

Chapter 2

Cross-Cultural Aesthetics: Analyses and Experiments in Verbal and Visual Arts

Abstract Aesthetics arise from arousal of the senses followed by appraisals that make sense of what was sensed. The underlying psychology of sense-making can be modeled mathematically as a reduction in entropy, from the initial entropy of arousal to residual entropy after appraisal. This theoretical framework is applied here to the aesthetics of verbal and visual arts across American and Japanese cultures. First, a computational model is proposed to analyze the aesthetics of humor in haiku form and amusing advertisements known as Burma-Shave jingles. These analyses demonstrate how aesthetic experiences can be computed in terms of entropy reduction, for both forms of verse. Then, an experimental study is performed to characterize preferences for complexity in abstract art across samples of American and Japanese populations. This experiment further illustrates how aesthetics can be computed in terms of entropy, establishing commonalities between the two cultures and uncovering differences in aesthetic preferences of individuals.

2.1 A Theoretical Framework for Computing Aesthetics

Aesthetics are often attributed to “*unity in variety*” [7, 9, 25, 27–29]—where unity refers to some measure of order, regularity, symmetry, or harmony; and variety refers to some measure of chaos, irregularity, asymmetry, or complexity. More formally, a number of authors have offered mathematical measures of unity and variety, along with equations for how these two factors are combined to compute aesthetics [6, 8, 23, 31, 37, 45]. Unfortunately, most of these equations are lacking a psychological foundation and empirical validation.

One alternative is a recent theory dubbed EVE', which models the aesthetic experience mathematically in terms of psychological expectations (E), violations (V) of expectations, and explanations (E') of violations. Formal analyses and human experiments have established that EVE' can explain and predict aspects of aesthetics in verbal, visual, and musical media [10–14].

This initial section of the chapter compares EVE' to other theories of aesthetics, in order to establish key differences and equivalences. In particular, EVE' is shown

to be similar to another equation that quantifies the audience's experience as an information rate (IR) in signal processing [17–19]. Then subsequent sections of the chapter apply EVE' to analyses and experiments in verbal and visual arts. The results of these studies offer insights into universal aspects of aesthetics across American and Japanese cultures, as well as personal preferences of individuals within each culture.

2.1.1 Birkhoff and Bense

Arguably the first formal theory of aesthetics was offered by George Birkhoff [8] as an equation that expressed his notion of “*unity in variety*”. Birkhoff's equation was $M = O/C$, where O is order, C is complexity, and M is a measure of aesthetics obtained from order (unity) in complexity (variety). Over the years this equation has been influential among scientists attempting to model aesthetics [6, 7, 20, 22–24, 29, 31, 32, 35, 37, 40, 45], despite some important limitations.

Birkhoff's equation was based only on personal introspection, without any psychological foundation or empirical validation to support the assumed relation between M , O , and C . For example, Hutcheson's [27, 45] notion of “*unity in variety*” implies an equation $M = O * C$, as proposed by Eysenck [23], as opposed to $M = O/C$. And actually empirical testing reviewed by Eysenck [22, 24] shows that $M = O * C$ offers a better match to psychological judgments of aesthetics than $M = O/C$. But a limitation of both equations lies in how the factors O and C are quantified. Birkhoff himself proposed only ad hoc methods for computing O and C from features of artworks, and neither he nor subsequent researchers ever tested those methods against human judgments of order and complexity. As a result, failure of Birkhoff's M to match human judgments of aesthetics could be attributed to the equation $M = O/C$, or to methods by which an experimenter computes values of O and C needed as input to this equation.

Several decades after Birkhoff, other scientists led by Max Bense [6, 32, 40] used concepts from Shannon's theory of communication [42] to more formally compute the factors O and C . In particular, Bense assumed that complexity (C) could be quantified by Shannon's entropy (H), which measures the information in a set of signals. This entropy is computed as $H = \sum_i P_i * -\log P_i$, where P_i is the probability (frequency) at which a signal s_i occurs and the sum is taken over the set of all signals $\{s_i\}$ in an artwork. The term $-\log P_i$, known as *surprisal* [15, 42], is a positive quantity that increases as P decreases. That is, an improbable signal s_i with low P_i and high $-\log P_i$ will be *surprising* when it is received. Therefore, Shannon's entropy H can be characterized as a weighted-average surprisal across all signals in an artwork, with surprisal $-\log P_i$ for each signal s_i weighed by the frequency P_i at which that signal appears in the artwork.

Several models are attributed to Bense and his contemporaries, but these models all assume Birkhoff's equation and differ only in how the factors O and C are quantified. One model [37] is $O = H_{\max} - H$, $C = H_{\max}$, and $M = 1 - H/H_{\max}$,

where H_{\max} is the maximum value of entropy that occurs when all signals in the set $\{s_i\}$ are equally probable. For a given set of possible signals, H_{\max} is a constant such that M increases as H decreases. Another model [32] is $O = (H_{\max} - H)/H_{\max}$, $C = H$, and $M = 1/H - 1/H_{\max}$. This model also predicts that M increases as H decreases, such that aesthetics M would be maximized when entropy H is minimized.

2.1.2 IR and EVE'

More recently, principles of information theory have been used to derive an expression [17–19] known as *information rate* (IR). Unlike Bense's formulation, which is concerned with a compositional set of signals that comprise an artwork (e.g., an image), IR is concerned with a temporal sequence of signals received by an audience (e.g., in music). Also unlike Bense's formulation, IR does not adopt Birkhoff's equation $M = O/C$. Instead IR computes a measure of aesthetics as mutual information, between a signal without any past context and the same signal given its past context. This mutual information is a difference between two entropies, $IR = H - H'$, where: $H = -\log P(\text{signal})$ is the entropy of a signal *independent* of past signals; and $H' = -\log P(\text{signal}|\text{past signals})$ is the entropy of the same signal *in the context* of past signals.

Other authors [3, 31, 45] have also argued that aesthetics involve a reduction in entropy. Indeed, Bense himself had this idea [40] in a model written as $M = O/C = (H_1 - H_2)/H_1$, where H_1 is an entropy computed from symbols that encode the stimulus, and H_2 is an entropy computed from super-symbols that re-code the stimulus. But IR differs in its approach to compression, by computing the entropy of a signal in the context of past signals.

Further insight into IR as a model of aesthetics can be gained by examining the underlying psychological processes of *arousal* and *appraisal* [7, 21, 33, 34, 41, 43, 44]. This perspective is promoted by the theory of EVE' [10–14], which models mental expectations (E), violations (V) of expectations, and explanations (E') of violations. EVE's equation for aesthetics is $X = Y * Z$, where: Y models arousal from the violation of an expectation (i.e., the V of E); and Z models appraisal from an explanation of the violation (i.e., the E' of V). Mathematically, Y is computed via information theory [15, 42] as the marginal entropy of a signal that is received, $-\log P(\text{signal})$; and Z is computed via Bayesian theory [5, 30] as the posterior probability of the meaning that most likely *explains* the signal that was received, $P(\text{meaning}|\text{signal})$.

As outlined above, EVE's aesthetic measure X is expressed as the product of an entropy and probability, whereas IR's aesthetic measure is expressed as a difference between two entropies. But an equivalence exists if one assumes that the context provided by past signals can *explain* the meaning of a present signal. Using H to denote the marginal entropy $-\log P(\text{signal})$ that quantifies a violation of expectation, and using Q to denote the posterior probability $P(\text{meaning}|\text{signal})$ that

quantifies explanation of the violation, EVE's product of arousal and appraisal is $X = Y * Z = H * Q$. An equivalence to IR is then seen by rewriting the equation as follows: $X = H - H * (1 - Q) = H - H' = IR$.

In this equation, H represents the magnitude of initial surprise experienced as arousal, and $H' = H * (1 - Q)$ represents the magnitude of residual surprise *not explained* in appraisal. Thus H' and $H * (1 - Q)$ are merely different ways of computing how much surprise has been experienced and *not explained*. So both IR and EVE' are measuring *how much surprise has been experienced and explained*.

Now returning to the equations of Birkhoff-Bense, we can distinguish an important difference from IR-EVE'. This is seen most easily from EVE's formulation $X = Y * Z$, where marginal entropy $Y = -\log P(\text{signal})$ is a measure of complexity akin to C , and posterior probability $Z = P(\text{meaning}|\text{signal})$ is a measure of order akin to O . By substitution, EVE's equation can be rewritten as $X = C * O$, which is equivalent to Eysenck's equation $M = O * C$ and different from the Birkhoff-Bense equation $M = O/C$.

2.1.3 Models of Memory

An important aspect of IR, as it relates to EVE', is the idea that past signals can explain the meaning of a present signal. Here one might wonder: If past signals (which have just been received prior to the present signal) are capable of *explaining* the present signal, then why were these same past signals not capable of *expecting* the present signal? In other words, if past signals can *reduce* surprise by an amount $H - H'$, then why do these past signals not *prevent* the same amount of surprise from occurring in the first place?

Here the psychological perspective of EVE' offers insight by modeling two types of memory involved in expectations (E) and explanations (E'), respectively. First, *expectations* (and violations of expectations) are governed by representations of meanings and signals held in *working memory*. The working memory of humans is known to be quite limited [4, 16], such that only a handful of so-called "chunks" are able to hold possible meanings and potential signals. On the other hand, *explanations* are based on extremely rich associations between meanings and signals that are represented in *long-term memory*.

In effect, the explanation of a present signal by past signals is a *re-cognition*—whereby associations between present and past signals are recalled from long-term memory *only after* the present signal is actually received and represented (along with past signals) in working memory. Therefore, past signals can be effective in explaining a present signal (via long-term memory), after it has been received; whereas past signals are less effective in expecting the same signal (via working memory), before it has been received. This difference in effectiveness, due to different memory systems, helps account for how past signals *can explain* but *not expect* a subsequent signal in a sequence of signals.

The nature of long-term memory also offers insight into EVE's notion of *meaning*, and IR's assumption that past signals can *explain* a present signal. That is, re-cognition entails associations between the present signal and past signals as these signals are all represented in long-term memory. But long-term memory also extends far beyond signals to include associated knowledge at a higher level of abstraction—which EVE' refers to as *meaning*. Thus together the past and present signals held in working memory will *evoke* associated meanings via long-term memory, and these evoked meanings (rather than past signals per se) are what actually *explain* the present signal.

This conceptual difference between *signals* and *meanings* is important for two reasons. First, explanations of signals (data) require associations to meaning (knowledge) at a higher level of abstraction, because signals at one level of abstraction cannot really *explain* other signals at the same level of abstraction. Second, the co-occurrence statistics of signals within artworks are often not sufficient to compute the higher-level semantics of meaning needed to *explain* signals. A specific example is provided in the next section, which uses EVE' to analyze the aesthetics of poetic verse.

2.2 A Computational Model of Verbal Aesthetics

In the domain of verbal arts, EVE' has been used to explain and predict aesthetics for two types of short verse. These two types were chosen to control for variables such as the total length of a verse and the grouping of words in lines, which can affect judgments of complexity and aesthetics. One type is haiku, a well-known form in Japanese tradition but also written in many other languages including English [1, 2, 26, 36, 38, 46]. The other type is rhyming “jingles”, used by an American company to advertise their shaving cream called Burma-Shave [39]. For both forms of verse, the aesthetic of interest here is one of humorous amusement.

2.2.1 *Haiku Humor*

In Japan, humor in haiku form is often referred to as *senryu*, after the name of a poet who popularized this comic genre in the 1700s. But outside Japan the term haiku is used more broadly, and in English all poems structured in three lines with $5 + 7 + 5 = 17$ syllables are properly referred to as haiku [36]. The following is an example that illustrates the theory of EVE':

an apple a day
will keep the doctor away
said devil to Eve

In analyzing this example [11], each line can be considered a signal received in temporal sequence as the verse is read. But besides these three *signals* (s_1, s_2, s_3) that serve as *evidence*, an analysis must also consider the *meanings* that an audience entertains as *hypotheses* to explain the signals. For example, after the first two lines a reader will think it is very likely that the verse is referring to a well-known proverb about how apples are healthy and hence good to eat. So after s_1 and s_2 , but before s_3 , the reader will be thinking $P(A|s_1, s_2) = \text{high}$ and $P(\sim A|s_1, s_2) = \text{low}$, where: A = “apples are good to eat” is a hypothesized meaning associated with the well-known proverb; and $\sim A$ is a catch-all hypothesis that includes other possible but unspecified meanings of the verse.

Although many specific meanings may be represented in long-term memory and used to *explain* signals, only a small set like $\{A, \sim A\}$ can be retained in working memory and used to *expect* the next signal. So with $P(A|s_1, s_2) = \text{high}$, a reader will expect the next signal s_3 to be in class “a” consistent with A = “apples are good to eat”, such that $P(a|s_1, s_2) = \text{high}$ and $P(\sim a|s_1, s_2) = \text{low}$. Once again, notice that only a small set $\{a, \sim a\}$ of potential signals is represented in working memory and used to *expect* the next signal.

Finally, after expecting a signal $s_3 = \text{“a”}$, the experienced signal is actually $s_3 = \text{“}\sim a\text{”}$. That is, the last line of the verse refers to characters in the context of a bible story where apples are bad to eat, which is inconsistent with A = “apples are good to eat”. This experience of s_3 (after s_2 and s_1) produces a high violation of expectation, or *arousal*, which can be quantified as $Y = -\log P(\sim a|s_1, s_2) = \text{high}$. The arousal, in turn, sparks a search through long-term memory for some meaning that can explain the signal, in a process of *appraisal* to “make sense” of the surprise.

A reader who “gets it” will realize that the meaning of the verse is not A but B , where B = “apples are bad to eat and the devil is using the proverb as a way to tempt Eve”. Notice that B comes from background knowledge about bible stories and the devil’s intent to deceive Adam and Eve in the Garden of Eden. This knowledge (recalled from long-term memory) is not contained in signals of the verse itself, because none of the signals makes mention of any individual’s aims or intentions. But based on a reader’s background knowledge, the meaning B makes sense of the surprising signal s_3 after previous signals s_2 and s_1 , so the Bayesian posterior probability is $P(B|s_1, s_2, s_3) = \text{high}$.

Finally, EVE’ computes the magnitude of aesthetic experience as $X = Y * Z$, where $Y = -\log P(\sim a|s_1, s_2) = \text{high}$ and $Z = P(B|s_1, s_2, s_3) = \text{high}$. This yields $X = \text{high}$, so the haiku is humorous to those who experience and understand it.

In contrast, consider two alternative endings that yield low X , one due to low Y and the other due to low Z . For the first case, assume $s_3 = \text{“said the old proverb”}$. Here we have little surprise (arousal), because s_3 is in class “a” consistent with A , so s_3 is expected after s_1 and s_2 . The resulting verse makes sense (high Z), but $X = Y * Z$ is low because Y is low. For the second case, assume $s_3 = \text{“said postman to Eve”}$. Here we have little meaning (appraisal), because background knowledge offers no apparent reason for a postman to be reciting the apple proverb to Eve (of bible fame,

or other Eve). So s_3 may be surprising (high Y), but $X = Y * Z$ is low because Z is low.

As described above, a high aesthetic X requires both high surprise Y and high meaning Z together. Now with respect to IR, this same product $Y * Z$ might be modeled as an entropy difference $H - H'$, where: $H = -\log P(s_3|s_1, s_2) = \text{high}$, to model the initial surprise experienced before explanation of s_3 ; and $H' = -\log P(s_3|s_1, s_2, B) = \text{low}$, to model the residual surprise remaining after explanation of s_3 . But notice the context B that explains the signal s_3 is not past signals (s_1 and s_2) themselves, because the past signals were received before s_3 , and actually these past signals need to be explained along with the present signal in order to make sense of the verse. Instead the context that explains s_3 (and s_2 and s_1) is an associated meaning B at a higher level of abstraction. This meaning is *re-cognized* only after all three signals (s_1 , s_2 , and s_3) are received and have evoked long-term memories of the apple proverb and bible story. In fact, re-cognition involves *re-combination* of these memories, in *creation* of an explanation for the novel situation (i.e., devil reciting the apple proverb) that was not previously stored in long-term memory.

This example illustrates commonalities as well as a key difference between IR and EVE'. A common aspect of both approaches is to quantify surprise in terms of entropy, denoted H in IR and Y in EVE'. The difference then comes in computing how this surprise is explained or not by an audience. IR models the amount of surprise that is *not explained by past signals*, via the term $H' = H(\text{signal}|\text{past signals})$. This makes IR applicable to artworks for which surprise can be explained by the co-occurrence statistics of signals. EVE' models the fraction of surprise that *is explained by some meaning*, via the term $Z = P(\text{meaning}|\text{signals})$. This makes EVE' applicable to other artworks, like the apple haiku, for which surprise is explained by meanings that cannot be computed from the co-occurrence statistics of signals themselves.

2.2.2 Serious Semantics

Of course not all haiku are humorous, and EVE' also applies when the semantics are serious. As an example, consider a famous verse by Bashō:

furu ike ya
kawazu tobikomu
mizu no oto

which is translated by Addiss [1] as follows:

old pond
a frog jumps in
the sound of water

Here the first two lines describe a sight, but then the last line describes a sound—which is surprising because it involves “sense-switching” [36]. The surprise is explained by combining sight and sound into a coherent re-cognition of the familiar frog-pond scene, based on long-term memories of natural experiences. As such, the aesthetic arises from meaning assigned to signals that initially appear incongruous but eventually are found harmonious.

Thus like high amusement for humorous haiku, high aesthetics for serious haiku are the product of high surprise (Y) and high meaning (Z). And once again, the meaning assigned to signals comes not from signals contained within the artwork itself—but rather from common knowledge shared by a culture of artists and audiences who live in the same natural world.

In today’s state of the art, it remains an unsolved challenge for artificial intelligence to model the meanings that humans assign to signals as they experience artworks. In language if not other media, this requires rich representations of human knowledge about the real world—including hypothetical meanings of evidential signals, along with prior probabilities of the form $P(\text{meaning})$ and conditional probabilities of the form $P(\text{signal}|\text{meaning})$. These representations are needed to compute the marginal probability of a signal across the set of possible meanings in *working memory*, $P(\text{signal}) = \sum P(\text{meaning}) * P(\text{signal}|\text{meaning})$, in EVE’s factor $Y = -\log P(\text{signal})$. They are also needed to compute posterior probabilities of possible meanings from *long-term memory*, $P(\text{meaning}|\text{signal}) = P(\text{meaning}) * P(\text{signal}|\text{meaning})/P(\text{signal})$, for EVE’s factor $Z = P(\text{meaning}|\text{signal})$.

In short, models of *meanings* for signals are required if artificial intelligence is to credibly compute aesthetic experiences—which are governed by how much surprise is experienced and *explained*.

2.2.3 Amusing Advertisements

Unfortunately, the needed knowledge structures are not currently available from artificial intelligence. As a consequence, it is not feasible to compute EVE’s factors Y and Z from first principles, at least not for narrative or figurative arts in which rich semantics are involved. But EVE’ can still be applied to these arts by asking humans for their judgments of meaning (Z) and surprise (Y), from which aesthetics (X) can be computed as $X = Y * Z$. The model-predicted values of X can then be compared to human-reported judgments of aesthetics, as a test of EVE’s equation $X = Y * Z$.

This approach was applied [13] in a study of advertising jingles made famous by the Burma-Shave company [39]. These jingles are similar to haiku in that they have a fixed form and short length. More specifically, each jingle is written in five lines with a total number of syllables similar to the 17 of haiku. In Burma-Shave’s original advertising campaign, each line of a jingle was painted on a separate sign and the signs were spaced along roadways so that readers in vehicles would experience a several-second delay between lines. An example is the following verse:

he's the boy
 the gals forgot
 his line
 was smooth
 his chin was not

A total of 20 prototypical jingles were chosen from among 600 published verses [39] originally used as advertisements by the Burma-Shave company. These 20 jingles were then presented in a survey [13] that asked human participants to make three judgments (on a scale of 1 to 5) for each jingle. The judgments were answers to the following three questions: *Did you understand the verse? Did the ending surprise you? Was the jingle creative?* At the end of the survey, all participants strongly or very strongly agreed that their judgments of creativeness were based on feelings of amusement. Thus answers to the three questions provide direct measures of EVE's variables Z (meaning), Y (surprise), and X (pleasure), respectively.

The data collected by this survey were used to compute average values of Z, Y, and X across all participants ($N = 81$), for each jingle. The average human values of Y and Z were then used to compute aesthetic predictions for five models: $X = Y * Z$; $X = Z/Y$; $X = Y/Z$; $X = Y$; and $X = Z$. Notice that the first model $X = Y * Z$ is EVE's equation and analogous to Eysenck's equation $M = C * O$, whereas the second model $X = Z/Y$ is analogous to the equation $M = O/C$ of Birkhoff and Bense.

These model-predicted values of X were compared to the human-reported (average) values of X for all 20 jingles. The results [13] showed that EVE's model $X = Y * Z$ was by far the best, with model-predicted X accounting for a remarkable 70 % of the variation in human-reported X across jingles. The Birkhoff-Bense model $X = Z/Y$ was the worst of all five models tested. In fact, it was the only model to predict a negative correlation across jingles, where model-predicted X decreased as human-reported X increased.

These results strongly suggest that the term Y representing complexity C belongs in the numerator of $X = Y * Z$, as modeled by EVE' and Eysenck's equation $M = O * C$; rather than in the denominator, as modeled by the Birkhoff-Bense equation $M = O/C$.

2.3 An Experimental Study of Visual Aesthetics

The previous section provided empirical support for EVE's equation $X = Y * Z$ over other mathematical models of aesthetic experience. However, testing required human judgments of Y and Z as input to the equation, and a more thorough test of EVE' would compute Y and Z directly from first principles.

As noted earlier, computing Y and Z from first principles requires detailed models of human knowledge that are not currently available from artificial

intelligence. This knowledge of *meaning* extends far beyond the data in *signals* comprising artworks themselves. But for some forms of art, it appears that the magnitude of an explanation can be estimated—at least to a first approximation—without explicitly modeling the meanings by which signals are explained.

One example is in music, where IR uses the associative context provided by past signals to explain present signals, without an explicit model of meaning. Using this approach, empirical testing of two contemporary music compositions [18, 19] showed that IR could predict 20–40 % of the variation in human judgments for “emotional force”, as measured continuously while subjects listened to these compositions. Another example is abstract art [12], which is analyzed by EVE’ and empirically tested below. Abstract art is much like music, and unlike other visual or verbal media, because the features (signals) of an abstract design do not represent objects (meanings) outside the artwork itself.

2.3.1 Abstract Artworks

In the realm of abstract art, a constrained form of composition using only vertical and horizontal grid lines [12] was used to simplify the analysis of semantics in EVE’s factor Z. For these artworks, it is reasonable to assume that the meaning assigned to signals (i.e., grid lines and shapes) is one of overall *coherence* (i.e., order or balance) that characterizes the composition. With this assumption, Z (coherence) can be modeled as an inverse function of Y (complexity): $Z = (Y_{\max} - Y)/Y_{\max}$. Notice that this is basically the same assumption as we saw earlier in a model by one of Bense’s contemporaries, where order was written as $O = (H_{\max} - H)/H_{\max}$. But for EVE’ the approach applies only to abstract art in which semantics are extremely simplified. As discussed below, the same grid-line compositions that enable use of this simplified model for Z also enable EVE’s factor Y to be computed directly from the features of an artwork, without any knowledge of real-world objects outside the artwork.

Unlike IR, which models signals in a temporal and sequential fashion, an audience’s processing of a visual artwork is spatial and parallel. Although such artworks are inspected in a temporal sequence via eye saccades, the resulting sequence is constrained more by the audience than by the artwork itself. Because this sequence will vary with the viewer, and because signals are received in parallel as well as sequentially, the process can be modeled as if all signals of an artwork are received together and then explained together. In that case IR equals the difference between initial entropy (of all signals) before explanation and residual entropy (of all signals) after explanation, which EVE’ models as $X = Y * Z$.

A novel aspect of EVE’s approach to these artworks lies in modeling Y as a visual entropy V_E . Details are discussed in [12], but the main point here is that V_E measures spatial complexity in a manner that differs from classical entropy H. In particular, EVE’ computes visual entropy as $V_E = 1/N * \sum_n -\log A_n$, where the sum is taken over all N white shapes (rectangles) bounded by black lines in the grid, and

A_n represents the area of an individual white shape computed as a fraction of the total area. Thus per this V_E , total surprisal for an artwork is computed as the average surprisal across all white shapes (bounded by black lines), with surprisal for each white shape given by $-\log A_n$.

As described above, V_E depends on the number of black lines and spacing between lines, which define the number of white shapes and area of each shape, respectively. In contrast, classical entropy H is computed as $H = \sum_i P_i * -\log P_i$, where i refers to a pixel color (black or white) and P_i is the frequency of that pixel color in the image. Unlike V_E , H depends only on the number of grid lines in an image (of a fixed size) and not the spacing between lines, because the frequencies of black and white pixels are determined only by the number of black lines in a grid.

The experiment [12] involved five pages of designs, shown in Fig. 2.1. For each page, which presented five panels in random order across the page, human participants provided two types of rankings. First, the five panels on a page were ranked from most (rank = 5) to least (rank = 1) *visually complex*. Then, the same five panels were ranked from most (rank = 5) to least (rank = 1) *aesthetically pleasing*. For each page, the average ranking across $N = 148$ human participants in an American population was used to test EVE' and other models.

The human rankings for visual complexity appear in Fig. 2.2, and the human rankings for aesthetic quality appear in Fig. 2.3. The human rankings for visual complexity (Fig. 2.2) were used to test EVE's model for entropy V_E , and also to test the model of classical entropy H . The human rankings for aesthetic quality (Fig. 2.3) were used to test EVE's prediction of aesthetic optimality as a function of visual entropy. The same human rankings for aesthetic quality were also used to test the prediction of Birkhoff-Bense that M will be maximized when classical entropy H is minimized.

The human rankings for visual complexity (Fig. 2.2) showed that EVE's visual entropy V_E was a much better model than classical entropy H . Using Spearman's rank-order correlation coefficient, averaged over all five pages of the experiment, V_E scored a 98 % match to human judgments of visual complexity, whereas H scored only 60 %. The low score for H was because classical entropy depends only on the number of lines in a design, whereas human judgments are clearly sensitive to both the number of lines and the spacing between lines.

The human rankings for aesthetic quality (Fig. 2.3) showed that EVE's model X was much better than the Birkhoff-Bense measure M . Using Spearman's rank-order correlation coefficient, X scored a 91 % match to human judgments, whereas M scored only 42 %. This low score for M was because the Birkhoff-Bense model always predicts the simplest panel on a page will be ranked as most aesthetic, whereas humans typically found panels of intermediate complexity to be most aesthetic.

For example, according to EVE's equation $X = Y * Z = Y * (Y_{\max} - Y)/Y_{\max}$, the optimal aesthetic ($dX/dY = 0$) will occur at $Y_{\text{opt}} = Y_{\max}/2$. As seen from Fig. 2.3, this prediction is generally consistent with the intermediate level of complexity found to be optimal (on average across subjects) in the experiment.

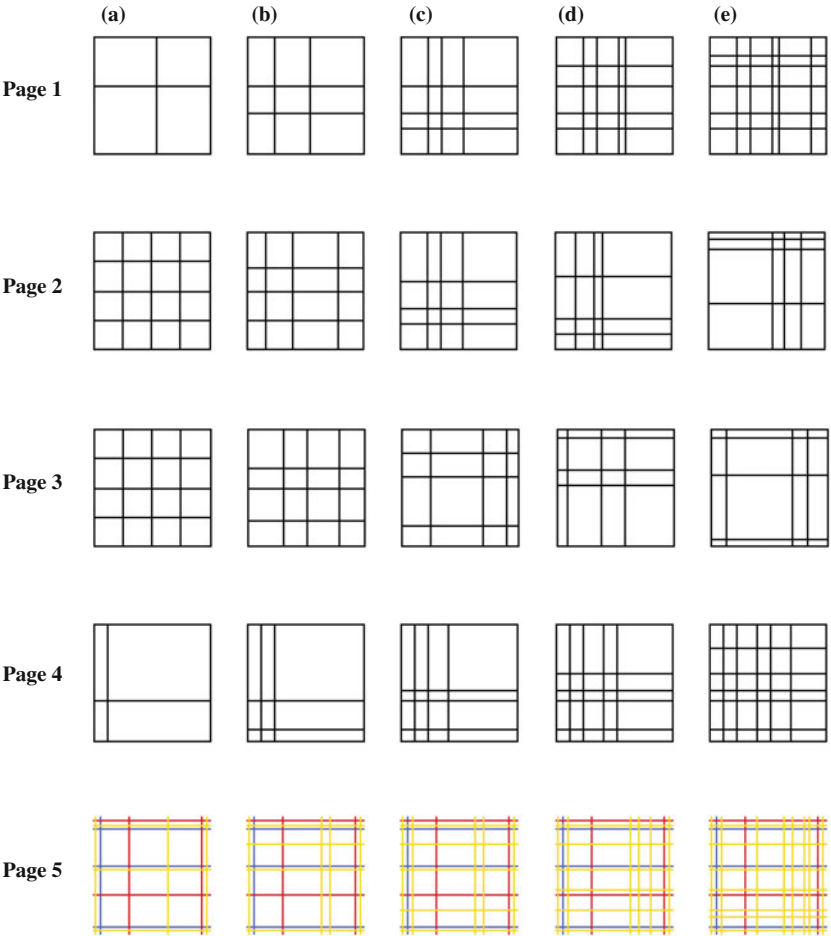


Fig. 2.1 Five panels (*a, b, c, d, e*) for the five pages (*1, 2, 3, 4, 5*) of an experiment. Pages were shown one at a time, with panels arranged in random order across the page. For each page of five panels, participants provided rankings of visual complexity and aesthetic quality

These results are noteworthy for two reasons. First, the scores show that EVE’s equation for X is much better than the aesthetic measure M of Birkhoff and Bense. Second, the experiment also tested factors V_E and H that are needed as inputs to equations for X and M , respectively. Although limited to a simple domain of abstract designs, V_E and H were computed from first principles based only on the visual features of these graphic designs. [Note that human judgments of Z or O were not collected, because pilot testing showed that humans judged coherence (i.e., order or balance) as the logical opposite of complexity.] Thus unlike other attempts to test aesthetic equations of Birkhoff and Bense, this experiment tested not only the equations but also the factors that are input to those equations.

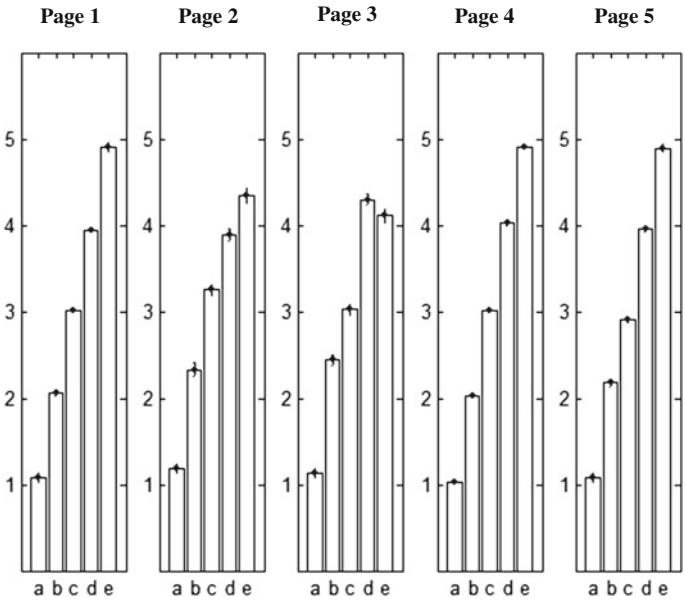


Fig. 2.2 Average human rankings of visual complexity, for five panels (*a, b, c, d, e*) on each page (1, 2, 3, 4, 5) of the experiment. Error bars represent standard errors of the mean. Data are from $N = 148$ American participants

Modeling these inputs is a matter of semantics, because *signals* like “black lines” and “white shapes” can themselves be characterized as *meanings* for lower-level signals like black and white pixels. Here instead EVE’ models meaning as the overall coherence of a composition, and treats black lines and white shapes as signals. The experiment showed that visual entropy V_E captured human judgments of signal complexity, using these black lines and white shapes as signal categories; whereas classical entropy H failed to capture human judgments of signal complexity, using individual pixel colors as signal categories.

Recall that a similar situation regarding the semantics of signals arose earlier in analysis of haiku humor. That is, in theory a line of verse might be treated as the meaning for words in that line. But instead EVE’ models a higher level of abstraction, where lines are considered signals in computing the meaning of a verse. The point here is that any application of a model like EVE’ or IR requires assumptions about the appropriate level at which to model signals and meanings. And those assumptions are essentially semantic, because meanings at one level of abstraction can serve as signals at a higher level of abstraction.

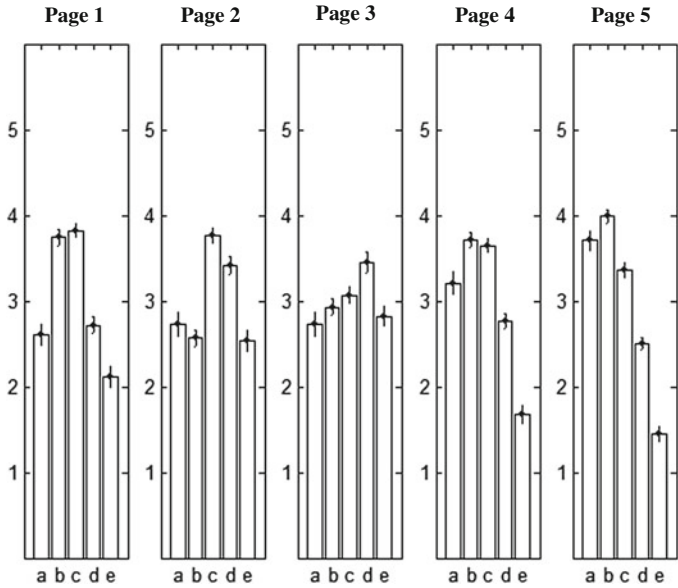


Fig. 2.3 Average human rankings of aesthetic quality, for five panels (*a*, *b*, *c*, *d*, *e*) on each page (1, 2, 3, 4, 5) of the experiment. Error bars represent standard errors of the mean. Data are from $N = 148$ American participants

2.3.2 Personal Preferences

Besides the average rankings of complexity and aesthetics used to test EVE' and other models, individual human judgments from the experiment were used to analyze personal preferences. More specifically, the aesthetically optimal level of visual complexity C_{opt} for each participant was obtained from the rankings provided by that participant. The average value of C_{opt} , over the four pages of black-and-white designs, was then used to characterize a participant's aesthetic preference for visual complexity ranging from $C_{opt} = 1$ to $C_{opt} = 5$.

Figure 2.4 shows the results for C_{opt} across all $N = 148$ Americans that were tested. A total of 11 subjects had $C_{opt} = 1$, which means they always preferred the simplest design on a page. Only one subject had $C_{opt} = 5$. Most subjects had C_{opt} between 2 and 4, and the average value across all subjects was $C_{opt} = 2.76$.

As part of the study, participants were also asked to answer demographic questions about their gender (male or female), age group (<48 years or ≥ 48 years), training (some or no training in art), and taste (like or dislike abstract art). Student's *t*-tests were performed to compare mean values of C_{opt} for each binary distinction. Results showed that none of the demographic variables was a significant predictor of C_{opt} . Gender was the only variable that even approached significance ($p = 0.08$), with females having a higher C_{opt} (mean = 2.90) than males (mean = 2.64).

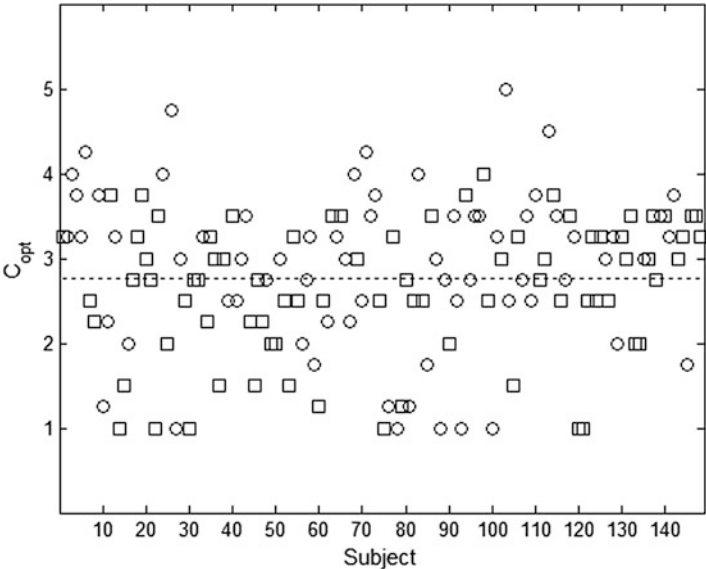


Fig. 2.4 Individual values of C_{opt} , representing each participant’s aesthetically optimal level of visual complexity. C_{opt} is measured on a scale of 1–5, as an average over the first four pages of the experiment. *Squares* denote male subjects and *circles* denote female subjects. Data are from $N = 148$ American participants

These results show that aesthetic preferences for visual complexity are quite varied, even among a culture consisting of all Americans. Moreover, the individual differences are difficult to predict, even by factors such as art taste and training that would be expected to affect a person’s preferences.

2.3.3 Cultural Comparison

After the above experiment with an American population, the same abstract designs were used to test a Japanese population—using verbal instructions translated into their language. Both average and individual results were analyzed.

Figures 2.5 and 2.6 show the average rankings for visual complexity and aesthetic quality across $N = 51$ Japanese participants. Qualitatively, the results are similar to those for $N = 148$ Americans presented in Figs. 2.2 and 2.3. Like the Americans, the Japanese judged visual complexity consistent with visual entropy (V_E) and found aesthetic optimality at an intermediate level of complexity.

Figure 2.7 shows the individual values of C_{opt} for Japanese participants. A comparison to American participants in Fig. 2.4 shows the distributions are similar. Although the mean value of C_{opt} is higher for Japanese (2.93) than for Americans (2.76), a t-test showed this difference was not significant.

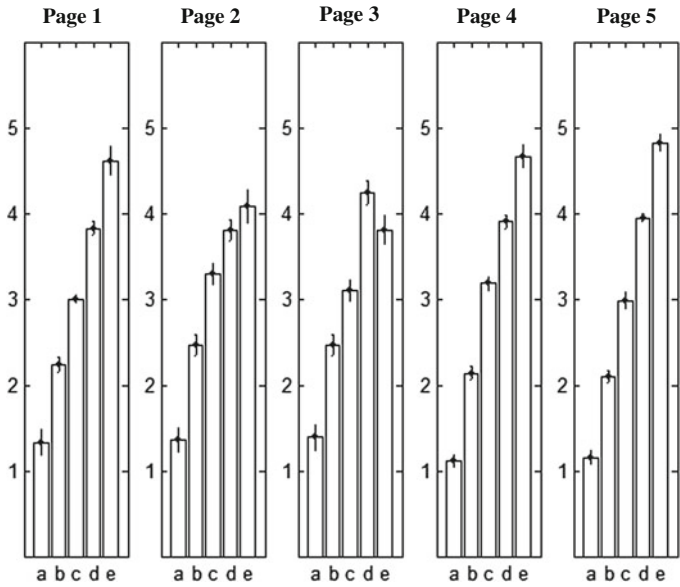


Fig. 2.5 Average human rankings of visual complexity, for five panels (*a, b, c, d, e*) on each page (1, 2, 3, 4, 5) of the experiment. Error bars represent standard errors of the mean. Data are from N = 51 Japanese participants

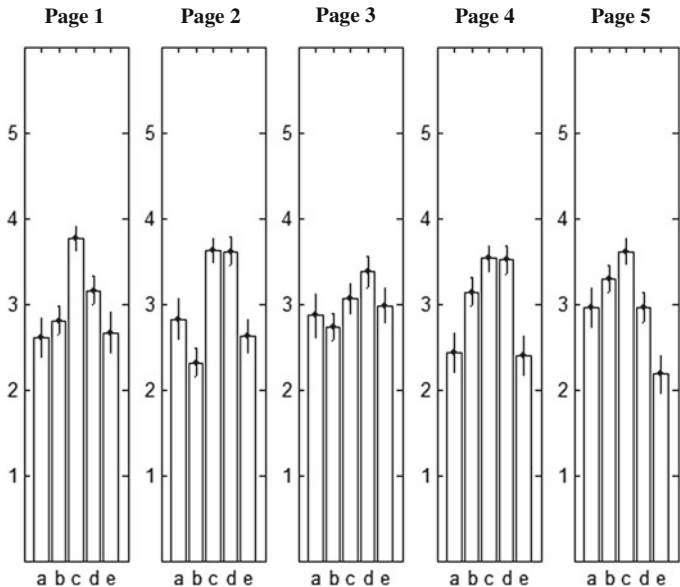


Fig. 2.6 Average human rankings of aesthetic quality, for five panels (*a, b, c, d, e*) on each page (1, 2, 3, 4, 5) of the experiment. Error bars represent standard errors of the mean. Data are from N = 51 Japanese participants

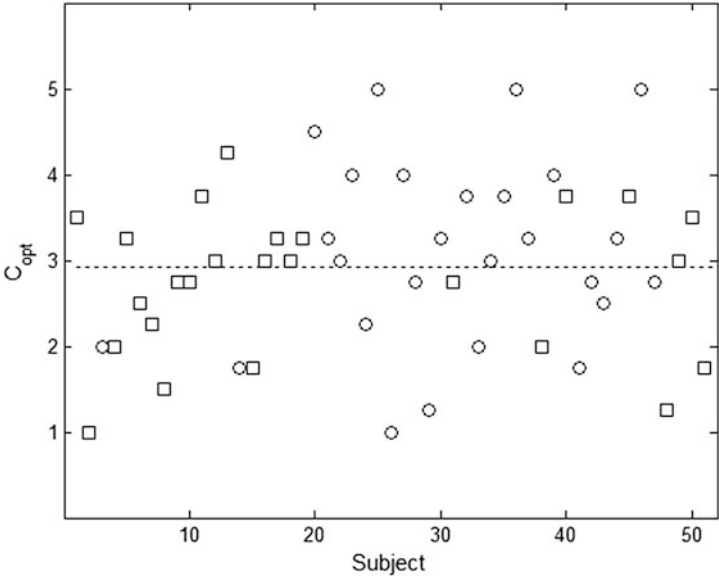


Fig. 2.7 Individual values of C_{opt} , representing each participant's aesthetically optimal level of visual complexity. C_{opt} is measured on a scale of 1–5, as an average over the first four pages of the experiment. *Squares* denote male subjects and *circles* denote female subjects. Data are from $N = 51$ Japanese participants

Finally, the distribution of C_{opt} within the Japanese population of participants was analyzed as a function of the same demographic variables analyzed earlier for Americans. Like the American results, mean C_{opt} was higher for females (3.10) than for males (2.74). But t-tests showed that none of the demographic variables, including gender, was significant.

It should be noted that 47 of the 51 Japanese participants were <28 years old. This demographic differs from the American survey, which included 84 participants <48 years old and 64 participants ≥ 48 years old. However, the American results showed no significant difference due to age.

In sum, the aesthetically preferred level of complexity was higher for Japanese than Americans, and higher for females than males in both cultures, but these differences were not significant. This suggests that aesthetic preferences for complexity in abstract art vary as much within a culture as between the cultures, and that preferences cannot be predicted by demographic variables such as gender, age group, training (some or no training in art), or taste (like or dislike abstract art).

2.4 The Fundamental Challenge of Computing Semantics

In conclusion, this final section of the chapter highlights what has been done and what remains to be done for artificial intelligence to credibly compute aesthetic experiences within and across cultures.

First, what has been done here is to characterize and compare several mathematical models of aesthetics. The models known as IR and EVE' were shown to be essentially equivalent, under the assumption that past signals can explain a present signal received in a temporal sequence of signals. In that case both IR and EVE' compute aesthetics as an incremental reduction in entropy, equal to *how much surprise has been experienced and explained*. This formulation by IR and EVE' was contrasted to earlier equations by Birkhoff and Bense, and tested against human judgments of aesthetics across examples of verbal and visual arts—using participants from American and Japanese populations. The results strongly support IR and EVE' over the equations of Birkhoff and Bense.

In the visual domain of abstract art, empirical testing went further to address the computation of entropy (complexity) needed as input Y to EVE's equation $X = Y * Z$. Results showed that visual entropy V_E was much better than classical entropy H as a model of how humans judge visual complexity. This highlights the importance of modeling signals at levels of abstraction consistent with those of human perception.

Next, what remains to be done for computing aesthetics lies in the problem of computing semantics. IR addresses semantics only implicitly, as signals are explained by past *signals*. EVE' addresses semantics more explicitly, as signals are explained by *meanings* evoked from long-term memory associations. However, EVE' has not encoded the vast human knowledge needed to compute these meanings and their probabilities in any specific domain.

Instead EVE' has either asked humans for their judgments of meaning (in a verbal domain); or else analyzed abstract artworks for which meaning could be modeled as a simple function of signal features (in a visual domain). This was done in the spirit of what Scha and Bod [40] suggested 20 years ago, as follows: “*For a specific, narrowly defined class of inputs (such as line drawings or grids), such a process-model [of aesthetic perception] might be worked out. But it would be absolutely out of the question to accomplish this in the context of a complete simulation of all possibilities of human visual perception. Things get even more difficult when we introduce the semantic dimension...*”

The value of models like IR and EVE' was also foreshadowed by Scha and Bod [40] as follows: “*For the time being, we cannot work out such a semantic model in any detail. But it will become more concretely imaginable as soon as a very limited purely syntactic model would show interesting results. Thus, the ultimate benefit of the computational approach to the esthetic will not lie in the models that can be implemented and validated—but in the more speculative and encompassing models which they make thinkable.*” Experiments with IR [18, 19] and EVE' [12, 13] have shown interesting and valid results—i.e., more consistent with human judgments

than competing models by Birkhoff, Bense, and their followers in the field of information aesthetics [32, 37, 40]. Thus IR and EVE' may help make computational models of aesthetics more “thinkable” by artificial intelligence.

In particular, both IR and EVE' suggest that aesthetics are governed universally by *how much surprise is experienced and explained*. Mathematically, this can be modeled as a reduction in entropy, similar to “work” done in the realm of a thermodynamic system. But psychologically, which is the realm in which art “works”, the amount of surprise actually experienced and explained will be determined by meanings that exist in the minds of audiences. These semantics are especially important for addressing cultural differences, because common knowledge and value structures are what underlie aesthetic agreements within a culture—and hence what distinguish its artworks and aesthetics from those of other cultures. Therefore, future advances in computing aesthetics will require a shift in the focus of artificial intelligence—from computing *information in signals* of artworks to computing *explanations of meanings* by audiences.

Acknowledgments We would like to acknowledge the work by Mr. Shunsuke Hananoi from Keio University who kindly translated the survey and collected responses from Japanese participants.

References

1. Addiss S (2012) The art of haiku: its history through poems and paintings by Japanese masters. Shambhala, Boston
2. Addiss S, Yamamoto F, Yamamoto A (2007) Haiku humor: wit and folly in Japanese poems and prints. Weatherhill, Boston
3. Arnheim R (1971) Entropy and art: an essay on order and disorder. University of California Press, Berkeley
4. Baddeley A (1992) Working memory. *Science* 255:556–569
5. Bayes T (1763) An essay toward solving a problem in the doctrine of chances. *Philos Trans* 53:370–418
6. Bense M (1965) Aesthetica, einföhrung in die neue aesthetik. Agis-Verlag, Baden-Baden
7. Berlyne D (1971) Aesthetics and psychobiology. Appleton Century Crofts, New York
8. Birkhoff G (1933) Aesthetic measure. Harvard University Press, Cambridge
9. Blinderman C (1962) T. H. Huxley's theory of aesthetics: unity in diversity. *J Aesthetics Art Criticism* 21:49–55
10. Burns K (2006) Atoms of EVE': a Bayesian basis for esthetic analysis of style in sketching. *Artif Intell Eng Des Anal Manuf* 20:185–199
11. Burns K (2012) EVE's energy in aesthetic experience: a Bayesian basis for haiku humor. *J Math Arts* 6:77–87
12. Burns K (2015) Entropy and optimality in abstract art: an empirical test of visual aesthetics. *J Math Arts* 9:77–90
13. Burns K (2015) Computing the creativeness of amusing advertisements: a Bayesian model of Burma-Shave's muse. *Artif Intell Eng Des Anal Manuf* 29:109–128
14. Burns K, Dubnov S (2006). Memex music and gambling games: EVE's take on lucky number 13. Papers from the AAAI workshop on computational aesthetics, WS-06-04. AAAI Press, Menlo Park, pp 30–36

15. Cover T, Thomas J (1991) Elements of information theory. Wiley, New York
16. Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24:87–114
17. Dubnov S (2003) Non-Gaussian source-filter and independent components generalizations of spectral flatness measure. In: Proceedings of the international conference on independent components analysis, pp 143–148
18. Dubnov S (2010) Information dynamics and aspects of musical perception. In: Argamon S, Burns K, Dubnov S (eds) The structure of style: algorithmic approaches to understanding manner and meaning. Springer, Berlin, pp 127–157
19. Dubnov S, McAdams S, Reynolds R (2006) Structural and affective aspects of music from statistical audio signal analysis. *J Am Soc Inform Sci Technol* 57(11):1526–1536
20. Ekárt A, Joó A, Sharma D, Chalakov S (2012) Modeling the underlying principles of human aesthetic preference in evolutionary art. *J Math Arts* 6:107–124
21. Ellsworth P, Scherer K (2003) Appraisal processes in emotion. In: Davidson R, Scherer K, Goldsmith H (eds) Handbook of affective sciences. Oxford University Press, Oxford, pp 572–595
22. Eysenck H (1941) The empirical determination of an aesthetic formula. *Psychol Rev* 48:83–92
23. Eysenck H (1942) The experimental study of ‘good gestalt’. *Psychol Rev* 49:344–364
24. Eysenck H (1957) Sense and nonsense in psychology. Penguin, Harmondsworth
25. Fry R (1920) Vision and design. Chatto and Windus, London
26. Higginson W, Harter P (1985) The haiku handbook: how to write, teach, and appreciate haiku. Kodansha, New York
27. Hutcheson F (1729) An inquiry into the original of our ideas of beauty and virtue. J and J Knapton
28. Huxley T (1882) On science and art in relation to education. Collected essays Volume III
29. Kreidler H, Kreidler S (1972) Psychology of the arts. Duke University Press, Durham
30. McGrayne S (2011) The theory that would not die: how Bayes rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy. Yale University Press, New Haven
31. Moles A (1966) Information theory and esthetic perception. University of Illinois Press, Urbana
32. Nake F (2012) Information aesthetics: an heroic experiment. *J Math Arts* 6:65–75
33. Oatley K (2003) Creative expression and communication of emotions in the visual and narrative arts. In: Davidson R, Scherer K, Goldsmith H (eds) Handbook of affective sciences. Oxford University Press, Oxford, pp 481–502
34. Parsons M (1987) How we understand art: a cognitive developmental account of aesthetic experience. Cambridge University Press, Cambridge
35. Phillips F, Norman J, Beers A (2010) Fechner’s aesthetics revisited. *Seeing and Perceiving* 23:263–271
36. Reichhold J (2002) Writing and enjoying haiku: a hands-on guide. Kodansha, Tokyo
37. Rigau J, Feixas M, Sbert M (2008) Informational aesthetics measures. *IEEE Comput Graphics Appl* 28:24–34
38. Ross B (2002) How to haiku: a writer’s guide to haiku and related forms. Tuttle, Tokyo
39. Rowsome F (1965) The verse by the side of the road: the story of Burma-Shave signs and jingles with all 600 of the roadside rhymes. Plume, New York
40. Scha R, Bod R (1993) Computacionele esthetica. *Informatie en Informatiebeleid* 11(1):54–63. <http://iaaa.nl/rs/compestE.html>. Accessed 11 June 2014
41. Scherer K (1999) Appraisal theory. In: Dalglish T, Power M (eds) Handbook of cognition and emotion. Wiley, New York, pp 637–663
42. Shannon C, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

43. Silvia P (2005) Emotional responses to art: from collation and arousal to cognition and emotion. *Rev Gen Psychol* 9:342–357
44. Silvia P (2006) *Exploring the psychology of interest*. Oxford University Press, New York
45. Stiny G, Gips J (1978) *Algorithmic aesthetics: computer models for criticism and design in the arts*. University of California Press, Berkeley
46. Yasuda K (1957) *The Japanese haiku: its essential nature, history, and possibilities in English*. Tuttle, Boston

Cross-Cultural Multimedia Computing

Semantic and Aesthetic Modeling

Dubnov, S.; Burns, K.; Kiyoki, Y.

2016, XII, 62 p. 31 illus., Softcover

ISBN: 978-3-319-42871-0