

Entropy and Optimality in Abstract Art: An Empirical Test of Visual Aesthetics

Kevin Burns

The MITRE Corporation, 202 Burlington Road, Bedford, MA, USA, kburns@mitre.org

Abstract: How can aesthetic responses to artworks be computed? Previous authors have proposed governing properties, including symmetry and complexity, along with equations for quantifying these properties and combining them into an overall measure of aesthetics. But existing mathematical models have not been well-motivated by psychological theories or well-validated by empirical testing. An alternative model is derived here, using a novel measure of visual entropy to quantify graphic complexity and compute aesthetic optimality. This model is tested against human judgments of complexity and aesthetics using abstract designs composed of horizontal and vertical grid lines. The empirical results support the mathematical model of entropy and optimality, but also highlight difficulties associated with computing aesthetics for other abstract and figurative artworks. Implications of the data and model are discussed with regard to current and future efforts in the field of computational aesthetics, aimed at automating the evaluation of aesthetics and generation of artworks.

Keywords: abstract art, aesthetics, entropy, experiments, probability, psychology, semantics, surprise.

AMS Subject Classifications: 62C10; 62F15; 91E10; 91E30; 94A15; 94A17

1. Introduction

Connections between mathematics and the arts can be drawn in two major directions. On the one hand, mathematical ideas or algorithms may be used in *synthesizing* artworks; and on the other hand, mathematical models or measures may be used in *analyzing* artworks. Sometimes the two directions are combined, but usually the analysis and/or synthesis is limited to mathematical properties of artworks without addressing the mathematical psychology of aesthetics.

As an example, the first sentence of the first paper of the first issue of *Journal of Mathematics and the Arts* [61] states: "*The creation of engaging mazes requires both mathematical and aesthetic considerations*". Although this statement acknowledges the importance of aesthetics, the article is about mathematical algorithms for maze construction and only one paragraph considers aesthetics, as follows: "*We have also experimented with the aesthetic possibilities that arise in mazes of interconnected vortices... [and] found that even in this very limited, abstract domain, we can still produce mazes that are also attractive drawings*".

This treatment is typical of papers published in *Journal of Mathematics and the Arts*, where evaluations of aesthetics are usually authors' opinions and not supported by mathematical modeling or empirical testing. Several exceptions appear in the special issue on *Mathematical Models Used in Aesthetic Evaluation* [24], although a psychological perspective is missing from most of those papers. This absence is important for reasons reported in the first article of *Journal of Mathematics and the Arts* [61], as follows: "*Both complexity and aesthetics rely not only on measurable geometric properties..., but also on poorly understood facts about human perception and psychology*".

The present paper advances an understanding of complexity and aesthetics, through mathematical modeling and empirical testing of psychological judgments. The approach includes: reviewing limitations of previous efforts to compute aesthetic quality as a function of visual complexity (Section 1); deriving a novel measure of visual entropy to quantify graphic complexity (Section 2); proposing a formal model of aesthetic optimality as a function of visual entropy (Section 3); obtaining empirical data to test the validity of the model (Section 4); and creating new artworks based on the model and data (Section 5).

Throughout the present paper, all designs are limited to an abstract domain of horizontal and vertical grid lines. This provides a theoretical focus and empirical control needed for rigorous analyses and experiments. The results of these analyses and experiments are then used to synthesize new designs in the same genre – similar to works by well-known artists such as Piet Mondrian. In an often-quoted letter to a fellow Dutch painter, Mondrian wrote of his work in this genre as follows: *"I believe it is possible that, through horizontal and vertical lines constructed with awareness, but not with calculation, led by high intuition, and brought to harmony and rhythm, these basic forms of beauty, supplemented if necessary by other direct lines or curves, can become a work of art, as strong as it is true."*

Although it is common for artists (and others) to contrast *"calculation"* and *"intuition"*, intuition is a psychological phenomenon and hence can be modeled by calculation – much like any other psychological judgment or decision studied in the field of mathematical psychology. Here the judgments of interest are judgments of aesthetics, made by artists and their audiences; and the decisions of interest are decisions made by artists, about where to place the lines in their grid-line compositions. The present study of these judgments and decisions is aimed at advancing an understanding of artistic intuition and audience appreciation, via mathematical calculation and empirical validation.

To further inform the approach, the remainder of this introduction reviews existing theories of aesthetics from philosophical, mathematical, and psychological perspectives. Special attention is paid to mathematical models of psychological processes that underlie aesthetics, as motivation for the model derived and discussed in subsequent sections.

1.1. Unity and variety

Arguably the most universal theory of art and aesthetics is that of *"unity in variety"* [6,29], expressed by Hutchesen [28] as follows: *"the figures which excite in us the ideas of beauty seem to be those in which there is uniformity amidst variety"*. Similarly Fry [21] wrote of artworks: *"the first quality that we demand in our sensations will be order, without which our sensations will be troubled and perplexed, and the other quality will be variety, without which they will not be fully stimulated"*.

Related theories have been expressed throughout history by many authors including Plato, Aristotle, Plotinus, Kant, Hegel, Dewey, and Heidegger [26]. Berlyne [3] reviews a number of these philosophies, along with scientific theories from psychology and physiology such as Fechner's [19] *"principle of unitary connection of the manifold"*. Berlyne and his colleagues also performed many empirical studies to advance a science of aesthetics [4], but ultimately his experiments and two-factor (reward and aversion) theory of *"arousal potential"* [3] did not go much beyond earlier philosophies in specifying how aesthetic responses can be computed.

Before Berlyne, Birkhoff [5] proposed a quantitative theory of aesthetic measure M as the ratio of perceived order O to complexity C , writing: "*the well known aesthetic demand for 'unity in variety' is evidently closely connected to this formula*". Although Birkhoff offered "*A Mathematical Argument*" as a basis for the formula $M = O / C$, his argument contains a number of critical errors outlined here in *Appendix A*. Also Birkhoff never tested his equation against human judgments of aesthetics, so his theory is lacking both a theoretical foundation and empirical validation. Nevertheless, the formula $M = O / C$ has historically been remarkably influential and remains so even in current efforts to compute aesthetics [2,3,15,17,18,24,32,36,39,42,46,49,50,51,57].

Over the years, various researchers have tested $M = O / C$ against human judgments of aesthetics, and on balance the equation has not been supported; for reviews see [17,18,32,39]. Instead a large number of empirical studies reviewed by Eysenck [17,18] suggests a better model would compute M as a product, $M = O * C$, rather than a quotient, $M = O / C$. However all these studies are sensitive to how the two independent variables (O and C) are computed by experimenters from features of stimuli, and Birkhoff himself proposed only ad hoc methods for doing so. For example, Birkhoff measured the complexity C of a polygon by the number of line segments in its outline, without any regard for the relative lengths, orientations, or relations between the line segments. Similarly, he measured order O by counting various symmetries (vertical, rotational, and so forth), which were all weighted equally, and he assumed an arbitrary list of "*unsatisfactory*" features that reduced the magnitude of O . Therefore it is not clear how the theory can be applied and tested on graphic designs other than the isolated polygons that Birkhoff himself analyzed.

1.2. Information theory

After Birkhoff and around the time of Berlyne, a number of authors attempted to apply Shannon's mathematical theory of information and communication [53] to art and aesthetics; for a review see [42]. These authors, led by Bense [2] and Moles [41], used a quantity known as *entropy* to measure complexity C . Following Shannon's formulation, a message (artwork) can be characterized as a set of discrete signals which occur with some probability (frequency), and entropy measures the overall *unpredictability* of signals in the message. The entropy is lowest when one signal is very probable and all other signals are very improbable. The entropy is highest when no signal is more probable than any other signal, and entropy increases as the number of equally probable signals increases.

To quantify the entropy of a visual stimulus, one must first define the categories of signals in a repertoire that will be used to compute probabilities. Typically this is done by treating each individual pixel as a signal [49,50,51]. For example, in each panel (a, b, c, d, e) of Figure 1 there are only two categories of pixels (black, white) and hence two probabilities (P_{black} , P_{white}). Thus by Shannon's definition [53], the entropy (H) of a panel (a, b, c, d, or e) is computed as follows:

$$H = -\sum_i P_i * \log P_i,$$

where i is an index over the two pixel categories (black, white), and the log is taken to base 2 such that H measures entropy in "bits". Numerically, the entropy H in panels a, b, c, d, and e of Figure 1 increases from a to e, because the probability of black pixels P_{black} is increasing toward that of white pixels P_{white} . Visually, inspection of Figure 1 suggests that the perceived *complexity* is also increasing from a to e.

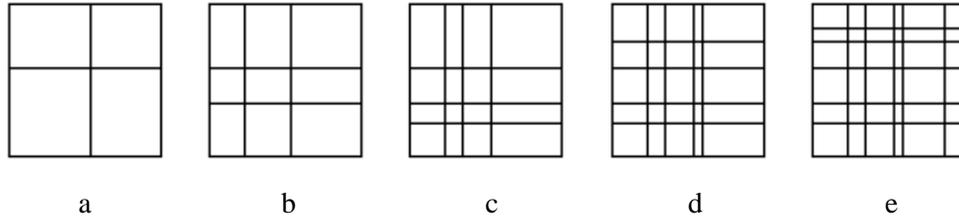


Figure 1. H increases: (a) $H = 0.19$; (b) $H = 0.24$; (c) $H = 0.28$; (d) $H = 0.33$; (e) $H = 0.36$. V_E increases: (a) $V_E = 2.0$; (b) $V_E = 3.3$; (c) $V_E = 4.3$; (d) $V_E = 5.0$; (e) $V_E = 5.4$. See text for definitions of H and V_E .

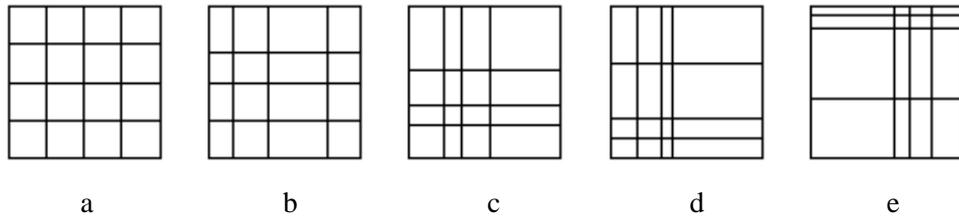


Figure 2. $H = 0.28$ is constant. V_E increases: (a) $V_E = 4.0$; (b) $V_E = 4.1$; (c) $V_E = 4.3$; (d) $V_E = 4.6$; (e) $V_E = 4.9$. See text for definitions of H and V_E .

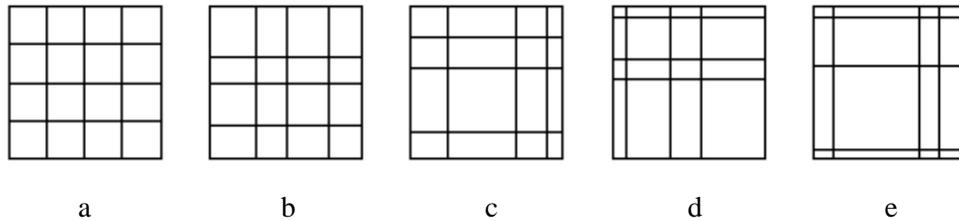


Figure 3. $H = 0.28$ is constant. V_E increases: (a) $V_E = 4.0$; (b) $V_E = 4.1$; (c) $V_E = 4.3$; (d) $V_E = 4.6$; (e) $V_E = 4.9$. See text for definitions of H and V_E .

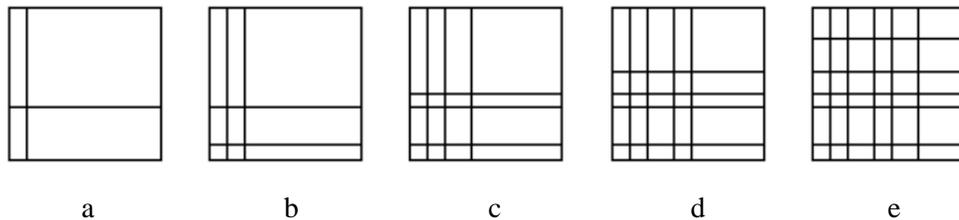


Figure 4. H increases: (a) $H = 0.19$; (b) $H = 0.24$; (c) $H = 0.28$; (d) $H = 0.33$; (e) $H = 0.36$. V_E increases: (a) $V_E = 2.7$; (b) $V_E = 4.2$; (c) $V_E = 4.8$; (d) $V_E = 5.1$; (e) $V_E = 5.4$. See text for definitions of H and V_E .

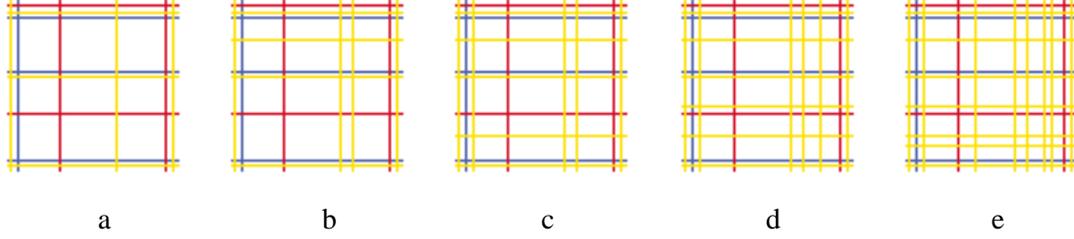


Figure 5. H increases: (a) $H = 0.47$; (b) $H = 0.52$; (c) $H = 0.56$; (d) $H = 0.60$; (e) $H = 0.66$. V_E increases: (a) $V_E = 6.5$; (b) $V_E = 6.7$; (c) $V_E = 7.0$; (d) $V_E = 7.3$; (e) $V_E = 7.6$. See text for definitions of H and V_E .

Mathematically, the quantity $-\log P_i$ is a measure of *improbability* > 0 (because $0 < P_i < 1$), sometimes called self-information [12,53]. The same quantity is also called *surprisal* [12,53], because when $-\log P_i$ is large (P_i is small) a signal of class i is surprising when it is seen, and when $-\log P_i$ is small (P_i is large) a signal of class i is not surprising when it is seen. Thus H can be characterized as a sum of surprisals, each weighed by the frequency (P_i) at which it occurs in the stimulus.

In applying Shannon's theory to information aesthetics, authors have typically adopted Birkhoff's equation $M = O / C$ and then used entropy H to quantify complexity C and order O . For example, a recent version [49,50] of Bense's formulation [2,51] assumes complexity is given by H_{max} and order is given by $H_{max} - H$ such that:

$$M = O / C = (H_{max} - H) / H_{max} = 1 - H / H_{max},$$

where H_{max} is the entropy computed when each signal category has an equal probability. For example, in the case of black and white pixels $H_{max} = 1$ when $P_{black} = P_{white} = 0.5$. In another equation [42] attributed to contemporaries of Bense, complexity is given by H and order is given by $(H_{max} - H) / H_{max}$ such that:

$$M = O / C = (H_{max} - H) / (H * H_{max}) = 1 / H - 1 / H_{max}.$$

Although these two equations differ by a factor of H , they both predict the same qualitative behavior for M as a function of H when the signal repertoire and hence H_{max} is held constant (such as $H_{max} = 1$ for the case of black and white pixels). That is, both equations imply M increases monotonically as H decreases, such that the aesthetic measure M would be maximized when entropy H is minimized. As noted in [51]: "*information aesthetics... inherits Birkhoff's preference for minimalist structures*". In fact the first model discussed above predicts maximal M at minimal H even if Birkhoff's equation $M = O / C$ is replaced by Eysenck's equation, because in that case we have: $M = O * C = (H_{max} - H) * H_{max}$.

If the above models are correct, then aesthetics should decrease monotonically from panel a to e in Figure 1 (and Figure 4), because entropy H increases from panel a to e. Also the aesthetic measure should be constant across all panels of Figure 2 (and Figure 3), because entropy H is constant across the panels. Later, in Section 4 *Validation*, these predictions will be tested formally in an experiment that includes Figures 1-4 as stimuli. But at this point even informal inspection of Figures 1-4 casts doubt on the predictions of information aesthetics. In fact inspection of Figures 2 and 3, where entropy H is constant, suggests that classical entropy H is not even an adequate measure of visual complexity itself. This will also be tested formally in Section 4 *Validation*.

1.3. A cognitive hierarchy

The obvious problem with classical entropy computed from pixels is that it only quantifies an overall frequency for each pixel category, such as P_{black} and P_{white} . By ignoring where the black and white pixels actually appear in a figure, relative to other black and white pixels, H fails to capture more meaningful aspects of similarity, proximity, and continuity along black lines and within white shapes at a higher level of abstraction.

As a simple example, consider the statement "rectangle with area A " in referring to a region of Figure 1a. This is a statement of *meaning*, connoting the shape and size for a subset of *signals* (which are white pixels bounded by black pixels) that appear in the figure. Cognitively this connotation stems from perceptual grouping processes, which aggregate lower-level signals like individual black and white pixels into higher-level meanings like lines and shapes, based on Gestalt principles [55] of similarity, proximity, and continuity.

In fact human vision actually produces a hierarchy of such feature abstractions [27,34,37], which can be characterized as follows: meanings at one (lower) level serve as signals to be assigned meanings at the next (higher) level; and then meanings at that level serve as signals to be assigned further meanings at the next level; and so forth; until meanings are assigned at some highest level as well as all lower levels. For example, a simple abstraction hierarchy with only two levels can be characterized as follows for the grid-line designs in Figures 1-4: At the lower level, signals are black pixels and white pixels, and meanings are white shapes bounded by black lines. At the higher level, signals are white shapes bounded by black lines, and meanings are connotative features of the overall composition that evoke *feelings* such as balance (associated with unity, symmetry, regularity, similarity) and tension (associated with variety, asymmetry, irregularity, diversity). In that respect, our affective responses to these graphic designs depend on cognitive appraisals [16,32,44,45,48,51,52,54] of *meanings* for *signals* at the higher level.

Of course one might argue that classical entropy H can instead be computed over some different set of signal categories, besides black and white pixels. In fact Moles [41] and Bense [2] suggested this in a notion of "*superization*". For example, consider Figure 2a and assume that the repertoire contains "*super-signals*" defined as white shapes and black lines, rather than signals defined as white pixels and black pixels. In that case we might count 16 white shapes and 10 black lines, such that $P_{white\ shapes} = 0.62$ and $P_{black\ lines} = 0.38$. However, these probabilities and hence entropy H would still be the same for all panels of Figure 2 (and Figure 3), so even with superization the classical entropy calculation predicts no change in visual complexity or aesthetic quality across the panels of Figure 2 (and Figure 3).

These examples illustrate two limitations of information aesthetics, which stem from classical entropy (H) and Birkhoff's equation ($M = O / C$). First, information aesthetics are limited by the assumption that classical entropy H is an adequate model of visual complexity (and order). Second, information aesthetics are limited by the assumption that Birkhoff's equation $M = O / C$ is an adequate model of how aesthetics vary as a function of complexity and order. These two limitations are addressed in the remaining sections of this paper as follows: Section 2 *Formulation* derives a novel measure of visual entropy to model graphic complexity; Section 3 *Optimization* develops a formal model of how aesthetic quality varies with visual entropy; and Section 4 *Validation* tests these mathematical models (as well as H and $M = O / C$) against psychological judgments of visual complexity and aesthetic quality.

2. Formulation

2.1. Fractional entropy

Assuming the repertoire of signals in Figures 1-4 includes only white shapes and black lines, a straightforward extension of Shannon's equation to spatial information would be to compute an entropy of areas [58]. That is, probabilities representing fractional white areas A_n (each > 0 and ≤ 1) bounded by black lines would replace white and black pixel frequencies P_i in the previous equation for H . The resulting entropy of areas, here denoted U , is as follows:

$$U = -\sum_n A_n * \log A_n,$$

where n is an index ranging from 1 to N over the N white shapes (bounded by black lines).

Like H , U tends to increase as the number of signal categories increases, which in this case would be the number N of discrete areas A_n . But for a given number N , U would be maximized when all N areas have the same A_n , such that variety among the white areas is minimized (see panel a of Figures 2 and 3). Therefore U is actually a measure of numerosity and *unity*, rather than a measure of numerosity and *variety* as needed in a model of complexity.

2.2. Visual entropy

Unlike the fractional entropy U , a *visual entropy* V_E should decrease as the variety among N shapes decreases, such that V_E is minimized (not maximized) when a set of N shapes exhibits area uniformity $A_1 = A_2 = A_3 = A_4 \dots = A_N$. Also a visual entropy V_E should decrease as the number N of areas decreases. To obtain such a quantity, we can define V_E as the sum of a *relative entropy* V_R (measuring only variety) and a *reference entropy* V_0 (measuring only numerosity), where V_R is measured relative to V_0 . An expression for this quantity $V_E = V_R + V_0$ is derived here as follows:

First, consider an area probability distribution $\{A_n\} = \{A_1, A_2, \dots, A_N\}$ along with a *reference* probability distribution of uniform areas (no variety) given by $\{A_{ave}\} = \{A_{ave}, A_{ave}, \dots, A_{ave}\}$ where $A_{ave} = 1 / N$. The entropy of distribution $\{A_n\}$ *relative* to the reference distribution $\{A_{ave}\}$ is given by an information-theoretic quantity known as the Kullback-Leibler divergence [33]:

$$V_R = -\sum_n A_{ave} * \log (A_n / A_{ave}).$$

By a law of logs, this equation is expanded to obtain:

$$\begin{aligned} V_R &= [-\sum_n A_{ave} * \log A_n] - [-\sum_n A_{ave} * \log A_{ave}] \\ &= (1 / N) * \sum_n -\log A_n - V_0, \end{aligned}$$

where V_0 is the reference entropy computed as follows:

$$V_0 = -\sum_n A_{ave} * \log A_{ave} = \log N.$$

Finally, we can write:

$$V_E = V_R + V_0 = (1 / N) * \sum_n -\log A_n.$$

Notice that this visual entropy V_E differs from the fractional entropy U , because each surprisal ($-\log A_n$) in the sum is multiplied by a probability $1/N$ rather than by a probability A_n . This follows from the choice of white shapes (bounded by black lines) as signals in the calculation of V_E , because at this higher level of abstraction we have N white shapes each appearing once in the image.

As specific examples, values of V_E are computed and reported in captions of Figures 1-5 for all five panels (labeled a, b, c, d, e) appearing in each figure. For Figures 1, 4, and 5, V_E increases with increasing numerosity of white shapes. For Figures 2 and 3, V_E increases with increasing variety among the white shapes. By comparison, classical entropy H is constant across panels of Figures 2 and 3 (and fractional entropy U actually decreases as the variety increases across panels of Figures 2 and 3). These two models, V_E and H , are tested against human judgments of visual complexity in Section 4 *Validation*.

2.3. Psychological complexity

From a psychological perspective, one might wonder why human vision would be more surprised by small white areas (where small A_n yields large $-\log A_n$) than large white areas (where large A_n yields small $-\log A_n$), as implied by the formulation of visual entropy V_E . A basis for this behavior [56] can be found in the function of human vision, which has evolved to exploit *meaningful changes*. That is, our visual system is designed to detect changes in local image intensity (*signals*) that represent changes in global object properties (*meanings*), such as surface orientation along the edge of one object, or depth of position from one object to another.

These changes are represented as short "lines" by neurons in early stages of the visual system, and then further aggregated to represent the continuous edges of objects at higher levels of the visual hierarchy [27,34,37]. In fact this neural coding helps explain how humans, even those who are blind from birth [31], can so readily interpret 2-D line drawings of 3-D shapes and scenes [47]. In effect, our vision of a line drawing assumes that each line depicts a change – which implies a surprise (discontinuity) across the line and lack of surprise (continuity) along the line. By that logic, there would be more surprise when one line is close to another line than when the two lines are far apart, so small white areas A_n (bounded by close lines) would represent more surprise than large white areas A_n .

Of course there is an upper limit on surprise measured by $-\log A_n$, based on the lower limit of resolution by which human vision can discriminate between lines. Also, when multiple lines are in close proximity they may be perceived at a higher level of abstraction as representing a region of texture or shading [30], rather than as individual edges or depth changes. And even when lines are relatively sparse, a viewer's surprise upon seeing one spacing between lines may depend on expectations formed after seeing other spacings between lines. For example, after seeing two parallel lines in close proximity a viewer may be less (or more) surprised to see two more parallel lines in close proximity. Thus the surprise experienced by a viewer actually depends on more than just the proximity of one line to another line. But to a first approximation, based on the design of our visual system, it is reasonable to assume that the surprise associated with seeing an area A_n would increase as the distances between the lines that bound A_n are decreased – hence as A_n itself decreases.

As described above, the surprise modeled by V_E is a key component of the aesthetic experience and akin to *variety* (or complexity). The other component, akin to *unity* (or order), is addressed in the next section, which presents an integrated model of how complexity and order combine to affect aesthetic optimality.

3. Optimization

Referring to the philosophy of "*unity in variety*", our underlying intuition is that aesthetics arise from a combination of competing properties, namely: *unity* (order, balance, simplicity, predictability) and *variety* (chaos, tension, complexity, unpredictability). This presents a tradeoff because these two properties are logical opposites, so the only way to get more of one is by giving up some of the other. In this section I develop a mathematical model of the tradeoff, dubbed EVE' [7-10], in which aesthetics are modeled by a psychological progression of expectations (E), violations (V) of expectations, and explanations (E') of violations.

3.1. EVE's equations

According to EVE', the aesthetic experience is a combination of two competing effects that each provide a form of *pleasure*. One effect is successful expectation (E) of *signals*, where pleasure occurs because one's expectation is satisfied as the expected signals are received. The other effect is successful explanation (E') of *meaning*, where pleasure-prime results from the resolution of tension that was caused by a violation (V) of expectation. Thus the total aesthetic pleasure, denoted X , is computed as follows:

$$X = g * E + g' * E',$$

where g is the gain in pleasure from a unit of successful expectation E , and g' is the gain in pleasure-prime from a unit of successful explanation E' .

The central component of EVE' is a violation (V) of expectation (E), which is measured by the surprisal associated with receiving some signal or set of signals, s . For the visual stimuli in Figures 1-5, this surprisal would be given by the visual entropy V_E derived in Section 2. Logically a violation (V) is the opposite of expectation (E), so E is computed by the additive inverse of V as follows:

$$E = -V = -V_E.$$

Finally, a measure of explanation (E') for a violation (V) is computed as a fraction $P(m|s)$ of V :

$$E' = P(m|s) * V = P(m|s) * V_E,$$

where $P(m|s)$ is a probability that reflects the extent to which the violation is resolved by an explanation of *meaning* m for the *signals* s . Substituting these equations into the earlier equation for X yields:

$$X = [g * (c * P(m|s) - 1)] * V_E,$$

where $c = g' / g$. Thus the equation for X can be seen as a *product* of two factors, much like the aesthetic measure $M = O * C$ suggested by Eysenck [17,18], where the first factor (in brackets) is analogous to order O and the second factor (V_E) is analogous to complexity C . This equation for X is clearly different from Birkhoff's ratio given by $M = O / C$, and the two factors appearing in EVE's equation for X are also different from previous information-theoretic models of order and complexity. In particular, the second factor is computed as a visual entropy V_E that differs from classical entropy H , and the first factor is computed using a conditional probability of *meaning* for *signals*, $P(m|s)$.

No quantity like the conditional probability $P(m|s)$ appears in previous models of information aesthetics, and in fact no quantity like $P(m|s)$ even appears in Shannon's information theory itself. As discussed in Weaver's introduction [59] to Shannon's famous paper [53], information theory addresses communication only at the level of probabilistic *signals*, without any regard for the semantic *meanings* that those signals may carry from the creator of a message to a consumer of the message. EVE' differs in modeling the appraisal of meaning, via conditional probability $P(m|s)$, as well as arousal from signals, via visual entropy V_E .

Besides addressing *meanings* as well as *signals*, EVE' is also novel in addressing differences between individuals in an audience. In particular, the parameter $c = g' / g$ models how much more pleasure-prime is gained from a unit of E' compared to the pleasure gained from a unit of E . Any individual will have $c > 1$, or else there would be no affective incentive for them to seek surprise and hence experience arousal from artworks. But the magnitude of $c > 1$ may vary among individuals, and psychological research has established that people do differ in a personality trait known as "sensation seeking" [62] that affects aesthetic preferences. For example, people who are high in sensation seeking (modeled here by high c) tend to prefer more complex and abstract art [62] than people who are low in sensation seeking.

Similarly, the magnitude of $P(m|s)$ may vary among individuals because P depends on the meanings (m) that an individual assigns to signals (s) in appraisal of an artwork [16,32,44,45,48,51,52,54]. This is especially true for figurative artworks depicting objects, humans, actions, and other aspects of the world where rich semantics may make meanings quite variable between cultures and between individuals within those cultures. For that reason, the present study uses only abstract designs for which it is feasible to estimate $P(m|s)$ by making a single simplifying assumption. This assumption is that more meaning will be extracted and hence higher $P(m|s)$ will be computed when a viewer is processing signals that are relatively simple and ordered – as opposed to signals that are relatively complex and disordered. Using a linear function bounded by $P(m|s) = 0$ when V_E reaches some maximum value $V_{E,max}$, and $P(m|s) = 1$ when $V_E = 0$, we can write the assumption as follows:

$$P(m|s) = (V_{E,max} - V_E) / V_{E,max}.$$

Substituting this expression for $P(m|s)$ into the previous equation for X , we obtain:

$$X = [g * (c * (V_{E,max} - V_E) / V_{E,max} - 1)] * V_E.$$

3.2. EVE's prediction

Now notice what the above equation for X as a function of V_E implies about an *optimal* aesthetic experience, which is computed by setting the first derivative equal to zero, $dX / dV_E = 0$. The result is:

$$V_{E,opt} = [1/2 * (1 - 1 / c)] * V_{E,max}.$$

According to this equation, aesthetic optimality occurs at an intermediate visual entropy $V_{E,opt}$ between 0 and $V_{E,max}$, and $V_{E,opt}$ increases as the personal preference for complexity modeled by c increases. Clearly this prediction is much different from the prediction of Birkhoff's formulation $M = O / C$ and associated equations of information aesthetics. As noted in Section 1.2, those models predict that aesthetics will increase monotonically as entropy decreases, such that aesthetic optimality will occur at a minimal entropy $H_{opt} = \epsilon$ for all individuals in an audience.

4. Validation

An experiment was performed to test EVE' and competing models against human judgments, by collecting two types of responses using Figures 1-5 as stimuli. First, judgments of *visual complexity* were obtained to test the model of visual entropy (V_E) as well as classical entropy (H). Second, judgments of *aesthetic quality* were obtained to test EVE's model of X versus V_E as well as $M = O / C$.

4.1. Materials and methods

Many artists have created artworks using only horizontal and vertical lines, albeit often filling areas with color and/or using lines that do not run the complete length or width of the image. Perhaps the most well-known is Mondrian, whose artworks include a series of designs similar to Figures 1-4 but painted with colored lines, titled New York City I, II, and III, which can be found by searching images on the world wide web. An approximation of New York City I (1942) was created for this study and appears as panel e of Figure 5. The remaining panels (d, c, b, and a) of Figure 5 were created by progressively removing yellow lines from Mondrian's design, with the simplest panel containing a total number of lines comparable to the most complex panel of Figures 1 and 4.

All other designs in the present study use grids of black lines on a white background, as illustrated in Figures 1-4. The positions of lines in these designs were generated at random by a computer program, subject to a constraint that parallel lines not be closer than 5% of the panel length or width (similar to the designs of Figure 5). The experimental stimuli in Figures 1-4 were chosen from a large set of designs that were generated in this fashion, to obtain numerical ranges of entropies needed to test competing models.

The experimental stimuli included five pages (numbered 1, 2, 3, 4, 5) as illustrated in Figures 1-5. On each page, the five panels (labeled a, b, c, d, e) were arranged in random order horizontally across the page. Each participant was presented with the five pages, one at a time, and asked to provide two sets of responses as follows: first, to rank the five panels on the page from most to least *visually complex*; then, to rank the five panels on the page from most to least *aesthetically pleasing*. The experiment was administered as a survey over the internet, and participants were volunteers who each reported their judgments on all pages of the survey. At the end of the survey, participants were also asked to answer demographic questions about their gender, age group, training (no art training; some art training), art taste (do not like abstract art; do like abstract art), and lifestyle (enjoys surprises; prefers to plan things).

4.2. Empirical data

A total of 148 adults volunteered for the experiment, typically taking about five minutes to complete all five pages of the online survey. These 148 participants were fairly well-balanced with respect to the demographic questions noted above: 80 male and 68 female; 84 younger age group (< 48 years old) and 64 older age group (≥ 48 years old); 94 no art training and 54 some art training; 61 do not like abstract art and 87 do like abstract art; 61 enjoy surprises and 87 prefer to plan things.

For each page of the survey (presenting Figure 1, 2, 3, 4, or 5), participants ranked the panels (a, b, c, d, e) from most to least *visually complex* (hereafter denoted C), and from most to least *aesthetically pleasing* (hereafter denoted A). Integer scores of 1 (lowest rank) to 5 (highest rank) were then assigned to these rankings in order to analyze the judgments quantitatively. As noted above, panels on each page were presented to participants in random order across the page. However, for purposes of presenting the data

here, all results are reported for panels labeled a, b, c, d, and e as illustrated in Figures 1, 2, 3, 4, and 5. For each ranking (C and A) of panels (a, b, c, d, e) on each page (Figure 1, 2, 3, 4, 5), mean judgments across participants were computed for use in testing models against data (see Section 4.4).

The mean rankings are illustrated in bar graphs of Figure 6 (rankings for C) and Figure 7 (rankings for A). The significance of these mean rankings was established by performing Friedman's non-parametric analysis of variance (ANOVA), which is appropriate for data that are rank-ordered. As summarized in Table 1, Friedman's test showed that all mean rankings were significant, with $p < 0.001$ for each ranking (C and A) of each figure. The data for C and A were also used to compute Kendall's coefficient of concordance (W) as a measure of agreement between participants, ranging from zero agreement ($W = 0.0$) to perfect agreement ($W = 1.0$). Table 1 summarizes the results denoted W_C (agreement in rankings of C) and W_A (agreement in rankings of A) for Figures 1-5. The W_C values > 0.50 for Figures 1-5 show that agreement among participants was quite high for all rankings of visual complexity (C). The W_A values < 0.50 for Figures 1-5 show that agreement among participants was lower for all rankings of aesthetic quality (A). These results are to be expected, because judgments of aesthetic quality are presumably a function of visual complexity as well as other factors, so rankings of A would necessarily be more variable than rankings of C .

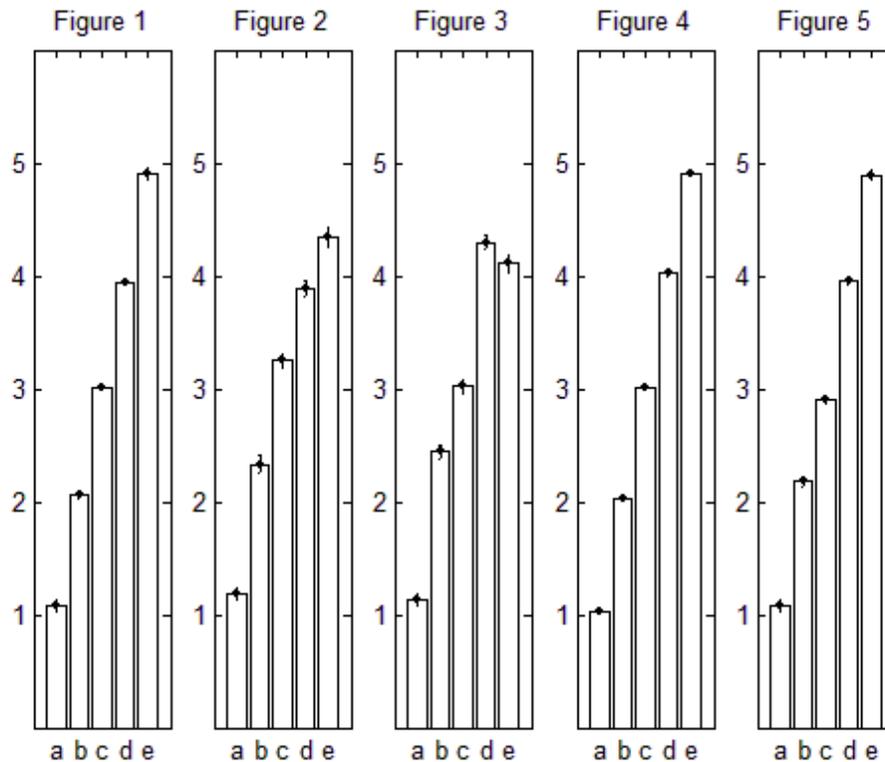


Figure 6. Mean rankings of visual complexity (C) for panels (a, b, c, d, e) of Figures 1, 2, 3, 4, 5. Error bars represent standard errors of the means.

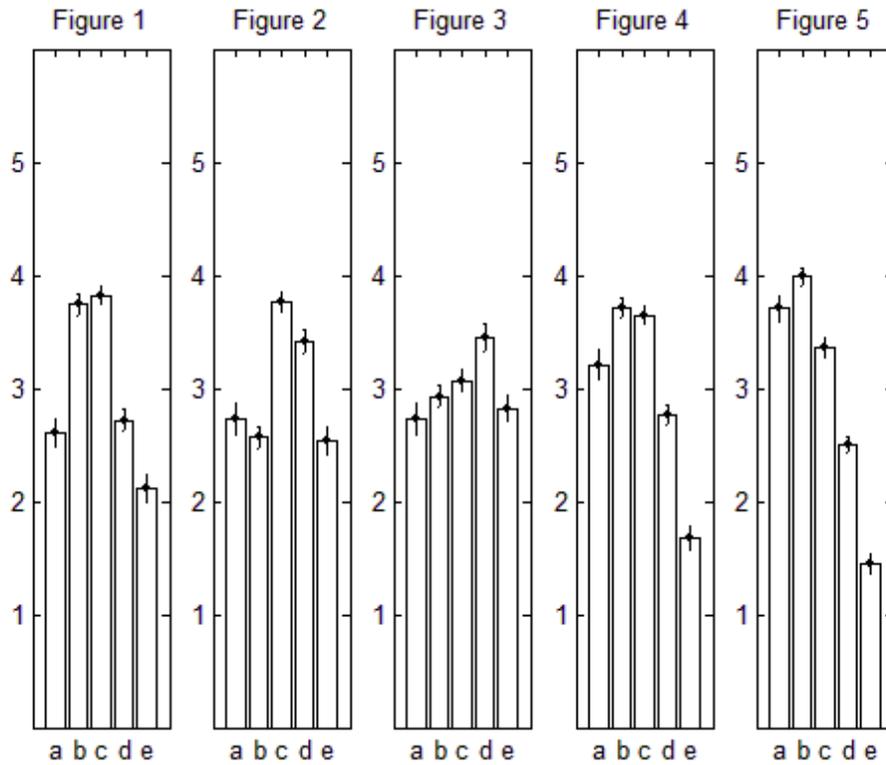


Figure 7. Mean rankings of aesthetic quality (A) for panels (a, b, c, d, e) of Figures 1, 2, 3, 4, 5. Error bars represent standard errors of the means.

Table 1. Levels of significance (p values), from Friedman's non-parametric analysis of variance, for mean rankings of visual complexity (C) and aesthetic quality (A). Values of Kendall's coefficient of concordance (W), for mean rankings of visual complexity (C) and aesthetic quality (A).

	<u>Figure 1</u>	<u>Figure 2</u>	<u>Figure 3</u>	<u>Figure 4</u>	<u>Figure 5</u>
p_C	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
p_A	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
W_C	0.91	0.64	0.67	0.95	0.88
W_A	0.22	0.12	0.03	0.28	0.42

Table 1 shows that W_C is lower for Figures 2 and 3 than for Figures 1 and 4; and W_A is lower for Figures 2 and 3 than for Figures 1 and 4. These results are also to be expected, because the range of visual entropy (V_E) across panels (a, b, c, d, e) was about three times higher for Figures 1 and 4 than for Figures 2 and 3. However, as established by Friedman's test (noted above), mean rankings of C and A were significant ($p < 0.001$) for all figures, so these mean rankings offer useful data for testing models as discussed below.

4.3. Model predictions

In order to compare models to data, mean rankings for C (Figure 6) and A (Figure 7) are first converted to integer values. In cases where the ranks are within error bars representing standard errors of the means, equal non-integer values are assumed. Comparable rankings are also predicted for each model as follows:

For visual complexity, visual entropy (V_E) predicts a ranking 1, 2, 3, 4, 5 (panels a, b, c, d, e) for each page (Figure 1, 2, 3, 4, 5), consistent with the captions of Figures 1-5. For classical entropy (H), the ranking is also 1, 2, 3, 4, 5 for Figures 1, 4, and 5, but the ranking is 3, 3, 3, 3, 3 for Figures 2 and 3, again consistent with the captions of Figures 1-5. Note that values of H reported in the caption of Figure 5 were computed assuming {white, red, yellow, blue} pixels as the signal repertoire, but the same ranking (1, 2, 3, 4, 5) is obtained when H is computed assuming {white, non-white} pixels as the signal repertoire.

For aesthetic quality, EVE' predicts that human rankings will depend on the personal parameter c discussed in Section 3. Individual differences between participants are addressed in Section 4.6, but here the mean prediction for Figure 1 is that optimal aesthetics will occur at a visual entropy $V_{E,opt}$ given by the average V_E across panels (a, b, c, d, e). Because this average value lies between the value of V_E for panel b and panel c, the ranking predicted by EVE' is 2.5, 4.5, 4.5, 2.5, 1 (for a, b, c, d, e, respectively). The same logic produces the same ranking for panels (a, b, c, d, e) of Figure 4, and produces the ranking 1, 2.5, 4.5, 4.5, 2.5 for panels (a, b, c, d, e) of Figures 2 and 3. For Figure 5, the total number of lines in the simplest design (panel a) is comparable to that of the most complex design (panel e) of Figures 1 and 4. Therefore EVE' predicts a ranking 5, 4, 3, 2, 1 for panels (a, b, c, d, e) of Figure 5. Now for the model $M = O / C$, where complexity C is computed from classical entropy H , the prediction is that M increases monotonically as H decreases. Thus the predicted ranking is 5, 4, 3, 2, 1 for Figures 1, 4, and 5, but 3, 3, 3, 3, 3 for Figures 2 and 3 where H is constant across panels. For another model, using V_E instead of H as a measure of complexity, the model $M = O / C$ predicts a ranking 5, 4, 3, 2, 1 for the panels (a, b, c, d, e) of Figures 1, 2, 3, 4, and 5. These two models are denoted O/H and O/V , respectively.

4.4. Model performance

Referring to Table 2, data for judgments of visual complexity C are used to test the model of visual entropy V_E as well as the competing model of classical entropy H . For each model, Spearman's rank correlation coefficient (ρ) is computed as a measure of how well the model rankings (V_E or H) match the human rankings (C). Spearman's ρ can range from -1.0 (representing an inverse correlation) through 0.0 (representing a lack of correlation) to 1.0 (representing a perfect correlation). As seen in Table 2, visual entropy V_E predicts subjective complexity C almost perfectly for each page (Figures 1-5). By comparison, classical entropy H predicts subjective complexity C only for Figures 1, 4, and 5, and fails to predict the significantly non-uniform rankings of C for Figures 2 and 3.

Table 2. Results of comparing models to data, using Spearman's rank correlation coefficient (ρ). Results for ρ_V and ρ_H test models of visual entropy and classical entropy, respectively. Results for $\rho_{EVE'}$, $\rho_{O/H}$, and $\rho_{O/V}$ test models of aesthetic quality.

	<u>Figure 1</u>	<u>Figure 2</u>	<u>Figure 3</u>	<u>Figure 4</u>	<u>Figure 5</u>
ρ_V	1.00	1.00	0.90	1.00	1.00
ρ_H	1.00	0.00	0.00	1.00	1.00
$\rho_{EVE'}$	1.00	0.88	0.80	0.97	0.90
$\rho_{O/H}$	0.53	0.00	0.00	0.67	0.90
$\rho_{O/V}$	0.53	-0.22	-0.16	0.67	0.90

Also referring to Table 2, data for judgments of aesthetic quality A are used to test the model of EVE'. These same data are used to test two models based on $M = O / C$, one denoted O/H that uses H as a measure of complexity, and the other denoted O/V that uses V_E as a measure of complexity. Once again, Spearman's rank correlation coefficient (ρ) between model and data is computed in order to obtain a comparable score for each model.

As seen in Table 2, EVE' predicts the human rankings of aesthetics A for Figures 1-5, all with a highly positive score of $\rho_{EVE'} = 0.80$ or greater. O/H scores much lower for Figures 1 and 4, and has zero correlation with the data for Figures 2 and 3. However, the human data for C (discussed above) clearly showed that H is a poor model of subjective complexity for Figures 2 and 3. This suggests the low score for O/H may be due to H and that better results might be obtained if complexity was modeled by V_E . However, Table 2 shows that O/V produces scores that are even lower and actually negative for Figures 2 and 3. In effect, O/H scored higher than O/V for judgments of A only because H did not predict human judgments of C . This result highlights the importance of empirically measuring the independent variable(s), in this case C , as well as the dependent variable (in this case A) – unlike previous efforts to test Birkhoff's model $M = O / C$.

4.5. Additional influences

For the case of Figure 5, Table 2 shows that EVE' and $M = O / C$ both predict the human ranking of aesthetic quality (A) quite well with $\rho = 0.90$. One reason is that complexity is increasing from panel a to panel e in Figure 5, as measured by data C and as modeled by both H and V_E . The other reason is that the simplest design (panel a) of Figure 5 is comparable to the most complex designs of Figure 1 (panel e) and Figure 4 (panel e), and these designs in turn are more complex than designs with optimal aesthetics (see Figure 7). This may help explain why Birkhoff's model has had some limited success predicting aesthetic judgments in other studies. That is, when all stimuli are more complex than the optimal complexity, then $M = O / C$ will correctly predict that aesthetics increase monotonically as complexity decreases. For example, one could easily design an experiment in which all or most stimuli have greater than the optimal level of complexity. One might then conclude that Birkhoff's model is a good model based on empirical results like those obtained here for Figure 5. But to adequately assess models, stimuli must be carefully designed to test the model predictions in realms where those predictions are theoretically falsifiable.

With regard to testing different models, one might wonder why Figure 5 was even included in the experiment, as all three models make the same predictions. One reason was to ensure that participants were not just, as a matter of convenience or conservatism, always choosing a design with average complexity (among panels a, b, c, d, e) as most aesthetic. If that were the case then the success of EVE's predictions could be considered an artifact of the experiment itself. But results for Figure 5 show that the highest ratings for aesthetics A were clearly skewed toward minimal rather than average complexity. This suggests that participants were indeed making deliberate judgments of aesthetics, rather than just choosing a design of average complexity as most aesthetic based on convenience or conservatism.

Another reason for using Figure 5 in the experiment was to include at least one design (panel e in Figure 5) that approximates an actual work of art by a well-known artist; also see [35,38,43,60]. Previous studies have often used stimuli such as Birkhoff's polygons, which are arguably not compositions of the sort that artists might create as art and audiences might display as art. In that regard it is interesting that the most complex panel of Figure 5, designed to approximate an original Mondrian composition in the number, placement, and color of lines, was ranked the least aesthetic of any design (see Figure 7). Moreover, aesthetic agreement among participants was greatest for Figure 5, as seen from the high value of $W_A = 0.42$ in Table 1. Instead of the Mondrian design, most people preferred much simpler variations produced by removing yellow lines such that the numbers of red, yellow, and blue lines were approximately equal.

Of course panel e of Figure 5 only approximates the actual Mondrian titled New York City I (1942), as the original is much larger and contains subtle modulations in colors, brushstrokes, and other features of the artist's style [11]. Also Mondrian's intent may have been to evoke a tense feeling in his painting, similar to what a person might actually feel in the buzz of New York City. In that case knowledge of the painting's title would likely influence an audience's aesthetic judgment – because the audience would be extracting more *meaning* from the image based on their *knowledge* of its content and the artist's intent. Other empirical studies have shown that knowledge of a painting's title alone can significantly enhance the aesthetic experience [40]. For that reason, only non-descript labels (a, b, c, d, e) were assigned to the abstract designs presented to participants in this study (Figures 1, 2, 3, 4, 5). The point here is to acknowledge that this control of experimental stimuli, needed for rigorous testing of competing theories, may in fact miss a major aspect of art in more natural settings – where semantic titles and artistic intents are of much importance to the aesthetic experience.

4.6. Individual differences

As discussed in Section 3 *Optimization*, EVE' includes a personal parameter c that reflects an individual's desire for "sensation seeking" [62] and affects the level of complexity C at which that individual will experience aesthetic optimality. To gain further insight into this parameter, individual judgments were used to identify the level of complexity C_{opt} (that is, rank of C ranging from 1 to 5) at which each participant reported the maximum aesthetic quality (that is, rank of $A = 5$) for Figures 1-4. The resulting value of C_{opt} for each participant, averaged across Figures 1-4, represents a measure of that participant's aesthetic preference for complexity ranging from $C_{opt} = 1$ (simplest) to $C_{opt} = 5$ (most complex). These values of C_{opt} spanned the complete range from 1 to 5, with a mean $C_{opt} = 2.76$. Only one individual had $C_{opt} = 5$, whereas 11 individuals had $C_{opt} = 1$, and the vast majority of all 148 participants had C_{opt} in the range of 2 to 4. Figure 8 plots a histogram of these C_{opt} values in eight intervals.

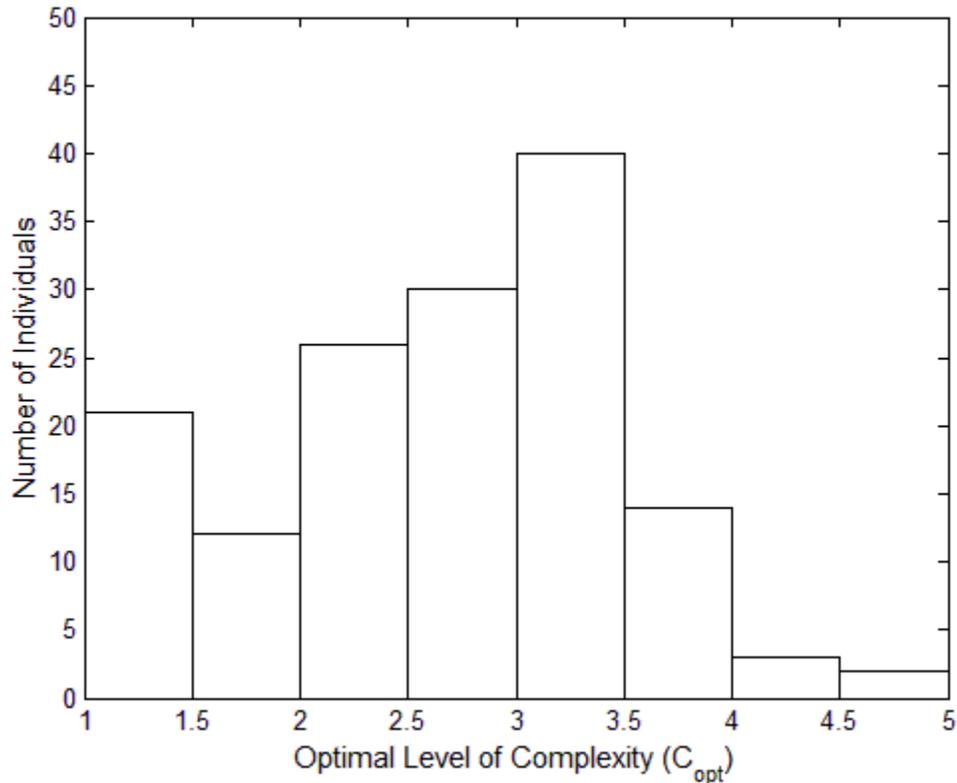


Figure 8. Histogram illustrating the number of individuals (y-axis) with optimal level of complexity (C_{opt}) in each of eight intervals (x-axis).

Further analysis was performed to establish whether any of the personal factors measured by demographic questions could account for these individual differences. To do so, mean values of C_{opt} for each binary distinction (male versus female; younger versus older; no art training versus some art training; do not like abstract art versus do like abstract art; enjoys surprises versus prefers to plan things) were compared using Student's t-test. The results showed that none of the factors was significant (all $p > 0.05$). Gender was the only factor approaching significance ($p = 0.08$), with a mean $C_{opt} = 2.90$ for females higher than the mean $C_{opt} = 2.64$ for males.

Contrary to some speculation [39], even art training did not predict a person's preferred level of visual complexity. Similarly, despite the findings of previous research on sensation seeking [62], neither art taste (such as liking abstract art) nor lifestyle (such as enjoying surprises) predicted a person's optimal level of complexity in the present study. These results demonstrate the importance of individual differences, and highlight the difficulty of predicting any one person's aesthetic preferences.

5. Conclusion

This article offers both theoretical and empirical contributions to the field of computational aesthetics [20,22,23,24,25,51]. The theoretical contribution is a novel formulation of visual entropy, and a mathematical model of aesthetic optimality as a function of visual entropy. The empirical contribution is formal testing of this model, dubbed EVE', including individual differences between members of an audience.

Compared to competing mathematical models of information aesthetics, EVE' is unique in addressing both affective arousal from probabilistic *signals* and cognitive appraisal of semantic *meanings*. But for the present study the semantics were extremely simplified by limiting the visual stimuli to abstract patterns of grid lines. This enabled computation of a visual entropy V_E , needed as input to EVE', and also allowed a consistent comparison between EVE' and existing models of information aesthetics that do not address semantics. The empirical testing demonstrated that V_E and EVE' offer much better models of how humans judge complexity and aesthetics, respectively, compared to classical entropy H and the Birkhoff-based equations of information aesthetics.

Beyond the abstract designs of this study, EVE' also applies to other abstract and figurative artworks in the visual domain [7] – as well as to narrative verse [8,9] and musical works [10]. But in any domain, EVE's mathematical modeling of arousal and appraisal will depend on the psychological processes by which humans perceive *signals* and assign *meanings*. More specifically, arousal depends on surprisal from signals, which in turn depends on perceptual grouping processes that govern which features of an artwork will be sensed as signals in the first place. Similarly, appraisal depends on world knowledge and value structures by which humans infer the meanings of artworks and assess the intents of artists. One insight from EVE' is that no automaton can credibly compute the aesthetics that humans will experience, unless the machine can at least approximately model the perceptual processes and semantic structures of human audiences.

Modeling of these perceptual processes and semantic structures lies largely beyond the current state of the art in artificial intelligence [1], and this limits what can feasibly be accomplished in the field of computational aesthetics. For it is not enough to quantify mathematical properties of artworks themselves, without modeling the psychological processes by which *aesthetics* arise in an *audience* experiencing those artworks. That is why the present study was limited to a narrow domain in which it was feasible to model what human beings would find *surprising* [14] and *meaningful* [13].

Although limited to a narrow domain of abstract designs, the present study does offer general insights into how aesthetics are *evaluated* by a human audience, as well as practical insights into how artworks might be *generated* by an automaton working in the domain. In particular, empirical results (shown in Figure 7) for Figures 1 and 4 suggest that designs in this genre are most aesthetic when constrained to a 3x3 or 4x4 grid of white shapes bounded by black lines. Additional results (shown in Figure 7) for Figures 2 and 3 suggest that most humans prefer the 4x4 designs of Figure 2c and Figure 3d. Referring to the captions of Figure 2c and Figure 3d, this implies an optimal range of $4.3 \leq V_E \leq 4.6$ for 4x4 designs. These results can be used in an algorithm to generate designs that human viewers will likely find most aesthetic, as follows:

1. Define the number of white shapes bounded by black lines to be drawn in a grid as $4 \times 4 = 16$.
2. Select the positions of black lines at random, subject to a requirement that no parallel lines can be closer than 5% of the panel length or width. This constraint is imposed for consistency with the experimental stimuli from which empirical results (used in step 4 below) were obtained.
3. Compute the visual entropy V_E of the resulting design.
4. Accept the design as a candidate artwork only if V_E is in the range $4.3 \leq V_E \leq 4.6$.

This algorithm was used to generate a number of designs that are nearly optimal based on the theoretical model and empirical data. From a set of such candidates, I selected three designs for a triptych to be constructed in wood and exhibited as a wall sculpture. A photograph of the resulting artwork appears in Figure 9.

In composing this triptych from the computer-generated designs, I made slight modifications to two horizontal lines in the middle panel. These lines were nearly at the same vertical positions as horizontal lines in the left and right panels. The adjustment was to make the alignments exact. The visual entropy of the middle panel changed only slightly, from $V_E = 4.37$ to $V_E = 4.42$, hence remained in the range of aesthetic optimality. But continuity along these two horizontal lines adds some unity amidst the variety of three different panels, hence the change is theorized to improve the overall aesthetic of the triptych.

The three panels were also given metaphorical titles [40] and numerical subtitles, which are meaningful to the artist (if not the audience). This is similar to the titles and subtitles *New York City I, II, and III* given to Mondrian's series of colored-line compositions. In the case of Figure 9, the titles are cities in which the artist has lived, and geographic features of those cities are suggested to the artist (if not the audience) by geometric features of the panels.

I also constructed two additional wall sculptures from computer-generated designs. But unlike the natural wood finish of the triptych, all inward-facing surfaces of these sculptures were painted either red, yellow, or blue. This introduces a faint colored glow to each bounded area when the artwork is hung on a white wall, as seen in the photographs of Figure 10. The painted edges also appear as bold colored lines when the artwork is seen from an oblique view, as in the photograph of Figure 11.

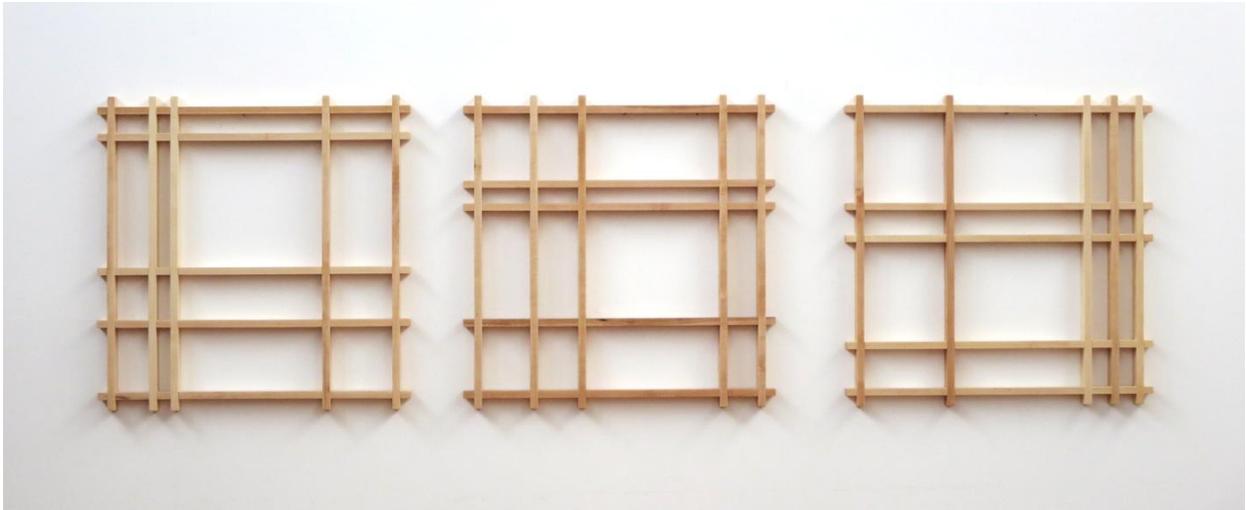


Figure 9. Triptych constructed from designs generated by a computer program. Each panel measures 31"x31". Panels are titled, from left to right: *Cleveland* ($V_E = 4.6$); *Worcester* ($V_E = 4.4$); *Cambridge* ($V_E = 4.5$).

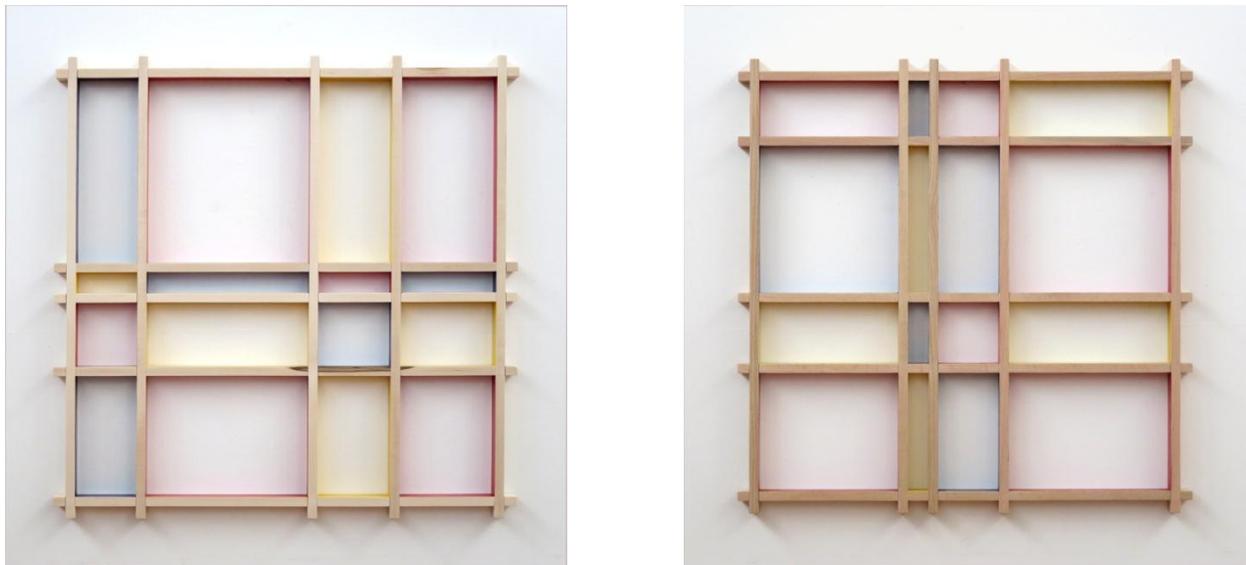


Figure 10. Artworks constructed from designs generated by a computer program. Each panel measures 36"x36". Panels are titled, from left to right: *Amsterdam* ($V_E = 4.5$); *Budapest* ($V_E = 4.5$).

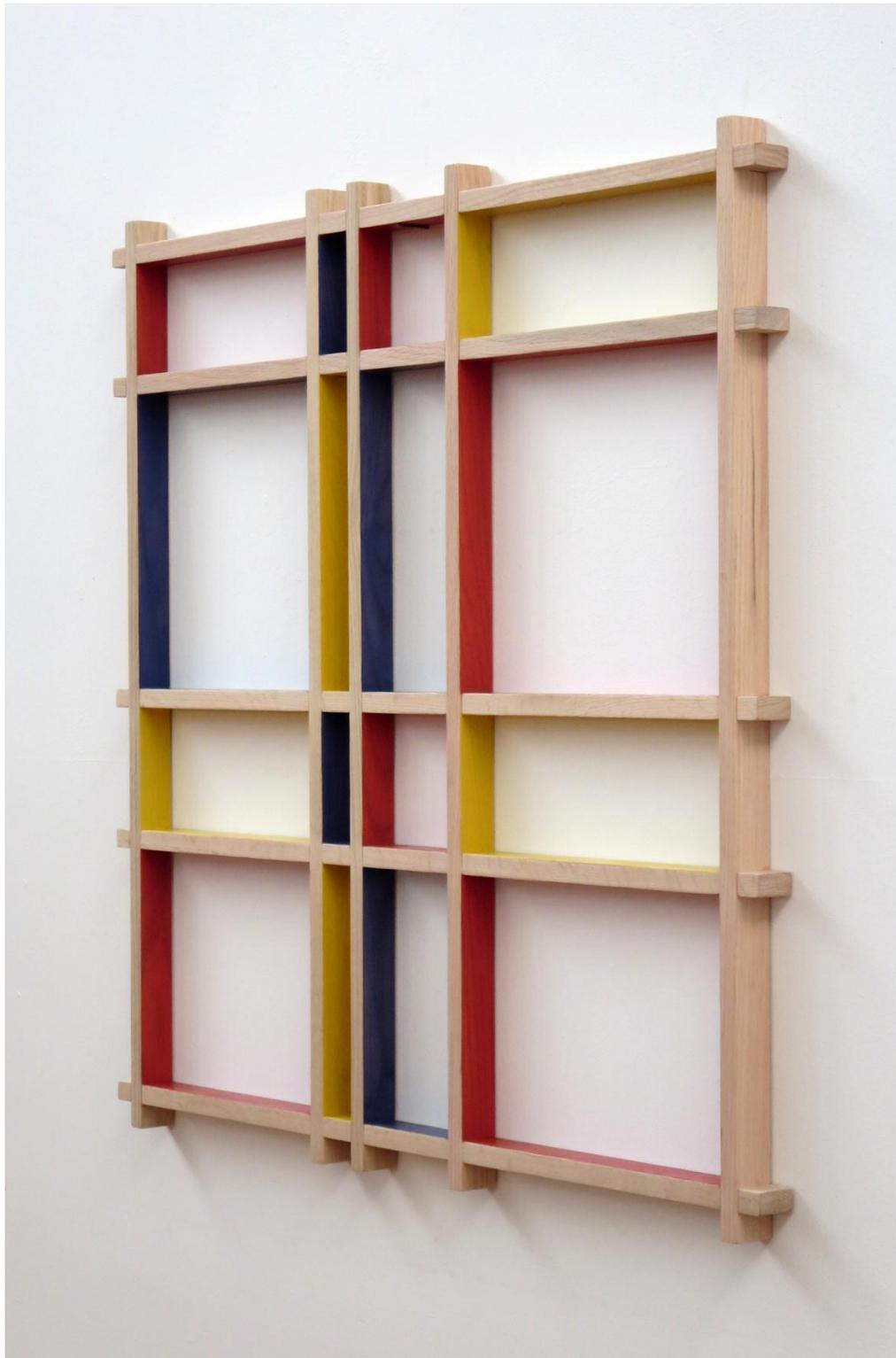


Figure 11. Oblique view, *Budapest* ($V_E = 4.5$).

References

- [1] S. Argamon, K. Burns, and S. Dubnov (eds.), *The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning*, Springer-Verlag, Berlin, 2010.
- [2] M. Bense, *Aesthetica. Einführung in die Neue Aesthetik*, Agis-Verlag, Baden-Baden, 1965.
- [3] D. Berlyne, *Aesthetics and Psychobiology*, Appleton Century Crofts, New York, 1971.
- [4] D. Berlyne, *Studies in the New Experimental Aesthetics*, Hemisphere, Washington, 1974.
- [5] G. Birkhoff, *Aesthetic Measure*, Harvard University Press, Cambridge, MA, 1933.
- [6] C. Blinderman, *T. H. Huxley's Theory of Aesthetics: Unity in Diversity*, *The Journal of Aesthetics and Art Criticism* 21 (1962), pp. 49-55.
- [7] K. Burns, *Atoms of EVE': A Bayesian Basis for Esthetic Analysis of Style in Sketching*, *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing* 20 (2006), pp. 185-199.
- [8] K. Burns, *EVE's Energy in Aesthetic Experience: A Bayesian Basis for Haiku Humor*, *Journal of Mathematics and the Arts* 6 (2012), pp. 77-87.
- [9] K. Burns, *Computing the Creativeness of Amusing Advertisements: A Bayesian Model of Burma-Shave's Muse*, *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing* 29 (2015), pp. 109-128.
- [10] K. Burns and S. Dubnov, *Memex Music and Gambling Games: EVE's Take on Lucky Number 13*, *Papers from the AAAI Workshop on Computational Aesthetics: Artificial Intelligence Approaches to Beauty and Happiness*, WS-06-04, AAAI Press, Menlo Park, CA, 2006, pp. 30-36.
- [11] H. Cohen, *Style as Emergence (from What?)*. S. Argamon, K. Burns, and S. Dubnov (eds.), *The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning*, Springer-Verlag, Berlin, 2010, pp. 3-20.
- [12] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.
- [13] A. Danto, *What Art Is*, Yale University Press, New Haven, CT, 2013.
- [14] C. Darwin, *The Expression of the Emotions in Man and Animals*, Third Edition, Oxford University Press, Oxford, 1998 (First published in 1872).
- [15] A. Ekárt, A. Joó, D. Sharma, and S. Chalakov, *Modeling the Underlying Principles of Human Aesthetic Preference in Evolutionary Art*, *Journal of Mathematics and the Arts* 6 (2012), pp. 107-124.
- [16] P. Ellsworth and K. Scherer, *Appraisal Processes in Emotion*. R. Davidson, K. Scherer, and H. Goldsmith (eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford, 2003, pp. 572-595.

- [17] H. Eysenck, *The Experimental Study of the 'Good Gestalt' – A New Approach*, Psychological Review 49 (1942), pp. 344-364.
- [18] H. Eysenck, *Sense and Nonsense in Psychology*, Penguin, Harmondsworth, UK, 1957.
- [19] G. Fechner, *Vorschule der Ästhetik*, Breitkopf & Härtel, Leipzig, 1876.
- [20] P. Fishwick (ed.), *Aesthetic Computing*, MIT Press, Cambridge, MA, 2006.
- [21] R. Fry, *Vision and Design*, Chatto & Windus, London, 1920, pg. 19.
- [22] P. Galanter, *Computational Aesthetic Evaluation: Past and Future*. J. McCormack and M. d'Inverno (eds.), *Computers and Creativity*, Springer-Verlag, Berlin, 2012, pp. 255-293.
- [23] G. Greenfield, *On the Origins of the Term "Computational Aesthetics"*. L. Neumann, M. Sbert, B. Gooch, and W. Purgathofer (eds.), *Computational Aesthetics in Graphics, Visualization and Imaging*, Eurographics, Girona, Spain, 2005, pp. 1-12.
- [24] G. Greenfield and P. Machado, *Guest Editor's Introduction, Special Issue on Mathematical Models Used in Aesthetic Evaluation*, Journal of Mathematics and the Arts 6 (2012), pp. 59-64.
- [25] F. Hoenig, *Defining Computational Aesthetics*. L. Neumann, M. Sbert, B. Gooch, and W. Purgathofer (eds.), *Computational Aesthetics in Graphics, Visualization and Imaging*, Eurographics, Girona, Spain, 2005, pp. 13-18.
- [26] A. Hofstadter and R. Kuhns, *Philosophies of Art and Beauty: Selected Readings in Aesthetics from Plato to Heidegger*, University of Chicago Press, Chicago, 1964.
- [27] D. Hubel, *Eye, Brain, and Vision*, W. H. Freeman, New York, 1988.
- [28] F. Hutcheson, *An Inquiry Into the Original of Our Ideas of Beauty and Virtue*, J. and J. Knapton, 1729, pg. 17.
- [29] T. Huxley, *On Science and Art in Relation to Education*, Collected Essays Volume III, 1882.
- [30] C. Kaplan and R. Bosch, *TSP Art*, Proceedings of Bridges: Mathematical Connections in Art, Music, and Science, Alberta, Canada, 2005, pp. 301-308.
- [31] J. Kennedy, *Drawing and the Blind: Pictures to Touch*, Yale University Press, New Haven, CT, 1993.
- [32] H. Kreitler and S. Kreitler, *Psychology of the Arts*, Duke University Press, Durham, NC, 1972.
- [33] S. Kullback and R. Leibler, *On Information and Sufficiency*, Annals of Mathematical Statistics 22 (1951), pp. 79-86.
- [34] M. Livingstone, *Vision and Art: The Biology of Seeing*, Abrams, New York, 2002.
- [35] M. Livio, *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*, Broadway Books, New York, 2002.

- [36] M. Livio, *The Equation that Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry*, Simon & Schuster, New York, 2005.
- [37] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*, W. H. Freeman, New York, 1982.
- [38] I. McManus, B. Cheema, and J. Stoker, *The Aesthetics of Composition: A Study of Mondrian*, *Empirical Studies of the Arts* 11 (1993), pp. 83-94.
- [39] H. McWhinnie, *A Review of Some Research on Aesthetic Measure and Perceptual Choice*, *Studies in Art Education* 6 (1965), pp. 34-41.
- [40] K. Millis, *Making Meaning Brings Pleasure: The Influence of Titles on Aesthetic Experiences*, *Emotion* 1 (2001), pp. 320-329.
- [41] A. Moles, *Information Theory and Esthetic Perception*, University of Illinois Press, Urbana, IL, 1966.
- [42] F. Nake, *Information Aesthetics: An Heroic Experiment*, *Journal of Mathematics and the Arts* 6 (2012), pp. 65-75.
- [43] A. Noll, *Human or Machine: A Subjective Comparison of Piet Mondrian's "Composition with Lines" (1917) and a Computer-Generated Picture*, *The Psychological Record* 16 (1966), pp. 1-10.
- [44] K. Oatley, *Creative Expression and Communication of Emotions in the Visual and Narrative Arts*. R. Davidson, K. Scherer, and H. Goldsmith (eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford, 2003, pp. 481-502.
- [45] M. Parsons, *How We Understand Art: A Cognitive Developmental Account of Aesthetic Experience*, Cambridge University Press, Cambridge, UK, 1987.
- [46] F. Phillips, J. Norman, and A. Beers, *Fechner's Aesthetics Revisited*, *Seeing and Perceiving* 23 (2010), pp. 263-271.
- [47] Z. Pizlo, *3D Shape: Its Unique Place in Visual Perception*, MIT Press, Cambridge, MA, 2008.
- [48] R. Reber, N. Schwarz, and P. Winkielman, *Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?* *Personality and Social Psychology Review* 8 (2004), pp. 364-382.
- [49] J. Rigau, M. Feixas, and M. Sbert, *Conceptualizing Birkhoff's Aesthetic Measure Using Shannon Entropy and Kolmogorov Complexity*. D. Cunningham, G. Meyer, L. Neumann, A. Dunning, and R. Paricio (eds.), *Computational Aesthetics in Graphics, Visualization and Imaging*, The Eurographics Association, 2007, pp. 105-112.
- [50] J. Rigau, M. Feixas, and M. Sbert, *Informational Aesthetics Measures*, *IEEE Computer Graphics and Applications* 28 (2008), pp. 24-34.
- [51] R. Scha and R. Bod, *Computationele Esthetica*, *Informatie en Informatiebeleid* 11 (1993), pp. 54-63.

- [52] K. Scherer, *Appraisal Theory*. T. Dalgleish and M. Power (eds.), *Handbook of Cognition and Emotion*, Wiley, New York, 1999, pp. 637-663.
- [53] C. Shannon, *The Mathematical Theory of Communication*. C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949, pp. 29-125.
- [54] P. Silvia, *Emotional Responses to Art: From Collation and Arousal to Cognition and Emotion*, *Review of General Psychology* 9 (2005), pp. 342-357.
- [55] B. Smith (ed.), *Foundations of Gestalt Theory*, Philosophia Verlag, Munich and Vienna, 1988.
- [56] S. Stevens, *Neural Events and the Psychophysical Law*, *Science* 170 (1970), pp. 1043-1050.
- [57] G. Stiny and J. Gips, *Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts*, University of California Press, Berkeley, CA, 1978.
- [58] P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich, *Viewpoint Selection Using Viewpoint Entropy*, *Proceedings of the Vision Modeling and Visualization Conference* (2001), pp. 273-280.
- [59] W. Weaver, *Introductory Note on the General Setting of the Analytical Communication Studies*. C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949, pp. 1-28.
- [60] A. Wolach and M. McHale, *Line Spacing in Mondrian Paintings and Computer-Generated Modifications*, *The Journal of General Psychology* 132 (2005), pp. 281-291.
- [61] J. Xu and C. Kaplan, *Vortex Maze Construction*, *Journal of Mathematics and the Arts* 1 (2007), pp. 7-20.
- [62] M. Zuckerman, *Behavioral Expressions and Biosocial Bases of Sensation Seeking*, Cambridge University Press, New York, 1994.

Appendix A.

As noted in Section 1.1, Birkhoff [5] offered "A *Mathematical Argument*" as a basis for his aesthetic measure $M = O / C$. This appendix examines the argument and identifies three important errors.

Birkhoff's argument applies to a "class of objects" [5, pg. 11] for which aesthetics can be "readily compared and arranged in order of preference". His argument assumes the aesthetic measure M is a function of order O and complexity C , and he uses a hypothetical experiment to establish the function by which M can be computed from O and C . Birkhoff's hypothetical experiment involves only a few steps [5, pp. 12-13], which are numbered and quoted here to expose the logic as follows:

1. "Suppose that we have before us a certain set of k objects of the class, all having the same order O and the same complexity C , and also a second set of k' objects of the class, all having the order O' and complexity C' . Let us choose k and k' so that $k'C'$ equals kC ."
2. "Now proceed as follows. Let all of the first set of objects be observed, one after the other; the total effort will be measured by kC of course, and the total tone of aesthetic feeling by kO . Similarly let all of the second set be observed. The effort will be the same as before, since $k'C'$ equals kC ; and the total tone of feeling will be measured by $k'O'$."
3. "If the aesthetic measure of the individual objects of the second class is the same as of the first, it would appear inevitable that the total tone of feeling must be the same in both cases, so that $k'O'$ equals kO . With this granted, we conclude at once that the ratios O' / C' and O / C are the same."
4. "In consequence the aesthetic measure only depends upon the ratio O to C : $M = f(O / C)$. The final step can now be taken. Since it is not the actual numerical magnitude of f that is important but only the relative magnitude when we order according to aesthetic measure, and since M must increase with O / C , we can properly define M as equal to the ratio of O to C ."

This argument contains three critical errors, as follows:

First, at step 2 Birkhoff assumes that the total "total tone of aesthetic feeling" is given by kO , yet in step 1 he defined O as order. The premise of the argument assumes M measures aesthetics, so the total tone of aesthetic feeling for each set of objects should be kM and $k'M'$, not kO and $k'O'$.

Second, at step 3 Birkhoff equates $k'O'$ and kO . But this holds true only if O denotes the "tone of [aesthetic] feeling" (which it does not), and "if the aesthetic measure" is the same (which it is not). In particular, there is no reason to assume $k'O' = kO$ at step 3 after having assumed $k'C' = kC$ at step 1, because O and C are independent variables in the proposed function $M = f(O, C)$. There is also no reason to assume "the aesthetic measure of the individual objects of the second class is the same as of the first", because M is an unknown function of the independent variables O and C .

Finally, even if we ignore the preceding errors, the equations $k'C' = kC$ and $k'O' = kO$ together imply two ratio equivalences, namely $C' / O' = C / O$ as well as $O' / C' = O / C$. More importantly, there is no reason to conclude $M = C / O$ or $M = O / C$, because the derivation only implies there are two ratio equivalences – not that either one represents a measure of aesthetics.