Kent Academic Repository
Full text document (pdf)

Citation for published version

Johnson, Colin G. and McCormack, Jon and Santos, Iria and Romero, Juan (2019) Understanding Aesthetics and Fitness Measures in Evolutionary Art Systems.   Complexity .   ISSN 1076-2787.   (In press)

DOI

Link to record in KAR

https://kar.kent.ac.uk/72848/

Document Version

Author's Accepted Manuscript

Understanding Aesthetics and Fitness Measures in Evolutionary Art Systems

Colin G. Johnson[1,+]**, Jon McCormack**[2,+]**, Iria Santos**[3,+]**, and Juan Romero**[3,+]

[1]School of Computing, University of Kent, Canterbury, UK. Email: C.G.Johnson@kent.ac.uk
[2]SensiLab, Monash University, Caulfield East, 3145, Australia. Email: Jon.McCormack@monash.edu
[3]University of A Coruña, A Coruña, Spain. Email: iry69.2@gmail.com, juan.romero1@udc.es
[+]these authors contributed equally to this work

ABSTRACT

One of the general aims of evolutionary art research is to build a computer systems capable of creating interesting, beautiful or creative results, including images, videos, animations, text, and performances. In this context it is crucial to understand how fitness is conceived and implemented to explore the 'interestingness', beauty or creativity that the system is capable of. In this paper we survey the recent research on fitness for evolutionary art related to aesthetics. We also cover research in the psychology of aesthetics, including relation between complexity and aesthetics, measures of complexity and complexity predictors. We try to establish connections between human perception and understanding of aesthetics with current evolutionary techniques.

Introduction

An ancient dream of humanity is to create models of itself. Ada Lovelace, often attributed as the first computer programmer, proposed to use computers for artistic tasks. Such tasks constitute a "grand challenge" since they present a series of subjective, social and emotional characteristics than are often considered exclusive to human cultures.

One of the main difficulties in addressing this challenge is in developing formal models of human aesthetic preference. Such models would allow computer systems to predict the aesthetic taste of a human being or adapt to the aesthetic tendencies of a human group: in simple terms, to be able to make aesthetic evaluations and choices.

The term 'Aesthetic' derives from the Greek *aisthesis*, denoting feeling or perception, and its original meaning referred to sensory impressions. In the 18th century it acquired a new meaning when Baumgarten's 'Meditationes Philosophicae de Nonnullis ad Poema Pertinentibus' was published in Germany. This identified the relation between sensory experience and knowledge, and gave the study of the knowledge of beauty the name *aesthetics* . From this moment, the term aesthetics is not restricted to the arts, but many of the things and experiences encountered in daily life. Hence aesthetic decisions affect many many aspects of human choice and action, beyond those traditionally associated with fine art for example.

Computational Aesthetics (CA) can be defined as 'the research of computational methods that can make applicable aesthetic decisions in a similar fashion as humans can.' (Hoenig, 2005). There are several papers that survey approaches to computational aesthetics, e.g. (Greenfield, 2005; Galanter, 2012; Birkin, 2010). The term 'computational aesthetics' is sometimes used in the sense of describing a particular class of artefacts made by computers, e.g. computer design and generative systems. However, in this paper we will refer to computational aesthetics only as computational models of human aesthetics.

CA and the Psychology of Aesthetics (PA) have studied human aesthetics using a variety of different approaches. In this paper we attempt to establish connections between these different approaches.

In the first section we analyse several aesthetic modes included in recent evolutionary computation systems. The second section explores research results from psychology of aesthetics that will be of interest to AI researchers. Finally, we then propose some connections between the efforts of human psychology and AI and outline the advantages of this collaboration for both groups.

Aesthetic Fitness in Evolutionary Art

Since classical times, if not before, philosophers have engaged in the study of aesthetics: attempting to understand the nature of art and its appreciation, and why people engage in a specific set of *aesthetic behaviours* around artworks and make particular statements that can be called *aesthetic judgements*. In more recent times, these philosophical investigations have been joined by experimental and observational methods from psychology, neuroscience and cognitive science, as well as research taking

a constructive stance by building machines that produce work of aesthetic value, or machines that can themselves exhibit aesthetic behaviour and make aesthetic judgements.

A number of major strands exist in aesthetic theory (Carroll, 1999; Gaut and McIver Lopes, 2013). Early classical work focused on attempts to understand the nature of beauty—not just pointing out specific examples of beauty, but identifying those aspects of objects that give rise to aesthetic appreciation. The earliest theories focused on the appreciation of skill in imitating the physical world, but later theorists found this lacking. In particular, once mechanical means for creating high-quality imitations were available such as photography and sound recording, new theories to explain the difference between simple reproductions and objects of aesthetic value became needed.

One major strand of aesthetic theory is based around the idea that aesthetic appreciation is a deliberate *act of expression* by the art-maker. In these theories, the art-maker is concerned with transmitting experiences and emotions that they have experienced to the audience for their work, in a way that cannot be readily done using more direct means such as purely descriptive text or diagrams. This gives rise to an immediate problem for computer art systems, which have no emotional qualia to form the grounds for expression. Nonetheless, even for machine-made art, we can sometimes recover some value from such expression-focused theories. One kind of computer art that can be said to be expressive is that which exploits the fact that computers are now almost universally are networked systems, and for the expression to be an expression of the zeitgeist around a particular area as discovered online. For example, one version of the *Painting Fool* system (Krzeczkowska et al., 2010) uses newspaper articles as the material that it 'expresses' in a visual art form; the importance and salience of the source material comes from the fact that is important enough to form a newspaper headline. Another way in which we can see expression explaining the impact of computer art systems is in those systems that act as a shaper and reinforcer of the user's interactions: not acting as an expressive device of the computer's own (absent) feelings, but allowing the user to explore and reinforce their expressions in a way that isn't possible without the machine.

Another major strand of aesthetic theory is concerned with ideas of form. These theories argue that what makes an aesthetically engaging object distinct from a mundane one are formal aspects such as the placement of objects in an image, the use of symmetry, and the balance between order and complexity. The content is less relevant—aesthetic objects still have content, of course, but broadly similar content arranged without regard to form will have little aesthetic interest. Such theories are appealing to explain how computer art systems can create aesthetically engaging objects, because aspects of form can be encoded algorithmically, measures of form can be used as fitness drivers within learning and evolution systems, and different aspects of form can be brought together using multi-criteria optimisation.

Another strand argues for the importance of social interactions and the social construction of aesthetic value, whether within a specific social discourse around art (e.g. Danto's (1964) idea of the *Artworld*) or by being influenced by, and influencing, wider social and political issues. This has been occasionally explored in computer art systems (Machado et al., 2004; Romero et al., 2003), by modelling a wider network of systems that create art and systems that critique and contextualise that art; however, there is much opportunity for more work in this area.

More recently, the focus has shifted from the wider world to the inner world of the brain and nervous system, examining the brain during aesthetic experiences (Chatterjee, 2013). Again, there are opportunities for this to be used in the context of fitness drivers for evolutionary art systems by modelling the potential audience response to art, much as user modelling (Fischer, 2001) model the user response to more prosaic systems.

In contrast with theories that argue aesthetics is a social phenomenon, other philosophers of aesthetics have taken a position that there are—at least at a very high level—some common features to aesthetic objects and to the act of aesthetic appreciation that remain constant over time. Dutton (2013), for example, lists seven 'aesthetic universals' that he claims form a feature of most social practices that are regarded as art. These are that:

- the production of art objects requires skill and expertise;

- the objects give pleasure in-and-of themselves, regardless of whether they satisfy a practical need;

- art is produced in styles that are socially developed and are primarily about form and composition;

- art exists in the context of a critical and analytical discourse;

- art objects imitate or symbolise aspects of the wider world;

- art objects are the subject of a special kind of attention and evoke particular behaviours towards them;

- audiences engage in art by using their faculties of imagination, and that artists make use of imagination in creating and developing artistic ideas and objects.

There is not necessarily a conflict between the idea of universals and the idea of social construction of aesthetics. It could be argued that whilst the broad categories of concepts that characterise art and aesthetic behaviour are broadly universal, specifics vary with time in a socially constructed way. Indeed, a major model of aesthetic appreciation and aesthetic judgement developed by Leder and colleagues uses an information-processing relationship between components that integrate into an aesthetic episode (Leder et al., 2004; Leder and Nadal, 2014). The model includes low-level 'universal' aesthetic properties, such as symmetry, complexity, contrast and grouping, but also social, cognitive and emotional components that all contribute in forming an aesthetic judgement.

Evolutionary Art Systems

Evolutionary art systems are computer systems that employ evolutionary computation (EC) methods to generate artworks (Lewis, 2008).Evolutionary art system have been devised that create drawings, designs, buildings, poetry, sounds, music, 3D forms, images, and even choreography. Typically, the way in which these systems vary from other applications of EC is in the fitness function; other aspects of EC (selection methods, crossover and mutation operators, etc.) are largely the same as in more traditional optimisation applications. Such fitness functions can give rise to aesthetic value in two main ways. The first is *explicitly*, where the fitness function drives the evolutionary search towards items of greater aesthetic value. The second is *endogenously*, where the fitness creates a process that is itself of aesthetic value. An example of the latter is the body of work in artificial life art and artwork based on simulated ecology (McCormack, 2012). These might reflect a shift back towards an aesthetics of imitation in a new way—by simulating processes that occur on a temporal or spatial scale that is inaccessible to naked-eye viewing, they imitate/represent natural processes in a scaled or abstracted way making them accessible to immediate perceptual apprehension. This allows unspecialised audiences to reflect on these processes which are otherwise only comprehensible to scientists.

An evolutionary system that aims to generate aesthetically engaging material explicitly should therefore have a fitness function that drives the evolution toward areas of a search space that are aesthetically valued. So, the fitness function should be grounded in some theory of aesthetics; perhaps one of the established theories, or perhaps a new kind of theory that is distinctive to computer art or evolutionary art. Johnson (2016) reviews a number of possible ideas on which such fitness functions could be built.

The most direct way to do this is via some kind of *aesthetic measure*. That is, the fitness function directly enacts some algorithmic method of scoring or ranking the aesthetic value of a specific work. This fits particularly well with aesthetic theories based around form—the most typical measures used are measures of formal aspects such as symmetry and complexity. In our discussion below on the psychology of aesthetics we will see that this is the dominant theory there too; much of the experimental work in this area explores correlations between formal aspects of visual images and the viewer's aesthetic or affective responses.

One of the most influential EC-art papers in recent times that uses aesthetic measure is that of den Heijer and Eiben (2010), which compares four different aesthetic measures as fitness functions for a EC system. The paper shows the results of the different functions in using them as fitness measures with an EC and by calculating the cross-evaluation of each function with the others. However, the problem with this type of approach is that whilst the functions proposed are useful as tools to explore the capabilities of EC, their connection to human aesthetic judgement is not clearly explained prior to their being employed as fitness functions. In some cases, the functions (called 'measures' in the paper) were employed as metrics in learning systems, so they can be used for aesthetic purposes, but not necessary alone. In fact, one of the metrics analysed in the paper—the one first proposed by Machado and Cardoso (1998)—was designed for monochrome images, but applied in this research to colour ones.

Another way to create this fitness function is via a corpus of examples, typically in the form of an *inspiring set* (Ritchie, 2007) of examples that the computer system should use to inspire work that is new but in a similar style. The features provided to the learning system from the examples will dictate what aesthetic theories are underpinning this use of the corpus. For example, a system that uses geometrical analysis of the corpus examples, or extracts features based on the histogram of colours in the image, is driving towards aesthetics based around form or colour distribution. By contrast, if the system were using sentiment analysis to extract emotional cues from the corpus, this can be seen as working closer to expression and perceived emotion theories.

Another way to assign fitness is for the system to use interaction with people in place of a fixed function (Takagi, 2001). In terms of aesthetic theories, this leaves the theory to the user—rather than a computational fitness function being used, the decision on fitness is referred to a human, who can apply their own aesthetic judgement without having necessarily to theorise it formally. One under-explored area for future work would be for the human making the judgement to provide a more detailed critique of the work rather than just a selection or score, in some computer-readable form. This fits into a recent trend in evolutionary computation which uses richer *fitness drivers* containing much more information than a simple score or ranking (Krawiec et al., 2016) for selection and focused mutations. We can see this as fitting into a more social, critic-based theory of aesthetics, where human critics engage in a discourse with established or emerging artwork traditions.

Some Findings from the Psychology of Aesthetics

There is some overlap between current research on PA and CA. As an example, there are some researchers in PA looking for measures of aesthetic value or visual complexity. But at the same time, looking at the cross-citation of both areas, there is little communication between them. This section will analyse some of the findings in PA from the point of view of an AI researcher. We hope that this can help in creating computer systems that work with concepts such us visual complexity, aesthetics, symmetry, and so on.

Firstly, a set of PA experiment only done with human beings are explored that relate aesthetic judgements to the complexity of the work produced. Next, we explore briefly some works that employ algorithmic measures of complexity, and other works that try to model visual complexity. Then we review research that relates measurable properties of images to visual perception in the form of fractal analysis. Finally, we focus on possible aesthetic tests from PA that could be useful in AI research.

Before researchers came up with practical experiments in the psychology of aesthetics, questions related to art and aesthetics were answered by means of theories and experiences of the theorists themselves. The majority of cases were based exclusively in the observation of the reactions of a few viewers contemplating artistic works. Such informal processes, whilst useful for clarifying ideas, do not provide a strong basis for implementable theories.

In 1876, Fechner published 'Elements of Aesthetics' (Fechner, 1876), where he described a study based on the observation of the diverse answers of representative subjects of distinct populations with different visual material. These investigations laid the foundations and experimental methods for hypothesis formulation in aesthetics and its verification under controlled conditions.

Once this experimental basis had been established (Fechner, 1871), the next step was to determine a method that was able to quantify the aesthetics of an object. The complex dimensions of a work of art for example (form and location of the lines, rhythmic sequences, variations of tone, etc.) are from this moment objects of measurement: first the mathematician Birkhoff and later Eysenck would propose the first formulas for aesthetic measure. They were used as a measure of the aesthetic 'value' in a number of different experiments and, as we shall see below, with contradictory results.

Experiments on Visual complexity and aesthetic

In the 1930's Birkhoff set out the first mathematical formula that was designed to measure aesthetic value. This formula asserts that for visual objects, the aesthetic measure of the object (M) is related to its order (O) and complexity (C), specified in the relationship:

$$M = \frac{O}{C} \tag{1}$$

Equation 1 proposes that the aesthetic measure of a image is correlated with the order and simplicity/complexity of its visual stimuli. Together with the presentation of that formula, Birkhoff (1933) defined complexity as an expression of multiplicity, such as the number of elements that make up an image, while the order describes the regularity of those elements (repetition and redundancy).

While Birkhoff provided many different visual examples, he did not carry out experiments to validate his hypothesis. Even so, there are several research papers on his theory, some of which offer widely differing results. On the one hand, Brighouse (1939) and Meier (1942) conclude that the theory of Birkhoff is empirically founded, while, on the other hand, Weber (1931), Beebe-Center and Pratt (1937), Davis (1936) and Eysenck (1942) are not in agreement with this hypothesis. The most complete study related to the Birkhoff hypothesis was carried out by Eysenck (1941b,a, 1942). Previously, Eysenck himself had carried out experiments related to this theory, exposing his disagreement with it. In order to be able to provide an alternative measure, he performed his own experiment in a controlled environment. A total of 11,000 participants, including those with art studies and without them (artists, students, teachers and psychologists), showed different series of polygons, and asked to sort them according to their aesthetic preferences. These polygons were part of the material provided by Birkhoff (1933). From the experiment, Eysenck presents a formula different from that of Birkhoff, although also based on ideas of order and complexity. In this case the relationship with complexity is positive, since both order and complexity were found to positively contribute to the appreciation of beauty.

It should be noted that the images employed by both Birkhoff and Eysenck are images with set of polygons, created for the experiment (not real world images) and that both researchers do not have exact, much less computational measures that allow quantifying order or complexity.

Berlyne (1963) proposed that judgements about the interest and liking of an image depend, fundamentally, on the judgement of the complexity of that stimulus (Berlyne et al., 1968). This, in turn, is related to factors such as the regularity of the model, the number of elements that make up the scene, its heterogeneity, or the irregularity of the forms (Berlyne, 1970). The optimum of aesthetic pleasure would be latent until a subject encountered stimuli of average complexity, in case of having a very moderate

Figure 1. The twelve most-liked (left) and least-liked (right) of Eysenck's (1941b) experiments.
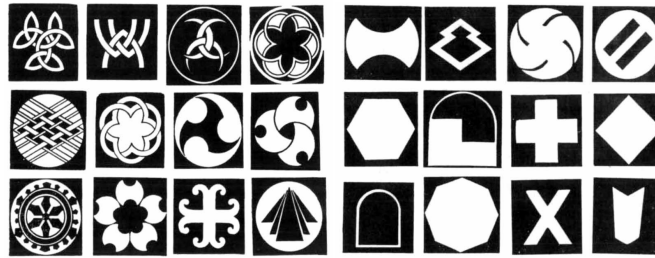
stimulation potential, or stimuli that imply a very high potential, but reducible by appropriate modifications. This optimum varies according to learning (Frances, 1985).

Aesthetic preferences and judgements of beauty have been the subject of numerous research experiments since the formulation of order and complexity by Berlyne. Their hypotheses have been the subject of study following two different approaches: one based on general visual stimuli and another on artistic stimuli. In the case of visual stimuli, Aitken (1974), Katz (2002) and Vitz (1966) use geometric objects while Heath et al. (2000), Ichikawa (1985) and Stamps III (2002) perform their experiments with artificially generated images. With a focus based on artistic stimuli, we highlight the work carried out by Krupinski and Locher (1988), Nicki and Moss (1975), and Osborne and Farley (1970) by means of abstract paintings, Nicki et al. (1981) with works of Cubist art, Messinger (1998) using figurative images and Saklofske (1975) by means of portraits.

The conclusions obtained across the experiments are contradictory, even within the same approach. Some find a distribution of preference in the form of an inverted U, with preference given to intermediate levels of complexity, whilst others observe a linear increase of aesthetic engagement with increasing complexity. A more detailed breakdown and analysis can be found in the paper by Nadal (2007).

Berlyne himself (1970) expressed a problem in conceptualisations of visual complexity. Attneave (1957) and Berlyne (1974) surveyed the subjective aspect of visual complexity. However, some experiments that use classification scales and other techniques confirm that collative variables and subjective information variables tend, as expected, to vary concomitantly with the corresponding objective measure of the classical theory of information (Cupchik and Berlyne, 1979). Hogeboom explained that the complexity perceived by each individual depends on the way the scene is organized (Hogeboom and van Leeuwen, 1997; Strother and Kubovy, 2003). This may be one of the reasons that the previous conclusions were contradictory.

Forsythe et al. (2008) demonstrated that the subjective image complexity measure can be conditioned by familiarity. In Nadal et al. (2010), a group of individuals rated the beauty and complexity of a set of images. The authors could not find any correlation between ratings. The researchers proposed three different types of complexity that can influence visual perception of complexity (asymmetry, the amount and variety of objects, and the way the objects are organised).

Also using ideas of priming and conditioning, Mallon et al. (2014) studied the changes in the evaluation of the perceived beauty in abstract artworks, and maintain that the perceived beauty increases after the exhibition of paintings that have been described as less beautiful, and diminishes after the exhibition of paintings that were described as the most beautiful, which again reinforce the idea of subjectivity in aesthetic appreciation.

Güçlütürk et al. (2016) call for a focus on individual differences in aesthetic preferences, and the adoption of alternative methods of analysis that take into account these differences, along with a re-evaluation of the established rules of aesthetic preferences in humans. The relationship between aesthetic taste and stimulus complexity is commonly defined as an inverted U-shaped curve; images that are too simple offer too little to appeal to the aesthetic sense, whereas excessively complex images present too many diverse stimuli to allow aesthetically engaging patterns to be identified. However, frequent individual differences between the preferences of the participants' complexity have been observed since the first studies on the subject. The usual use of methods of linear analysis that ignore these great individual differences in aesthetic preferences gives an impression of high level of coincidence between individuals. In their study they gather the qualities of taste and perception of the complexity of 30 participants for a set of 144 digitally generated grayscale images. In addition, an objective measure of the complexity of each image is calculated. The authors claim that the results show that the U-shaped relationship between the taste and the complexity of the stimulus is produced as the combination of different individual functions of taste. Specifically, after automatically grouping the participants in relation to their taste qualifications, they determine that a group of sample participants assigned increasingly lower quality of taste for more and more complex stimuli, while a second group of participants had scores of taste increasingly higher for more and more complex stimuli. The two groups differ as to whether they prefer complex or simple patterns, but not in the way in which they perceive the complexity. The group of participants who prefer the simplest patterns were faster in their taste assessments compared to the group that preferred complex patterns. These differences in the

assessment time were not found in the evaluation of complexity. A partial explanation of the results is provided by the theory of fluidity Reber et al. (2004), according to which experience in fluid processing has a positive effect on the stimulus, so a decrease in taste towards complex stimuli could be expected (and therefore processed with less fluidity) compared to simpler stimuli (and processed more fluidly). This would validate the results of a group, but not those of the other.

A recent framework by Graf and Landwehr (2015) called PIA (Pleasure-Interest model of Aesthetic liking), aims to provide a better explanation of the contradictory patterns of preference for aesthetic stimuli that are easy or difficult to process. According to the authors, an aesthetic object can be processed in two stages. In the first stage an automatic processing is carried out, and then, if the viewer is sufficiently motivated to continue the processing of the stimulus, there is a controlled processing. Similarly to the theory of fluidity, the PIA model predicted that purely automatic processing of the stimuli results in a decrease in taste as the complexity of them increases. The prognostic model furthermore that the controlled processing could give rise to an inverted U curve, if the levels of complexity of the stimuli are sufficiently high to cause disgust and confusion.

Measurements of image complexity

After exploring several PA ideas that try to analyse the relation between visual complexity and aesthetics using an ad hoc determination of complexity, we will move on in this section to survey some works that employ algorithmic measures of complexity.

As stated previously, perception of image complexity is subjective. The first method to calculate the complexity of a set of images is to relate complexity with another objective factor of the image. As an example, complexity could be related to the number of objects in an image. So, a constructed image of two triangles has a complexity of 2, while a constructed image with 9 triangles has a complexity of 9. Similar approaches use the number of *different* objects, and other objective qualities of constructed images. The first works analysed in the previous section employ this method by constructing the images used in the dataset (typically with combinations of polygons and other simple forms).

A different approach in order to determine the complexity of a image is to ask a group of people self-report the perceived complexity and calculate the average of the responses. This gives a complexity measure for images that were not created specifically for the experiment (such us paintings or real world photographs) (Bonin et al., 2003; Alario and Ferrand, 1999; Cycowicz et al., 1997). This method was employed on most of the papers presented in the previous section. While this method is not limited to any specific kind of image, it may present a significant time or resource cost if the image corpus is large.

A computer generated measure of complexity can be applied to images with relatively little cost so it can be used to feed computer systems that generate images or other novel images (Hochberg and Brooks, 1960). Moreover, it can allow us to determinate the factors (emotional, semantic, etc.) that affect the human perception of image complexity, through a proportionate objective measure. Hence it can be used to analyse the differences between objective and subjective values in different types of images. We will see a clear example of that later in the work of Jakesch and Leder (2015). Moreover, as we will see, some PA researchers such as Forsythe et al. (2017) suggest that the objective measure (based on calculated metrics) can be more useful to predict human aesthetic preference than the subjective one (based on human scores).

Hochberg and Brooks (1960) created a semi-automated measure of image complexity, based on the combination of number of interior angles, different angles and lines. García et al. (1994), developed an algorithm to measure the image complexity of icons using the number of lines (horizontal, vertical and diagonal), forms (open and closed) and letters in each icon. Mcdougall et al. (1999) employ the same measure for the complexity of a set of forms and achieve a correlation with the judgement of humans of Rs=0.73 for abstract icons.

Forsythe et al. (2003) created an automatic system to measure the complexity of icons based on edge information and structural variability. They found high correlation between their scores and those provided by Garcia et al (Rs=0.66 for edge information and Rs=0.65 for structural variability), and also for the studies of McDougall et al (Rs=0.64 for edge information and Rs=0.65 for structural variability). To our knowledge this system is the first example published in psychology that employs a computational metric to measure complexity.

In AI research, Machado and Cardoso (1998) propose visual complexity metrics based on the compression rate and error of JPEG and Fractal compression. This was based on ideas from Arnheim (1956, 1966, 1969) and Moles (1958). They base their measure of image complexity on findings from information theory. Other authors propose similar theories (Leeuwenberg, 1969; Simon, 1972; Schmidhuber, 1998), where the complexity is related to the unpredictability of the image (of the pixels in the image) (Salomon, 2006). As a highly unpredictable image is not easy to compress, they used the length of the compressed file and the degree of error as estimates for the predictability of the image. Equation 2 shows the formulation of the measure.

$$Visual\_Complexity\_Measure = \frac{RMS\_Error}{Compression\_Ratio} \tag{2}$$

In PA, Donderi and colleagues (Donderi et al., 2003; Donderi, 2006) were also inspired by algorithmic information theory. They used JPEG and ZIP compression as an approximation of the minimum code to describe an image, as a consequence

estimating the predictability of the image. An image with all pixels black is (in principle) easy to compress and is readily predictable. On the other hand, a random generated image with no relation between each pixel is not predictable at all and is also not compressible. In Donderi and McFadden (2005) the authors get a correlation of Rs=0.77 between the length of JPEG and ZIP compressed files and subjective image complexity.

Forsythe et al. (2008) presented four metrics based on perimeter, Canny, JPEG and GIF compression. They tested these metrics with a number of previous datasets, showing high correlations with subjective complexity. In 2011, Forsythe et al. (2011) analysed the correlation between the perceptual image complexity using several algorithmic measures: (i) length of JPEG compression (ii) length of GIF compression and (iii) perimeter detection measures. The authors employ a dataset of 800 images with 5 different categories: Abstract artistic, Abstract non artistic, Representative Artistic, representative non artistic and photographs. The results show a correlation of Rs=0.74 with the length of the GIF file for the Figurative Decorative category. Other categories had lower correlations (Abstract Decorative: Rs=0.6, Natural Pictures: Rs=0.55, Figurative Artistic: Rs=0.47 and Abstract Artistic: Rs=0.42).

Chikhman et al. (2012) test different measures of complexity. With a dataset of 15 Chinese hieroglyphs, they found that the best measure is the 'product of squared spatial-frequency median and the image areas'. For a set of 24 outline images of objects, they found that the best measure is the number of turns in the image. Their conclusion is that different complexity estimates are needed for different types of images.

Marin and Leder (2013) also analyse the correlation between computer-generated measures and perceptual image complexity. They use a subset of the International Affective Picture System (IAPS) (Lang, 2005), which contains a collection of images labelled with degrees of affective states that are expressed through those pictures. The correlation between length of the TIFF (Rs=0,53) and JPEG (rs=0,52) was higher than the one achieved using perimeter detection (Rs=0,44). The highest correlation found in this experiment was the RMS contrast, with a correlation of Rs=0.59. In a second experiment, done with a set of paintings, the correlation achieved was lower.

The differences in findings between Forsythe et al. (2011) and Marin and Leder (2013) could be explained by the datasets employed. The dataset in Forsythe et al. (2011) contained images with highly differing complexity in five different categories. On the other hand, the two datasets of Marin and Leder (2013) present less variation in complexity: the IASP dataset contains non professional photographs designed for exploring different emotions and the dataset of paintings offers a very similar degree of complexity.

Using the same datasets as Marin and Leder (2013), Marin et al. (2016) analyse the effect of presentation time on perceptual complexity of images. Seventy women classified 96 images from IAPS dataset, presented each for 1, 5 and 25 seconds. The correlations between the objective measures and the subjective ones get higher with the longer exposition time. As before, the experiment with paintings was less conclusive.

Cavalcante et al. (2014) propose the use of a combination of statistics of local contrast and spatial frequency as a measure of complexity. Their dataset contains 74 streetscape images from four cities, 40 daytime and 34 nighttime scenes. They compare the results of this metric with some of the state of the art ones, including perimeter and JPEG complexity, finding that their proposed metric is the more robust regarding different time scenarios.

Jakesch and Leder (2015) tested the role of ambiguity in human complexity perception. To do this they employed artworks with high degree of ambiguity, and modifications of artworks with a low level of ambiguity. While both sets present similar results regarding computer measures (Jpeg, GIF and perimeter detection), the perceptual complexity was different between the two sets. Humans considered those images with higher ambiguity to be more complex than the low ambiguity images.

Ciocca et al. (2015) analyse the role of colour in complexity. They found that subjective scores for colour images present a high correlation to those of greyscale images, suggesting that colour is not related to perception of complexity. They use a range of image features but do not find any one capable of predicting image complexity.

Marin et al. (2016) analyse the differences between three alternate ways to asses the 'hedonic tone' of an image: beauty, pleasantness or liking. They used two datasets, one with 96 representational paintings and other with 96 attractive environmental scenes converted into cartoons. The correlation between the three hedonic tone measures was higher in cartoons (Rs=0.85) than on paintings (Rs=0.73). With the dataset of paintings, correlation of complexity and beauty was Rs=0.26, with a 'pleasantness' Rs=-0.16 and not present for liking. In the cartoons dataset, correlation between complexity and the three hedonic tone measures was not found.

Friedenberg and Liby (2016) analysed the correlation between beauty and compression metrics. The datasets contain images that are patterns of different density created for the experiment. They reported high correlations between beauty and GIF complexity (0.56) and contour length (0.47). They found no correlations between beauty and numbers of parts. Building on this work and using the same datasets, Gauvrit et al. (2017) analysed the correlation between subjective beauty and several different complexity measures: density, number of blocks, GIF compression rate, edge length, entropy and algorithmic complexity. They found that the participants tend to have a preference for some types of complexity, but not for all. That can explain partially the differences between reported results related to image complexity. The authors propose that researchers should specify which

notion of complexity is behind each work.

Forsythe et al. (2017) evaluated human scores for beauty, complexity, familiarity, and encounter. The authors calculated two automatic measures of complexity based on GIF and JPEG compression. The results show a high correlation between automatic measures and human perception of complexity (Rs=.78 for GIF compression). The better predictor for human beauty was GIF complexity. The authors state: 'The data reported here suggests GIF complexity contributed in a small way to perceptions of beauty, but that beauty has no significant relationship with human judgements of visual complexity or familiarity with an image'. The authors consider computer measures more reliable and valid than human collected perceptions of complexity.

Following this line of research, Madan et al. (2018) found that emotional arousal and valence influence image complexity ratings. They found a correlation between arousal and visual complexity of Rs=.50, which was attenuated with bias-aware instructions to Rs=.40. Also, Forsythe et al. (2008) found that familiarity and learning also influences image complexity ratings.

Visual Complexity prediction

In this section we analyse several works that employ a set of metrics and a machine learning system to predict the visual complexity of images. Most of the systems are created by AI researchers but some are created by PA and CA researchers together, with one published in a psychology journal.

Machado et al. (2015) is the first attempt to create an automatic predictor of image complexity based on a combination of metrics. The dataset employed is the one used in Forsythe et al. (2011), consisting of 800 images in 5 different categories. In the first experiment the individual correlation is calculated between a large set of computer generated measures and the average perceptual image complexity. Higher correlation was obtained using a canny edge filter, with Rs=0.77. JPEG compression achieved a correlation of Rs=0.74. In the second experiment, the large set of measures was fed into a machine learning system based on Artificial Neural Networks (ANNs), which form a predictor of complexity. The correlation between the best predictor and the subjective image complexity was Rs=0.83. Edge density and JPEG compression error were the strongest predictors of human complexity rates. The predictor error was 0.09 (0.4 in a scale 1-5). The error was higher on 'Representational Artistic' and 'Photographs of Natural and Man-made Scenes' images, possibly due to more semantic meaning than Abstract (Artistic and Non-artistic) and Representational Non-artistic image categories.

Ciocca et al. (2016) used genetic programming to build an image complexity predictor, using four measures: roughness, number of regions, chroma variance, and memorability. They reported a correlation of Rs=0.890 on the training set, 0.728 on the validation set, and 0.724 on the test set, outperforming the results of each of the measures individually.

Gartus and Leder (2017), calculate a wide range of computational measures of complexity and combine them using a random forest (a standard machine learning technique) to predict image complexity. The images were a set of abstract patterns from the set used by Gartus and Leder (2017), with different numbers of triangles on a white background. The dataset contains 152 asymmetric and 76 symmetric patters for five types of symmetry. They found several computer metrics to have positive correlation with complexity. One metric based on GIF compression had the highest correlation with Rs=0.634 and mirror symmetry having a negative correlation of Rs=-0.578. Combining the metric based on GIF and mirror symmetry together, they reported a correlation of Rs=0.903.

Measuring visual concepts

In this section we focus on different visual concepts that relate to visual aesthetics and how they can be modelled using metrics. We begin with Fractal Dimension, then on principles of symmetry, colour gradient and low level processing.

The first work we are aware of to relate fractal dimension and aesthetics is that of Aks and Sprott (1996), who analysed the correlation between aesthetic preferences and (i) fractal dimension, and, (ii) Lyapunov exponent of abstract patterns. They found a preference for values of fractal dimension and Lyapunov exponents that are typical in natural objects.

Taylor et al. (1999) analysed the fractal dimension of paintings by the artist Jackson Pollock. Later, Taylor et al. (2002) demonstrated that the fractal dimension of Pollock's paintings increased almost linearly for a decade. From that moment the fractal dimension was considered a measure related to the image complexity and was employed on both psychological studies of aesthetics and artificial intelligence applied to aesthetics.

Spehar et al. (2003) found a consistent aesthetic preference for fractal images. They employed forced-choice method of paired comparison, and used images with different fractal dimension. They use three different datasets: (i) natural images, (ii) simulated coastlines and (iii) Pollock's images. The results showed a 'consistent trend for aesthetic preference to peak within the fractal dimension range 1.3–1.5 for the three different origins of fractal image'. The authors consider this range as typical for natural objects.

Taylor et al. (2011) analysed different responses to fractal patterns (visual preferences to physiological responses) in the work of painter Jackson Pollock. Jones-Smith and Mathur (2006), however, question the use of fractal dimension in the work of Pollock.

Street et al. (2016) presents a large scale analysis of aesthetic preferences involving fractal and complexity metrics. The dataset used was composed of 81 abstract monochrome fractal images. After calculating a series of complexity measures, they

found a strong negative correlation between Fractal dimension (FD) and GIF ratio complexity measure, $Rs = -0.93$. They also used two alternative forced choice analysis (TAFC) and obtained demographic information (age, gender, and continent of residence) from each participant. The results suggest strong differences related to continent and gender: in these experiments, females consistently preferred complex images over males.

In Spehar et al. (2016), the authors use a set of 27 synthetic fractal images: nine $1/f$ filtered greyscale images with spectral slopes ranging from 0.5 to 2.5 in increments of 0.25, their thresholded black and white images and edges only counterparts. In a second experiment, they employed two further variations of the filtered greyscale images, called 'mountain' (that simulate a binary view of a mountain) and 'terrain' (that simulates a satellite view of a field with altitude shown in greyscale). They found that the majority of participants exhibited a peak preference for the intermediate fractal-scaling characteristics while other participants exhibited either a linear increase (aprox 20%) in preference with increasing amplitude spectrum slope, or a linear decrease in preference with increasing amplitude spectrum slope (aprox 20%). The different tendencies were highly stable across all image types.

In his PhD Thesis, Patuano (2018) applied fractal dimension to landscapes. In order to do that, he employ several pre-processing stages to create a binary version of the image (using edges, silhouette outline, etc.) and then applied the box-counting method. The measure with highest correlation to human preference was the fractal dimension of the image's extracted edges.

In Viengkham and Spehar (2018) a set of images of tree levels of fractal dimension (low, medium and high) are present to a group of people, who are asked to rate liking, pleasantness, complexity, and interestingness. The study includes three types of synthetic fractal images and seven types of paintings. In most of the categories, a majority of participants prefer images with intermediate fractal dimension, with (40.13% compared with 33.05% of low fractal dimension and 26.82% of high FD).

Zipf (1949) proposed that many phenomena follow a distribution where the frequency of occurrence is inversely proportional to its rank in the frequency table. So, the largest city of a country has double the population of the second one, three times more than the third one and so on. Zipf's distribution is usual in language, but it can also describe city population sizes in a country, the number of people watching TV channels, and so on. Manaris and colleagues employ this distribution as metrics for music in several works, e.g. (Manaris et al., 2003, 2005). Machado et al. (2015) obtain a correlation with visual perceptual complexity of Rs=0.64.

The histogram of oriented gradients (HOG) counts occurrences of gradient orientation in localised portions of an image. The Pyramid Histogram of Orientation Gradients (PHOG) contains the HOG of the image with HOGs of parts of the image. Redies et al. (2012) propose two metrics based on PHOG: self-similarity and complexity. They calculated the metrics for different datasets and found that one of those datasets (containing images of art paintings) could be characterised by a specific combination of values of these metrics.

Lyssenko et al. (2016) found a correlation between subjective visual complexity and (i) PHOG Self-Similarity ($Rs = 0.56$) and HOG Complexity ($Rs = 0.682$). The dataset consisted of 79 abstracts artworks. They also found correlations between these metrics and subjective terms that participants use to describe the artworks.

There are some studies that have tried to establish relationships between aesthetic value and colour gamut. Nascimento et al. (2017) analyse the effects of changing the colour gamut of paintings to increase their aesthetic value. They asked a group of users to change the colour gamut of ten paintings. The maximum of the distribution was the same as the original, suggesting, unsurprisingly, that the chromatic compositions of the paintings employed matched the viewers' preferences.

Other works have investigated the relation between aesthetics and symmetry. Weichselbaum et al. (2018) tested symmetry preferences of participants over different levels of individual art expertise. They found that 'with higher art expertise, the ratings for the beauty of asymmetrical patterns significantly increased, but, again, participants preferred symmetrical over asymmetrical patterns'. Thömmes and Hübner (2018) analysed the relation between Instagram 'likes' and three computational measures: two measures of visual balance and the preference for curvature over angularity. They utilised 700 architectural photographs from Instagram accounts. They found a positive correlation between visual balance and likes in 3D photographs, and a negative correlation in 2D ones. To the best of our knowledge it is the first work that employees 'likes' as a measure of aesthetic appeal.

Psychological testing related to Aesthetics

There are a number of psychological tests related to aesthetic judgement. These tests are relatively objective, easy to reproduce and provide quantified results (Nadal, 2007; Attneave, 1957; Rump, 1968; Hall, 1969; Chipman, 1977; Chipman and Mendelson, 1979; Hogeboom and van Leeuwen, 1997; Strother and Kubovy, 2003). The main problem with these tests however, is the lack of consensus about them. The validity of concepts behind each test are debatable, typically being based on aesthetic principles proposed by the author of the test, but not accepted universally. The results of individual tests also vary between different studies, maybe due to selection of participants and other exogenous factors. As an example, Weichselbaum et al. (2018) show that artistic experience affects symmetry preferences.

Graves (1948) developed the *Design Judgement Test*. This test is based on theories of artistic creation and appreciation (Graves, 1951). The author claims that this test can estimate certain capabilities related to artistic and aesthetic evaluation. To

do this, the test estimates the degree of reaction to specific principles of aesthetics (according to the author) such us: unity, drive, predominance, variety, balance, continuity, symmetry, proportion and rhythm. Such principals may not be universally accepted or applicable (Eysenck, 1969; Eysenck and Castle, 1971; Uduehi, 1995). A test consists of ninety pages. Each page contents two or three similar designs. One of the designs obeys all the commented principles while the remaining ones break at least one of them. The task of the individual doing the test is to select those designs that do not break any of the principles.

The average results obtained by participants in this test vary between studies (Eysenck and Castle, 1971; Uduehi, 1995). Although this can be, at least partially, explained by the selection of participants and other exogenous factors, it makes it hard to understand what constitutes a good score in this test. In the test done by Graves, art students get a higher average score than students who did not study art Graves (1948). Graves concludes that the test can be used to differentiate between those two groups. Eysenck and Castle (1971) obtain very different results, showing only minor differences between artistic and no-artistic students (64,4% vs. 60%), and differences between males and females. Eysench explain that the different results regarding art students can be related to changes in artistic education that in 1971 promote more regularity and simplicity than in 1948. Götz and Götz (1974) report that '22 different arts experts (designers, painters, sculptors) had 0.92 agreement on choice of preferred design, albeit being critical of them' (Chamorro-Premuzic and Furnham, 2004).

Machado and Cardoso (1998) propose an aesthetic measure based on processing complexity and image complexity: 'images that are simultaneously visually complex and easy to process are the images that have higher aesthetic value'. Fractals are a easy example of very complex images but easy to process due to the self-similarity. Using metrics based on compression described below, and a fixed equation for aesthetics they obtain scores up to 66 (corresponding to a 73.3% success rate), which is larger than those obtained with fine art graduates. In Machado and Cardoso (2002), the authors employ a similar equation as fitness for a genetic programming engine that creates images. Romero et al. (2012) employs some metrics related to JPEG, fractal compression and Zipf's law and an ANN-based machine learning system to predict the answer of the test, resulting in an accuracy of 74.49%, similar to the previous study.

Hayn-Leichsenring et al. (2017) studied a relationship between objective image measures and the subjective evaluations of the JenAesthetics dataset. This dataset consists of 1628 high quality images of paintings (`http://www.inf-cv.uni-jena.de/en/jenaesthetics`). The objective measures are low level image statistics related to aesthetics in previous research, such as those from Braun et al. (2013), related to selfishness, anisotropy and complexity. The subjective evaluations were aesthetic (defined as artistic value) and beauty (defined as individual attachment). The results revealed that the paintings of each period present specific statistical properties of the images. Moreover, they show evidence of correlation between beauty and aesthetics, and correlation between aesthetics and some objective measures on different subsets. The highest correlation was found between self-similarity and the beauty of a subset of paintings of buildings with $Rs = 0.50$. They found differences between aesthetic and beauty scores.

Conclusions

The research from PA and CA have several main differences. First, we analyse some differences regarding the datasets and the results. The datasets used in the PA experiments usually contain a small number of images, due to the need to evaluate each of them by a group of human participants. In CA the ideal is to have a large dataset of images that can allow machine learning to have more complete and diverse information.

Some datasets used in CA work are based on website photographic collections that have a large number of contributed images (Ke et al., 2006; Datta et al., 2006, 2008). However, the images in these datasets were evaluated online in an uncontrolled environment and may have potential biases depending on relations with the author of the image, popularity reinforcement, display environment, and so on. Finally, as the information (images and evaluations) was provided from photographic websites, it is not clear what the users are evaluating (photograph quality, originality, visual aesthetic, liking). An interesting approach is to employ a game to obtain evaluations of images. That allows the researcher to provide clear choices for evaluation and may encourage participants to spend more time contributing to the research. In Hacker and von Ahn (2009) the authors employ a two player game where each participant should evaluate images following the taste of the other participant. They recruited thousands of players and have collected millions of judgements.

From PA research we learn that, even in PA experiments done in controlled conditions, users provide substantially different evaluations depending on the term and context of the question (Marin et al., 2016; Hayn-Leichsenring et al., 2017). Hence we propose that when using such website datasets it is necessary to experimentally test what users are evaluating. And if possible, the best way to proceed is to create datasets in collaboration with PA researchers, with evaluations done in controlled environments and with a number of images that support the use of machine learning techniques.

Many of the curated datasets of art images are restricted to Western or European art, raising issues of cultural bias. Likewise many of the reported studies are undertaken in Europe or North America, which may impact on diversity of study participants. With increasing scrutiny on how AI datasets are obtained for machine learning applications, researchers need to be aware of implicit or explicit bias in their selection of training data. This is an on-going issue for research in this field.

Finally, some recent PA research comments on the main differences between individuals in appreciation of visual aesthetics and complexity (Güçlütürk et al., 2016). A more detailed analysis of this issue could be very interesting. Moreover, it can be interesting to create a large set of images with individual evaluations of human beings, allowing the training of computer system to evolve to the aesthetic preferences of one individual human.

Regarding the results, CA research typically reports results using a success rate or RMS error, while psychologists are more likely to use correlation. This is not a major problem: some papers get good results in correlation employing ML systems that try to minimise RMS error (Machado et al., 2015), but future systems trained to maximise correlation can achieve better results.

Closer collaboration between PA and CA can give rise to results that advance both disciplines. Given the general quality of datasets (PA), enormous sets of computer metrics (CA) and ML techniques (CA), and posterior analysis (both), more powerful predictors of visual complexity can be built. AI researchers can even use AI methods to build new metrics that no one was thinking about (using GP programming such as in Ciocca et al. (2016), Artificial Neural Networks (Carballal et al., 2018), or deep learning (Lemarchand, 2018)). Predictors can be implemented that allow remote access (via a web page, for example), allowing any researcher to get a visual complexity value for a image or set of images on-line. This will help make the analysis of aesthetics and complexity objective metrics more accurate than individual feature analysis, and accessible for everyone. In this context, it is remarkable that the work of Forsythe et al. (2017) finds more correlation between aesthetics with objective complexity measures (GIF compression metric) than with subjective complexity. It could be interesting to undertake a similar analysis with a complexity predictor made by a combination of metrics. Additionally, the detailed analysis of the predictor then allows us to know more information about the relevant metrics related to complexity. Obviously, the definition of some standards: one dataset, one complexity predictor, etc. that is acceptable for everyone will help in this schema of common research.

Some CA researchers begin the research with the idea of creating a computer system able to create original and aesthetically valuable artworks (McCormack, 2017). Generative techniques such us genetic programming are very interesting for this research because they can be used to illustrate the results of a metric or combination of them (den Heijer and Eiben, 2010; Machado et al., 2008). However, better research results in complexity and aesthetic prediction are needed in order to advance image generation systems. Here a collaboration between PA and CA is needed in order to achieve the new research results required. Moreover, even without being used in conjunction with generative systems, computer aesthetics systems can have enormous real world applications.

Evolutionary art and computational aesthetics are relatively young areas of research. Yet, some authors may think that there is nothing new that can be done. Our purpose with this paper is to help the development of both areas by highlighting some possible under-explored pathways and to illustrate the exciting and valuable prior research from the psychology of aesthetics.

We hope for a future where several visual complexity and aesthetic predictors are accessible online, where evolutionary art tools are widely employ by people as ways of exploring their creative capacity, and where computer systems can convincingly create paintings in the style of any human artist and beyond.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

Data Availability

No data was used to support this study.

References

PP Aitken. Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology*, 103(2):240, 1974.

Deborah J Aks and Julien C Sprott. Quantifying aesthetic preference for chaotic patterns. *Empirical studies of the arts*, 14(1): 1–16, 1996.

F-Xavier Alario and Ludovic Ferrand. A set of 400 pictures standardized for french: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, 31(3):531–552, 1999.

Rudolf Arnheim. *Art and Visual Perception, a Psychology of the Creative Eye*. Faber and Faber, London, 1956.

Rudolf Arnheim. *Towards a Psychology of Art/Entropy and Art — An Essay on Disorder and Order*. The Regents of the University of California, 1966.

Rudolf Arnheim. *Visual Thinking*. University of California Press, Berkeley, CA, 1969.

Fred Attneave. Physical determinants of the judged complexity of shapes. *Journal of experimental Psychology*, 53(4):221, 1957.

John Gilbert Beebe-Center and Carroll Cornelius Pratt. A test of Birkhoff's aesthetic measure. *The Journal of General Psychology*, 17(2):339–353, 1937.

Daniel E Berlyne. Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 17(3):274, 1963.

Daniel E Berlyne. Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5):279–286, 1970.

Daniel E Berlyne. *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Hemisphere, 1974.

Daniel E Berlyne, JC Ogilvie, and LC Parham. The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 22(5):376, 1968.

Garrett Birkhoff. *Aesthetic Measure*. Harvard University Press, 1933.

Guy Birkin. *Aesthetic Complexity: Practice and Perception in Art & Design*. PhD thesis, Nottingham Trent University, 2010.

Patrick Bonin, Ronald Peereman, Nathalie Malardier, Alain Méot, and Marylène Chalard. A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, 35(1): 158–167, 2003.

Julia Braun, Seyed Ali Amirshahi, Joachim Denzler, and Christoph Redies. Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, 4:808, 2013.

Gilbert Brighouse. Variability in preferences for simple forms. *Psychological Monographs*, 51(5):68, 1939.

Adrian Carballal, Antonino Santos, Juan Romero, Penousal Machado, João Correia, and Luz Castro. Distinguishing paintings from photographs by complexity estimates. *Neural Computing and Applications*, 30(6):1957–1969, Sep 2018. ISSN 1433-3058. doi: 10.1007/s00521-016-2787-5. URL https://doi.org/10.1007/s00521-016-2787-5.

Noël Carroll. *Philosophy of Art: A Contemporary Introduction*. Routledge, 1999.

André Cavalcante, Ahmed Mansouri, Lemya Kacha, Allan Kardec Barros, Yoshinori Takeuchi, Naoji Matsumoto, and Noboru Ohnishi. Measuring streetscape complexity based on the statistics of local contrast and spatial frequency. *PLoS One*, 9(2): e87097, 2014.

Tomas Chamorro-Premuzic and Adrian Furnham. Art judgement: A measure related to both personality and intelligence? *Imagination, Cognition and Personality*, 24:3–25, 2004.

Anjan Chatterjee. *The Aesthetic Brain*. Oxford University Press, 2013.

Valeriy Chikhman, Valeriya Bondarko, Marina Danilova, Anna Goluzina, and Yuri Shelepin. Complexity of images: Experimental and computational estimates compared. *Perception*, 41(6):631–647, 2012.

Susan F Chipman. Complexity and structure in visual patterns. *Journal of Experimental Psychology: General*, 106(3):269, 1977.

Susan F Chipman and Morton J Mendelson. Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, 5(2):365, 1979.

Gianluigi Ciocca, Silvia Corchs, Francesca Gasparini, Emanuela Bricolo, and Riccardo Tebano. Does color influence image complexity perception? In *International Workshop on Computational Color Imaging*, pages 139–148. Springer, 2015.

Gianluigi Ciocca, Silvia Corchs, and Francesca Gasparini. Genetic programming approach to evaluate complexity of texture images. *Journal of Electronic Imaging*, 25(6):061408, 2016.

Gerald C. Cupchik and Daniel E. Berlyne. The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology*, 20(1):93–104, 1979. doi: 10.1111/j.1467-9450.1979.tb00688.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9450.1979.tb00688.x.

Yael M Cycowicz, David Friedman, Mairav Rothstein, and Joan Gay Snodgrass. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of experimental child psychology*, 65(2):171–237, 1997.

Arthur Danto. The artworld. *Journal of Philosophy*, 64(19):571–584, 1964.

Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision – ECCV 2006, 9th European Conference on Computer Vision, Part III*, LNCS, pages 288–301, Graz, Austria, 2006. Springer.

Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40:5:1–5:60, May 2008. doi: http://doi.acm.org/10.1145/1348246.1348248. URL http://doi.acm.org/10.1145/1348246.1348248.

Roland Clark Davis. An evaluation and test of Birkhoff's aesthetic measure formula. *The Journal of General Psychology*, 15(2):231–240, 1936.

E. den Heijer and A. E. Eiben. Comparing aesthetic measures for evolutionary art. In Cecilia Di Chio, Anthony Brabazon, Gianni A. Di Caro, Marc Ebner, Muddassar Farooq, Andreas Fink, Jörn Grahl, Gary Greenfield, Penousal Machado, Michael O'Neill, Ernesto Tarantino, and Neil Urquhart, editors, *Applications of Evolutionary Computation*, pages 311–320, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-12242-2.

Don C Donderi. Visual complexity: a review. *Psychological bulletin*, 132(1):73, 2006.

Don C Donderi and Sharon McFadden. Compressed file length predicts search time and errors on visual displays. *Displays*, 26(2):71–78, 2005.

Don C Donderi, PWGSC Contract No, and Sharon McFadden. *A complexity measure for electronic displays: Final report on the experiments*. Department of National Defence, Defence Research & Development Canada-Toronto, 2003.

Denis Dutton. Aesthetic universals. In Gaut and McIver Lopes (2013), pages 267–278.

H. J. Eysenck and M. Castle. Comparative study of artists and nonartists on the maitland graves design judgment test. *Journal of Applied Psychology*, 55(4):389–392, 1971.

Hans Jürgen Eysenck. The empirical determination of an aesthetic formula. *Psychological Review*, 48(1):83, 1941a.

Hans Jurgen Eysenck. 'type'-factors in aesthetic judgements. *British Journal of Psychology. General Section*, 31(3):262–270, 1941b.

H.J. Eysenck. *The Experimental Study of the Good Gestalt: A New Approach*. Lancaster Press, 1942. URL https://books.google.es/books?id=MhHaNwAACAAJ.

H.J. Eysenck. Factor analytic study of the maitland graves design judgement test. *Perceptual and Motor Skills*, 24:13–14, 1969.

G.T. Fechner. *Zur experimentalen Aesthetik*. Number v. 1 in Abhandlungen der Mathematisch-physischen classe der Königl. sächsischen gesellschaft der wissenschaften. IX. bd. S. Hirzel, 1871. URL https://books.google.es/books?id=AbzhQAAACAAJ.

Gustav Theodor Fechner. *Vorschule der aesthetik*, volume 1. Breitkopf & Härtel, 1876.

Gerhard Fischer. User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction*, 11:65–86, 2001.

Alex Forsythe, Noel Sheehy, and Martin Sawey. Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers*, 35(2):334–342, 2003.

Alex Forsythe, Gerry Mulhern, and Martin Sawey. Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, 40(1):116–129, 2008.

Alex Forsythe, Marcos Nadal, Noel Sheehy, Camilo J Cela-Conde, and Martin Sawey. Predicting beauty: fractal dimension and visual complexity in art. *British journal of psychology*, 102(1):49–70, 2011.

Alex Forsythe, Nichola Street, and Mai Helmy. Revisiting rossion and pourtois with new ratings for automated complexity, familiarity, beauty, and encounter. *Behavior research methods*, 49(4):1484–1493, 2017.

Robert Frances. *Psicología del arte y la estética*, volume 3. Ediciones AKAL, 1985.

Jay Friedenberg and Bruce Liby. Perceived beauty of random texture patterns: A preference for complexity. *Acta psychologica*, 168:41–49, 2016.

Philip Galanter. Computational aesthetic evaluation: Steps towards machine creativity. In *ACM SIGGRAPH 2012 Courses*, SIGGRAPH '12, pages 14:1–14:162, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1678-1. doi: 10.1145/2343483. 2343497. URL http://doi.acm.org/10.1145/2343483.2343497.

Mariano García, Albert N Badre, and John T Stasko. Development and validation of icons varying in their abstractness. *Interacting with computers*, 6(2):191–211, 1994.

Andreas Gartus and Helmut Leder. Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception. *PloS one*, 12(11):e0185276, 2017.

Berys Gaut and Dominic McIver Lopes, editors. *The Routledge Companion to Aesthetics*, 2013. Routledge.

Nicolas Gauvrit, Fernando Soler-Toscano, and Alessandro Guida. A preference for some types of complexity comment on "perceived beauty of random texture patterns: A preference for complexity". *Acta psychologica*, 174:48–53, 2017.

K. Götz and K. Götz. The maitland graves design judgement test judged by 22. *Perceptual and Motor Skills*, 39:261–262, 1974.

Laura K. M. Graf and Jan R. Landwehr. A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19(4):395–410, 2015. doi: 10.1177/1088868315574978. URL https://doi.org/10.1177/1088868315574978. PMID: 25742990.

Maitland Graves. *Design Judgement Test, Manual*. The Psychological Corporation, New York, 1948.

Maitland Graves. *The Art of Color and Design*. McGraw-Hill, New York, 1951.

Gary Greenfield. On the origins of the term "computational aesthetics". In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'05, pages 9–12, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association. ISBN 3-905673-27-4. doi: 10.2312/COMPAESTH/ COMPAESTH05/009-012. URL http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH05/009-012.

Yağmur Güçlütürk, Richard HAH Jacobs, and Rob van Lier. Liking versus complexity: decomposing the inverted u-curve. *Frontiers in human neuroscience*, 10:112, 2016.

Severin Hacker and Luis von Ahn. Matchin: Eliciting user preferences with an online game. pages 1207–1216, 10 2009. doi: 10.1145/1518701.1518882.

AC Hall. Measures of the complexity of random black and white and coloured stimuli. *Perceptual and motor skills*, 29(3): 773–774, 1969.

Gregor U Hayn-Leichsenring, Thomas Lehmann, and Christoph Redies. Subjective ratings of beauty and aesthetics: correlations with statistical image properties in western oil paintings. *i-Perception*, 8(3):2041669517715474, 2017.

Tom Heath, Sandy G Smith, and Bill Lim. Tall buildings and the urban skyline: The effect of visual complexity on preferences. *Environment and behavior*, 32(4):541–556, 2000.

Julian Hochberg and Virginia Brooks. The psychophysics of form: Reversible-perspective drawings of spatial objects. *The American Journal of Psychology*, 73(3):337–354, 1960. ISSN 00029556. URL http://www.jstor.org/stable/1420172.

Florian Hoenig. Defining computational aesthetics. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'05, pages 13–18, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association. ISBN 3-905673-27-4. doi: 10.2312/COMPAESTH/COMPAESTH05/013-018. URL http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH05/013-018.

Menno Hogeboom and Cees van Leeuwen. Visual search strategy and perceptual organization covary with individual preference and structural complexity. *Acta Psychologica*, 95(2):141–164, 1997.

Shintchi Ichikawa. Quantitative and structural factors in the judgment of pattern complexity. *Perception & psychophysics*, 38 (2):101–109, 1985.

Martina Jakesch and Helmut Leder. The qualitative side of complexity: Testing effects of ambiguity on complexity judgments. *Psychology of Aesthetics, Creativity, and the Arts*, 9(3):200, 2015.

Colin G. Johnson. Fitness in evolutionary art and music: a taxonomy and future prospects. *International Journal of Arts and Technology*, 9(1), 2016.

Katherine Jones-Smith and Harsh Mathur. Fractal analysis: Revisiting pollock's drip paintings. *Nature*, 444(7119):E9, 2006.

Bruce F Katz. What makes a polygon pleasing? *Empirical studies of the Arts*, 20(1):1–19, 2002.

Yan Ke, Xiaoou Tang, and Feng Jing. The Design of High-Level Features for Photo Quality Assessment. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:419–426, 2006.

Krzysztof Krawiec, Jerry Swan, and Una-May O'Reilly. Behavioral program synthesis: Insights and prospects. In R. Riolo, B. Worzel, M. Kotanchek, and A. Kordon, editors, *Genetic Programming Theory and Practice XIII*, pages 169–183. Springer, 2016.

Elizabeth Krupinski and Paul Locher. Skin conductance and aesthetic evaluative responses to nonrepresentational works of art varying in symmetry. *Bulletin of the Psychonomic Society*, 26(4):355–358, 1988.

A. Krzeczkowska, J. El-Hage, S. Colton, and S. Clark. Automated collage generation—with intent. In Dan Ventura et al., editors, *Proceedings of the International Conference on Computational Creativity*, pages 36–40, 2010.

Peter J Lang. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical report*, 2005.

H. Leder, B. Belke, A. Oeberst, and D. Augustin. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95:489–508, 2004.

Helmut Leder and Marcos Nadal. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode – developments and challenges in empirical aesthetics. *British Journal of Psychology*, 105:443–464, 2014.

EL Leeuwenberg. Quantitative specification of information in sequential patterns. *Psychological Review*, 76(2):216, 1969.

François Lemarchand. Fundamental visual features for aesthetic classification of photographs across datasets. *Pattern Recognition Letters*, 2018.

Matthew Lewis. Evolutionary visual art and design. In Juan Romero and Penousal Machado, editors, *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*, pages 3–37, 2008.

Nathalie Lyssenko, Christoph Redies, and Gregor U Hayn-Leichsenring. Evaluating abstract art: Relation between term usage, subjective ratings, image properties and personality traits. *Frontiers in psychology*, 7:973, 2016.

P. Machado and A. Cardoso. All the truth about NEvAr. *Applied Intelligence, Special Issue on Creative Systems*, 16(2): 101–119, 2002.

Penousal Machado and Amílcar Cardoso. Computing aethetics. In *Proceedings of the 14th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence*, SBIA '98, pages 219–228, London, UK, UK, 1998. Springer-Verlag. ISBN 3-540-65190-X. URL http://dl.acm.org/citation.cfm?id=645851.669151.

Penousal Machado, Juan Romero, María Santos, Amílcar Cardoso, and Bill Manaris. Adaptive critics for evolutionary artists. In Günther Raidl et al., editors, *Applications of Evolutionary Computing*, volume 3005 of *Lecture Notes in Computer Science*, pages 437–446. Springer Berlin / Heidelberg, 2004.

Penousal Machado, Juan Romero, and Bill Manaris. Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In Juan Romero and Penousal Machado, editors, *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*, pages 381–415. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-72877-1_18.

Penousal Machado, Juan Romero, Marcos Nadal, Antonino Santos, João Correia, and Adrián Carballal. Computerized measures of visual complexity. *Acta psychologica*, 160:43–57, 2015.

Christopher R Madan, Janine Bayer, Matthias Gamer, Tina B Lonsdorf, and Tobias Sommer. Visual complexity and affect: ratings reflect more than meets the eye. *Frontiers in psychology*, 8:2368, 2018.

Birgit Mallon, Christoph Redies, and Gregor Uwe Hayn-Leichsenring. Beauty in abstract paintings: perceptual contrast and statistical properties. *Frontiers in human neuroscience*, 8:161, 2014.

Bill Manaris, Juan Romero, Penousal Machado, Dwight Krehbiel, Timothy Hirzel, Walter Pharr, and Robert Davis. Zipf's law, music classification and aesthetics. *Computer Music Journal*, 29(1):55–69, 2005.

Bill Z. Manaris, Dallas Vaughan, Christopher Wagner, Juan Romero, and Robert B. Davis. Evolutionary music and the zipf-mandelbrot law: Developing fitness functions for pleasant music. In Günther R. Raidl, Jean-Arcady Meyer, Martin Middendorf, Stefano Cagnoni, Juan J. Romero Cardalda, David Corne, Jens Gottlieb, Agnès Guillot, Emma Hart, Colin G. Johnson, and Elena Marchiori, editors, *Applications of Evolutionary Computing, EvoWorkshop 2003: EvoBIO, EvoCOP, EvoIASP, EvoMUSART, EvoROB, and EvoSTIM, Essex, UK, April 14-16, 2003, Proceedings*, volume 2611 of *Lecture Notes in Computer Science*, pages 522–534. Springer, 2003. ISBN 3-540-00976-0. doi: 10.1007/3-540-36605-9_48. URL https://doi.org/10.1007/3-540-36605-9_48.

Manuela M Marin and Helmut Leder. Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS One*, 8(8):e72412, 2013.

Manuela M Marin, Allegra Lampatz, Michaela Wandl, and Helmut Leder. Berlyne revisited: evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers in human neuroscience*, 10:536, 2016.

Jon McCormack. Creative ecosystems. In Jon McCormack and Mark d'Inverno, editors, *Computers and Creativity*, pages 39–60. Springer, 2012.

Jon McCormack. Working with generative systems: an artistic perspective. In Jonathan Bowen, Nick Lambert, and Graham Diprose, editors, *Electronic Visualisation and the Arts (EVA 2017)*, Electronic Workshops in Computing (eWiC), pages 213–218. BCS Learning and Development Ldt., London, 10th-13th July 2017. doi: http://dx.doi.org/10.14236/ewic/EVA2017.47.

Siné JP Mcdougall, Martin B Curry, and Oscar de Bruijn. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, 31(3):487–519, 1999.

Norman Charles Meier. *Art in human affairs; an introduction to the psychology of art.* McGraw-Hill, 1942.

Sandra M Messinger. Pleasure and complexity: Berlyne revisited. *The Journal of Psychology*, 132(5):558–560, 1998.

Abraham Moles. *ThÈorie de L'Information et Perception EsthÈtique.* Denoel, 1958.

Marcos Nadal. *Complexity and Aesthetic preference for diverse visual stimuli.* PhD thesis, Departament de Psicologia, Universitat de les Illes Balears, 2007.

Marcos Nadal, Enric Munar, Gisèle Marty, and Camilo José Cela-Conde. Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2):173–191, 2010.

Sérgio MC Nascimento, João MM Linhares, Cristina Montagner, Catarina AR João, Kinjiro Amano, Catarina Alfaro, and Ana Bailão. The colors of paintings and viewers' preferences. *Vision research*, 130:76–84, 2017.

Richard M Nicki, Poh Lin Lee, and Virginia Moss. Ambiguity, cubist works of art, and preference. *Acta Psychologica*, 49(1): 27–41, 1981.

RM Nicki and Virginia Moss. Preference for non-representational art as a function of various measures of complexity. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 29(3):237, 1975.

John W Osborne and Frank H Farley. The relationship between aesthetic preference and visual complexity in absract art. *Psychonomic Science*, 19(2):69–70, 1970.

Agnes Patuano. *Fractal dimensions of landscape images as predictors of landscape preference*. PhD thesis, 2018.

Rolf Reber, Norbert Schwarz, and Piotr Winkielman. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, 8(4):364–382, 2004.

Christoph Redies, Seyed Ali Amirshahi, Michael Koch, and Joachim Denzler. Phog-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In *European Conference on Computer Vision*, pages 522–531. Springer, 2012.

Graeme Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17:67–99, 2007.

Juan Romero, Penousal Machado, Antonio Santos, and Amilcar Cardoso. On the development of critics in evolutionary computation artists. In Stefano Cagnoni et al., editors, *Applications of Evolutionary Computing: Evoworkshops 200'3*, volume 2611 of *Lecture Notes in Computer Science*, pages 559–569. Springer, 2003.

Juan Romero, Penousal Machado, Adrian Carballal, and João Correia. *Computing Aesthetics with Image Judgement Systems*, pages 295–322. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-31727-9. doi: 10.1007/978-3-642-31727-9_11. URL https://doi.org/10.1007/978-3-642-31727-9_11.

EE Rump. Is there a general factor of preference for complexity? *Perception & Psychophysics*, 3(5):346–348, 1968.

DH Saklofske. Visual aesthetic complexity, attractiveness and diversive exploration. *Perceptual and Motor Skills*, 41(3): 813–814, 1975.

David Salomon. *Data Compression: The Complete Reference*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 1846286026.

Juergen Schmidhuber. Facial beauty and fractal geometry, June 1998. URL http://cogprints.org/690/.

Herbert A Simon. Complexity and the representation of patterned sequences of symbols. *Psychological review*, 79(5):369, 1972.

Branka Spehar, Colin WG Clifford, Ben R Newell, and Richard P Taylor. Universal aesthetic of fractals. *Computers & Graphics*, 27(5):813–820, 2003.

Branka Spehar, Nicholas Walker, and Richard P Taylor. Taxonomy of individual variations in aesthetic responses to fractal patterns. *Frontiers in human neuroscience*, 10:350, 2016.

Arthur E Stamps III. Entropy, visual diversity, and preference. *The Journal of general psychology*, 129(3):300–320, 2002.

Nichola Street, Alexandra M Forsythe, Ronan Reilly, Richard Taylor, and Mai S Helmy. A complex story: Universal preference vs. individual differences shaping aesthetic response to fractals patterns. *Frontiers in human neuroscience*, 10:213, 2016.

Lars Strother and Michael Kubovy. Perceived complexity and the grouping effect in band patterns. *Acta psychologica*, 114(3): 229–244, 2003.

Hideyuki Takagi. Interactive evolutionary computation: fusion of the capabilities of ec optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.

Richard Taylor, Branka Spehar, Caroline Hagerhall, and Paul Van Donkelaar. Perceptual and physiological responses to jackson pollock's fractals. *Frontiers in human neuroscience*, 5:60, 2011.

Richard P Taylor, Adam P Micolich, and David Jonas. Fractal analysis of pollock's drip paintings. *Nature*, 399(6735):422, 1999.

Richard P Taylor, Adam P Micolich, and David Jonas. The construction of jackson pollock's fractal drip paintings. *Leonardo*, 35(2):203–207, 2002.

Katja Thömmes and Ronald Hübner. Instagram likes for architectural photos can be predicted by quantitative balance measures and curvature. *Frontiers in psychology*, 9:1050, 2018.

Joseph Uduehi. A cross-cultural assessment of the maitland graves design judgment test using u.s. and nigerian students. *Visual Arts Research*, 21(2):11–18, 00 1995.

Catherine Viengkham and Branka Spehar. Preference for fractal-scaling properties across synthetic noise images and artworks. *Frontiers in psychology*, 9, 2018.

Paul C Vitz. Preference for different amounts of visual complexity. *Behavioral Science*, 11(2):105–114, 1966.

CO Weber. The aesthetics of rectangles and theories of affection. *Journal of Applied Psychology*, 15(3):310, 1931.

Hanna Weichselbaum, Helmut Leder, and Ulrich Ansorge. Implicit and explicit evaluation of visual symmetry as a function of art expertise. *i-Perception*, 9(2):2041669518761464, 2018.

George Kingsley Zipf. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, 1949.