Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* **381**, 607–609.
- Parraga, C. A., Troscianko, T. and Tolhurst, D. J. (2000). The human visual system is optimised for processing the spatial information in natural visual images, *Curr. Biol.* **10**, 35–38.
- Redies, C. (2007). A universal model of esthetic perception based on the sensory coding of natural stimuli, *Spat. Vis.* **21**, 97–117.
- Redies, C., Amirshahi, S. A., Koch, M. and Denzler, J. (2012). PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects, in: *ECCV 2012 Ws/Demos, Part I, Lect. Notes Comp. Sci.* **7583**, pp. 522–531.
- Redies, C., Hänisch, J., Blickhan, M. and Denzler, J. (2007a). Artists portray human faces with the Fourier statistics of complex natural scenes, *Network* **18**, 235–248.
- Redies, C., Hasenstein, J. and Denzler, J. (2007b). Fractal-like image statistics in visual art: similarity to natural scenes, *Spat. Vis.* **21**, 137–148.
- Ruderman, D. L. and Bialek, W. (1994). Statistics of natural images — Scaling in the woods, *Phys. Rev. Lett.* **73**, 814–817.
- Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation, *Annu. Rev. Neurosci.* **24**, 1193–1216.
- Stolzenburg, A. (1997). *Picasso und die Lithographie*. “Ich mache ein Bild und zerstöre es”. Seemann, Leipzig, Germany.
- Taylor, R. P., Spehar, B., Van Donkelaar, P. and Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock’s fractals, *Front. Hum. Neurosci.* **5**, 60.
- Wehrli, U. (2004). *Kunst aufräumen*, 2nd edn. Kein und Aber, Königstein im Taunus, Germany.
- Wiltschut, J. and Hamker, F. H. (2009). Efficient coding correlates with spatial frequency tuning in a model of V1 receptive field organization, *Vis. Neurosci.* **26**, 21–34.