EVE's Energy in Aesthetic Experience: A Bayesian Basis for Haiku Humor

Kevin Burns

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

EVE' is a mathematical model of aesthetic experience, founded on Bayesian probability and grounded in cognitive psychology. The model addresses three stages of cognitive processing: starting with expectations (*E*) of signals, which produce pleasure; followed by violations (*V*) of expectations, which produce tension; and ending with explanations (*E*') of meaning, which produce pleasure-prime. This creates a tradeoff to be optimized in the design of artworks, because an audience can only obtain pleasure-prime from *E*' at the expense of pleasure from *E*. A total measure of aesthetic pleasure *X* is derived as the product of two terms, $X = Y * Z$, where *Y* is a measure of entropy in violations, and *Z* is a measure of how completely entropy is converted to energy in explanations. The model is applied to the art of humor in haiku form, by evaluating one poem in detail and by generating additional poems as exemplars.

Keywords: entropy, energy, expectation, explanation, information, emotion, haiku, humor

AMS Subject Classification: 62C10; 62F15; 91E10; 91E40; 94A15; 94A17

1. Introduction

All approaches to aesthetic evaluation imply some sort of optimization. Mathematically, there must be a set of variables that combine to produce peaks in a space of aesthetic responses, which will depend on the medium and vary from culture to culture as well as from person to person. But cutting across media specifics and individual differences are universal principles that apply to all art [15]. My aim in this article is to uncover some of those universals by dissecting one art form in detail – i.e., humor composed as haiku verse.

The aesthetic of haiku is often described as an "aha moment" of surprise, insight, and delight [30, 31]. This is similar in many respects to the way jokes work, and in fact the "hai" of haiku originally meant "playful" or "humorous". In Japanese, the term "haiku" is now reserved for serious poems about nature, whereas humorous poems about humans are called "senryu" – such as the following classic [31]:

she has just one eye
but it is a pretty one
says the matchmaker

All poems of this form written in English are properly referred to as haiku, and the writers of such poems often employ puns, wordplay, and other techniques of humor – see Reichhold's *Writing and Enjoying Haiku* [30]. Of course the delight experienced in a "haiku moment" is not always humorous. Likewise, the EVE' model I present applies in a broader context to other aesthetics and other artworks such as sketching [9], music [10], and gaming [8]. However, in this article I will focus on the art of haiku humor in order to illustrate the main elements of the model.

1.1 Order and Complexity

An early equation for aesthetics that applies across media was developed by Birkoff [6] in the 1920s. His famous formula is: $M = O / C$, where M is a measure of aesthetic value, O is order, and C is complexity. This formula captures notions of optimality, symmetry, and complexity that are ubiquitous among mathematical models of art. But Birkhoff's equation is inadequate because its variables are static and deterministic – whereas aesthetics are *experienced* [14, 29, 33] in a dynamic and probabilistic context as meaning and feelings unfold in the mind of an audience. As a result, it is not surprising that psychologists have found little support for the aesthetic measure $M = O / C$ in formal testing of human subjects [22]. In fact, based on dozens of experiments involving graphics, poems, vases, and music, Eysenck [16] concluded that aesthetics are positively correlated with complexity and order – such that a better equation would be $M = O * C$.

The invention of information theory in the 1940s led to new approaches based on the probabilistic concept of *entropy* [24]. Entropy is a formal measure of unpredictability, which relates to both order and complexity. One model by Bense [4] in the 1960s assumed complexity was equal to entropy (Y), and order was equal to a reduction in entropy ($Y - Y'$), such that Birkhoff's formula could be rewritten as $M = O / C = (Y - Y') / Y$. This equation implies a before-and-after dynamic, and reflects the probabilistic nature of information processing, but it still does not address *how* the mind reduces entropy, or *why* the brain produces aesthetics when entropy is reduced.

Starting in the 1980s and continuing to the present day, advances in technology have led to widespread interest in aesthetic computing [17]. The vast majority of this work is aimed at *generating* artworks, typically via algorithms written by the artists sometimes known as algorists [40]. But generating art requires a corresponding capability for *evaluating* the aesthetics of what is or might be generated [1].

The problem of aesthetic evaluation is difficult because it must address psychological workings of the mind – not just mathematical features of artworks or the computational algorithms by which those artworks might be generated. As a practical matter, aesthetic evaluation requires representations of the human knowledge and value structures by which cognitive appraisals [34] give rise to emotions [26]. This is because artworks themselves are merely *signals*, whereas aesthetics are *feelings* that arise as those signals are given *meaning* by an audience [27]. The basic problem was outlined over half a century ago in the introduction to Shannon's [35] seminal paper on information theory, where Weaver [39] identified the following three levels of communication: *"How accurately can the symbols of communication be transmitted? How precisely do the transmitted symbols convey the desired meaning? How effectively does the received meaning affect conduct in the desired way?"*

Shannon himself addressed only the first level, and yet clearly communication is much more than just the transmission of symbols. But one concept from information theory that does give a glimpse into the deeper levels is *surprise* [20]. As formulated by Shannon, the amount of information that a signal (s) carries is given by its *surprisal* and measured in "bits" by marginal *entropy* (Y). This quantity is defined by the function $Y = -lg\ P(s)$ plotted in Figure 1a, where lg denotes the logarithm to base 2, and $P(s)$ denotes the probability of receiving signal s. Similar to the notion of surprisal, some psychologists have suggested that *arousal* is the driving force behind aesthetics. A leader in this vein was Berlyne [5], who in the 1960s proposed that arousal potential increases as a stimulus becomes more complex, and that hedonic value (akin to aesthetic measure) can be modeled as the interaction between competing neural systems (positive and negative) that are activated by the arousal. But Berlyne is still missing a model of what happens after arousal, and especially how emotions and aesthetics arise from appraisals of meaning [36]. So the question remains: *What are the psychological mechanisms that give rise to an aesthetic experience, and how can those fundamental mechanisms be specified in a mathematical formulation?*

1.2 Meaning and Probability

The first part of an answer comes from a paper by the Reverend Thomas Bayes published in the mid 1700s [3]. Although Bayes Rule [23] was never intended to address aesthetics or any other psychological process, it can help explain how humans find meaning in signals – i.e., by reasoning about the conditional probabilities $P(H|e)$ of hypothetical *meanings* (H) in light of evidential *signals* (e). In words, $P(H|e)$ refers to the *posterior* probability of a hypothesis given some evidence. This is different from the probability $P(H)$ of a hypothesis before receiving the evidence, which Bayesians call the *prior* probability. And both are different from the conditional probability $P(e|H)$ of evidence assuming the hypothesis is true, which Bayesians call the *likelihood* function.

Bayes Rule can be derived simply by equating two expressions for the joint probability of H and e, denoted $P(e,H)$ or $P(H,e)$ as follows: $P(e,H) = P(e) * P(H|e) = P(H) * P(e|H) = P(H,e)$. Rearranging yields the final form in which the result is often expressed:

$$P(H|e) = P(H) * P(e|H) / P(e).$$

In words, the *posterior* probability $P(H|e)$ of a hypothesis H given evidence e, is proportional to the product of the *prior* probability $P(H)$ and the *likelihood* $P(e|H)$ of receiving the evidence assuming the hypothesis is true. The product $P(H_i) * P(e|H_i)$ for each H_i in a set of hypotheses $\{H_i\}$ is divided by the sum of the products, $P(e) = \Sigma_i P(H_i) * P(e|H_i)$, to obtain posterior probabilities that sum to 1 over the set.

The distinction between $P(H|e)$ and $P(e|H)$ is important not only mathematically but also psychologically, as will be discussed further in Section 3. For now the main point is that *likelihoods* of the form $P(e|H)$ are stored in long-term memory, as the associative strengths between various meanings (H) and the signals (e) that those meanings are likely to produce. On the other hand, *posteriors* of the form $P(H|e)$ are constructed in working memory, which is extremely limited. The capacity of working memory is known to be only a handful of "chunks" [2, 13], so only a few hypotheses can be retained at once while reasoning about the associated priors, likelihoods, and posteriors. As we will see later, it is the richness of likelihoods in our long-term memories, together with constraints on priors and posteriors in our working memories that enable us as humans to have the aesthetic experiences that we do. Thus the claim here is not that humans are perfect Bayesians – but rather we are bounded Bayesians, and our bounds constrain the art that we create and consume [21].

2. Formulation

By adopting a Bayesian approach, aesthetic evaluation can move beyond the entropy of *signals* to address *meaning* and *feelings* at deeper levels of processing. Below I outline my model, dubbed EVE', whose name refers to cognitive processes of expectation (E), violation (V), and explanation (E'). The model is shown to compute a measure of aesthetic energy (X) as a product of entropy (Y) and a Bayesian function (Z), which satisfy $X = Y * Z$.

2.1 Pleasure and Pleasure-Prime

According to the model of EVE' [9], aesthetics arise from subjective *success* in expecting signals and explaining meaning. An underlying assumption is that the brain would have evolved to reward itself with pleasure [7, 15] for these successes – simply because the same skills that lead to such successes would help an animal survive in the world. Thus the model considers two levels of reward: First, there is pleasure from success in *expecting* signals. Second, there is pleasure-prime from success in *explaining* signals that were not expected, i.e., by finding meaning at a deeper level.

More specifically, if a signal s is expected at probability $P(s)$, then success in expectation (E) can be measured by the lack of surprise when s is received, as follows: $E = lg\ P(s)$. This is illustrated by the lower dotted curve in Figure 1b. Conversely, failure in predicting the signal s would be measured by Shannon's surprisal, as follows: $V = -lg\ P(s)$, where V represents the violation of expectation when s is received. This is illustrated by the upper dotted curve in Figure 1b.

Taken together, the measure of success along with an opposite measure of failure would cancel each other out, if the pleasure from a unit of success was equal in magnitude to the displeasure from a unit of failure. But according to the EVE' model, a failure of expectation at the level of signals creates an opportunity for explanation at the deeper level of meaning. In effect, the failure is not felt directly as displeasure, but rather as tension – which can then be converted to pleasure-prime if and when the tension is resolved. Thus a violation (V) sets up the potential for success in explanation (E') as follows: $E' = R * V = R * -lg\ P(s)$, where the resolution ($R$) is a fraction that ranges from 0 to 1. This resolution is modeled as the Bayesian posterior for the most likely meaning (M) of the signal (s) that was received, $R = P(M|s)$, and hereafter R is written as $R(M|s)$ to highlight this dependence on both M and s.

Finally, the EVE' model expresses how expectations of *signals* and explanations of *meaning* combine to produce pleasurable *feelings* at the deepest level, i.e., by the weighted sum of successes in expectation (E) and explanation (E') as follows:

$$X = (G * E) + c + (G' * E').$$

Here the constants G and G' scale the amount of pleasure and pleasure-prime resulting from a unit of E and E', respectively, as the two combine in the aesthetic measure X. The constant c represents the pleasure obtained when a signal is perfectly predictable i.e., when $P(s) = 1$.

The equation is illustrated in Figure 1b for the case where $R(M|s) = 1$, assuming $G' = 3$, $G = 1$, and $c = 1$. These are generic values for G', G, and c, taken from previous studies [8, 9, 10] for purposes of illustrating the formula here. As will be discussed in Section 4, specific values for each parameter may vary between individuals, and may also vary with the signal (s) and meaning (M) even for the same individual. The value of $R(M|s)$ will vary with the signal (s) and meaning (M). Figure 1c plots X versus $P(s)$ for various values of $R(M|s)$ ranging from 0.1 to 0.9, assuming the generic values $G' = 3$, $G = 1$, and $c = 1$.

2.2 Entropy and Energy

The above equation for aesthetic measure X can be reformulated by substituting the earlier expressions for E and E', and then rearranging terms to obtain the following:

$$X = [-lg\ P(s)] * [G' * R(M|s) - G] + c = Y * Z + c.$$

Thus the result can be expressed (to within a constant c) simply as $X = Y * Z$, where X is the product of marginal entropy Y and a Bayesian factor Z. In this formulation, the Bayesian factor Z transforms the measure of entropy Y to a quantity X that measures how well the art "works". Thus by analogy to similar concepts in statistical mechanics, X can be characterized as a capacity for doing work – i.e., *energy*.

A similar conception of mental energy appears in many psychological theories of aesthetics – including a "Fundamental Law of Aesthetics" proposed by Eysenck [16], as well as Freud's psychoanalytic treatment of dreams and jokes [18]. The EVE' model adopts this notion of energy in referring to the pleasurable feelings that an audience experiences when art "works". However, no physical or neurological systems are explicitly modeled by the above equations for *entropy* or *energy*. Instead these terms refer only to informational and psychological concepts, which are applied here in the mathematical modeling of aesthetic experience.

To recap, the EVE' model of aesthetic experience considers arousal from marginal entropy as well as the cognitive appraisal by which that entropy (measured in bits) is converted to energy (also measured in bits). The combination is shown in Figure 1c, which offers a key insight not suggested by the models of Birkhoff, Bense, Berlyne, or others. That is: large entropy, which occurs at small $P(s)$, holds the potential for the most positive aesthetic experience as well as the most negative aesthetic experience, depending on the Bayesian resolution $R(M|s)$ by which entropy is converted to energy. When $R(M|s)$ is high then large entropy leads to very positive X, and when $R(M|s)$ is low then large entropy leads to negative X. This is because the energy X is governed by a tradeoff between how well signals were expected and how well meaning is explained, with the strongest positive experiences coming when signals were not well-expected but meaning is well-explained.

3. Evaluation

As one application of the above equation, the EVE' model is used here to analyze a joke written in haiku form. The format and phrasing of haiku serve to illustrate the *E*, *V*, and *E'* of the model, which also apply beyond haiku to a wide range of other artworks [8, 9, 10]. Within the constraints of the haiku form, I chose to focus on humor for three reasons, as follows:

First, humor is an important aesthetic that is often evoked by artists working in verbal, visual, and musical modalities. Second, compared to other aesthetics such as "beauty", humor is more directly measurable by behavioral responses such as a smile or laugh if the humor succeeds (or a frown or groan if the humor fails). Finally, humor highlights the role of *probability* in poetry, as espoused by Aristotle and formalized by my EVE' model. In his *Poetics* [12], which is arguably the single-most influential work in the philosophy of aesthetics, Aristotle writes (IX.3-5): *"... for poetry tends to express the universal... by the universal I mean how a person of a certain type will on occasion speak or act, according to the law of probability... in comedy this is already apparent: for here the poet first constructs the plot on the lines of probability, and then inserts characteristic names..."*.

When written in English, haiku typically contain three lines with 5, 7, and 5 syllables, respectively [31]. One subtle but important feature of these verses is that the three lines are grouped into two phrases [30], where the first two (or last two) lines comprise one phrase. This grouping creates a short pause to be felt between the two phrases, which typically punctuates an "aha" (surprise) – sometimes referred to as a "moment" [31] of insight and delight. A similar grouping of units in a so-called *"AAB pattern"* [32] is often found in music and humor, and for this reason haiku, especially when used to convey humor, can serve as a well-structured medium in which to analyze how EVE' works. The example I wish to consider is the following:

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

The first line of this haiku mimics the first few words of a well-known proverb, and thereby establishes expectations for the next line of the poem. When the second line is read it is consistent with the proverbial meaning, i.e., that apples are healthy and hence good to eat. The final line then introduces an incongruity, which is surprising. That is, in the Garden of Eden it was bad for Adam and Eve to eat apples from the Tree of Knowledge, and this meaning is opposite to the original meaning that apples are good to eat.

After the violation, additional reasoning is required to find an adequate explanation. In this case a savvy reader will realize that eating apples is not always good, because it depends on the context of the situation – and any advice on the matter of apples and eating, offered by another person, depends on the person's intentions. Thus such a reader will see that the devil is using the proverb as a trick to tempt Eve, who is naive and (as we know from the bible story) succumbs to temptation. A reader who reaches this meaning will get the joke, and at the same time feel like he himself would not fall prey to the devil's tempting because he knows the devil's intent. So the reader will feel empathy and superiority.

Feelings of empathy [26] and superiority [19] play a key role in any story and especially humor. Empathy makes the outcome feel relevant, like it could happen to you. Superiority makes you feel dominant, like it would not happen to you. The combination enables you to feel like you have met a challenge and succeeded in conquering it, and that is exactly the sort of imaginative success that should be rewarded with pleasure [7] by a brain that is adapting itself for survival [15]. In the case of humor, this reward is often accompanied by a release of energy in the form of a smile or snicker or laugh out loud.

The remainder of this section dissects the above haiku in detail, line by line, to identify and quantify the signals (s), meanings (M), and probabilities $P(s)$ and $R(M|s) = P(M|s)$ needed to compute energy X per the EVE' model. The analysis will continue to distinguish between *signals* and *meanings* by use of lower case and upper case notation. All signals will be denoted by lower case letters such as a_1, a_2, and b_3, where the letter refers to a class of signal (a or b) and the subscript refers to a line of the poem in which the signal appears (1, 2, or 3). Meanings will be denoted by upper case letters such as A or B.

3.1 The First Line: a_1 = an apple a day

To begin, the initial meaning of the poem is that apples are good to eat and the person who says so is being truthful. This meaning, denoted A, is inferred when the signal a_1 is recognized as repeating the first few words of a well-known proverb. The signal a_1 might suggest some other meanings, besides A, but the match to the proverb is so strong that no other meaning stands out. On that basis we can assume a reader mentally represents just two possible meanings in her initial set of hypotheses $\{A, \sim A\}$, where $\sim A$ denotes "not-A" and includes all unspecified hypotheses that are not A. Thus at this stage, where the probabilities of A and $\sim A$ are conditioned only on signal a_1, we can write the reader's beliefs as follows:

$$P(A|a_1) = \eta$$

$$P(\sim A|a_1) = 1 - P(A|a_1) = 1 - \eta$$

where η represents a moderate probability, because there are many unspecified hypotheses included in $\sim A$. For example $\eta \approx 0.5$ would reflect the belief that the proverbial meaning (A) is about as likely as all other possible meanings combined ($\sim A$), such that the reader's beliefs at this point can be written as $P(A|a_1) \approx P(\sim A|a_1) \approx 0.5$.

Now after reading the signal a_1, expectations for the next signal will be formed in working memory. These expectations will depend on the reader's current beliefs, $P(A|a_1)$ and $P(\sim A|a_1)$, as well as on the likelihoods that the hypothesized meanings in the set $\{A, \sim A\}$ would cause various possible signals in a set $\{a_2, \sim a_2\}$. Here a_2 refers to a signal in class a that is consistent with meaning A, whereas $\sim a_2$ refers to a signal in class $\sim a$ that is consistent with meaning $\sim A$. The two meanings are mutually exclusive, so a signal consistent with one meaning is unlikely to be (but could still possibly be) consistent with the other meaning. Thus the likelihoods can be written as follows:

$$P(a_2|A) = P(\sim a_2|\sim A) = \delta$$

$$P(a_2|\sim A) = P(\sim a_2|A) = \varepsilon$$

where $\delta \approx 1$ and $\varepsilon \approx 0$. Here δ and ε denote approximate (order-of-magnitude) values for probabilities, i.e., δ is a very high probability (but less than 1.0) and ε is a very low probability (but greater than 0.0). These same symbols will be used below to denote probabilities that are very high or very low without regard to the exact values, for two reasons. First, a reader of the poem would probably only represent approximate values for likelihoods in her mind. Second, only approximate estimates are needed to explain and predict the basic workings of the EVE' model.

The likelihoods noted above apply to the signals a_2 or $\sim a_2$ that the reader is expecting as possible signals in the next line of the poem. Together with these conditional likelihoods, a reader will use her current beliefs (formed after receiving a_1) about $P(A|a_1)$ and $P(\sim A|a_1)$ to compute the marginal probability of receiving a_2 or $\sim a_2$ in the next line of the poem, and thereby establish expectations as follows:

$$P(a_2) = P(A|a_1) * P(a_2|A) + P(\sim A|a_1) * P(a_2|\sim A) = \eta * \delta + (1 - \eta) * \varepsilon \approx \eta$$

$$P(\sim a_2) = P(A|a_1) * P(\sim a_2|A) + P(\sim A|a_1) * P(\sim a_2|\sim A) = \eta * \varepsilon + (1 - \eta) * \delta \approx 1 - \eta.$$

Because $\eta \approx 1 - \eta \approx 0.5$, these equations imply that $P(a_2) \approx 0.5$ and $P(\sim a_2) \approx 0.5$. In words, after receiving signal a_1 in the first line, the reader is expecting the signal in the next line to be either a_2 or $\sim a_2$, each with approximately equal probability.

3.2 The Next Line: a_2 = will keep the doctor away

After reading a_2, which is a signal of class a (consistent with meaning A) in line 2, the reader experiences some measure of success in expectation because a_2 was expected at probability $P(a_2) \approx \eta \approx 0.5$. But at the same time there is some measure of failure in expectation, which is a violation, because $\sim a_2$ was also expected at probability $P(\sim a_2) \approx 1 - \eta \approx 0.5$. According to the EVE' model, the success brings some pleasure while the failure creates tension, which fuels further effort to find an explanation. Per Bayes Rule, the explanation comes from updated (posterior) beliefs about the probabilities of hypotheses A and $\sim A$ in light of the new evidence a_2 (along with previous evidence a_1), as follows:

$$P(A|a_2,a_1) = P(A|a_1) * P(a_2|A) / P(a_2)$$

$$P(\sim A|a_2,a_1) = P(\sim A|a_1) * P(a_2|\sim A) / P(a_2)$$

where $P(a_2) = P(A|a_1) * P(a_2|A) + P(\sim A|a_1) * P(a_2|\sim A)$. Using Greek letters to represent the approximate probabilities:

$$P(A|a_2,a_1) = \eta * \delta / (\eta * \delta + (1 - \eta) * \varepsilon) \approx \delta$$

$$P(\sim A|a_2,a_1) = (1 - \eta) * \varepsilon / (\eta * \delta + (1 - \eta) * \varepsilon) \approx \varepsilon.$$

In words, after reading the signal a_2 in the second line (which in turn was read after the signal a_1 in the first line), the reader now has a very strong belief in A over $\sim A$, i.e., $P(A|a_2,a_1) >> P(\sim A|a_2,a_1)$ because $\delta >> \varepsilon$. The measure of resolution $R(M|s)$ for the meaning (M) of the complete signal ($s = a_2,a_1$) at this point is given by the posterior probability of the most probable hypothesis. So $R(M|s) = P(A|a_2,a_1) \approx \delta \approx 1$, and the resolution of the violation is nearly complete.

Using the above value for $P(s) = P(a_2) = \eta$, and $R(M|s) = P(A|a_2,a_1) \approx \delta$, the aesthetic measure X_2 (after the second line of the poem) is computed as follows based on the energy equation in Section 2.2:

$$X_2 = [-lg\ \eta] * [G' * \delta - G] + c \approx [-lg\ \eta] * [G' - G] + c.$$

For example, assuming the generic values $G' = 3$, $G = 1$, and $c = 1$, and assuming $\eta \approx 0.50 = 1/2^1$, we obtain $X_2 = 1 * (3 - 1) + 1 = 3$. Thus at this point the reader has experienced $X_2 = 3$ bits of positive energy. Also at this point, the reader forms new expectations for the signal to be read in the last line of the poem, using the posteriors $P(A|a_2,a_1)$ and $P(\sim A|a_2,a_1)$ from above as the new priors along with the conditional likelihoods of signals a_3 and $\sim a_3$. These conditional likelihoods are the same as those noted above for signals a_2 and $\sim a_2$, respectively. Thus the new priors and conditional likelihoods are used to compute to expectations for a_3 and $\sim a_3$ as follows:

$$P(a_3) = P(A|a_2,a_1) * P(a_3|A) + P(\sim A|a_2,a_1) * P(a_3|\sim A) = \delta * \delta + \varepsilon * \varepsilon \approx \delta$$

$$P(\sim a_3) = P(A|a_2,a_1) * P(\sim a_3|A) + P(\sim A|a_2,a_1) * P(\sim a_3|\sim A) = \delta * \varepsilon + \varepsilon * \delta \approx \varepsilon.$$

Here $P(a_3) >> P(\sim a_3)$ because $\delta >> \varepsilon$, so the reader strongly expects to read a_3 rather than $\sim a_3$ in the last line of the poem.

Of course at a higher level of reasoning the reader is actually expecting to read $\sim a_3$ rather than a_3, because she knows or at least suspects the haiku is a joke with only three lines, and she knows at least implicitly that $\sim a_3$ would be required to make it a joke. But apparently readers suspend this belief and pretend that $\sim a_3$ is unlikely, as a sort of convention between artists and audiences needed to make jokes work. This playful willingness to suspend belief seems central to a sense of humor. A similar suspension of belief is involved in repetition [25], when an artist/audience find a joke funny even after telling/hearing it over and over again – and the same would be true of all artwork that people enjoy even though they have already seen or heard it before.

3.3 The Last Line: b_3 = said devil to Eve

When the signal b_3 is received, the reader experiences a large violation of expectation because a_3 was highly expected and b_3 is in class $\sim a$ rather than class a. Of course b_3 is just one of many signals in the set $\{b_3, c_3, d_3, e_3, f_3, g_3,...\}$ of all possible $\sim a_3$ signals, and any single signal in this large but finite set has a probability much less than $P(\sim a_3)$. But as described above, a reader could not [2, 13] hence would not be representing all of these possibilities in working memory as she starts reading the last line. So when the signal b_3 is received, the violation of expectation would be measured as $-lg\ P(\sim a_3) = -lg\ \varepsilon$.

This large entropy produces high arousal, which holds potential for large energy from cognitive appraisal. It is here in the search for meaning that the art succeeds or fails, as the audience performs a creative leap much like that of the artist who wrote the joke in the first place. The leap is to "abduct" [38] a new meaning B that is consistent with the entire signal and at the same time consistent with the earlier portion of the signal. That is, B must not only explain the last line (b_3) but also the first lines (a_1,a_2) that were already explained by A, in order for B to explain the whole poem (a_1,a_2,b_3).

To resolve the violation, a set $\{H_i\}$ of hypothetical explanations would be activated from long-term memory and represented in working memory, based on the likelihoods $P(b_3,a_2,a_1|H_i)$ that each hypothesis (H_i) would cause the collective evidence (b_3,a_2,a_1). These likelihoods represent the associative strengths between various meanings and the signals they are likely to produce. Only a relatively small set of hypotheses with the highest likelihoods could feasibly be represented in working memory [2, 13], perhaps a couple or at most a handful. Along with each H_i, working memory would also represent corresponding values for each prior $P(H_i)$ and likelihood $P(b_3,a_2,a_1|H_i)$, as needed to compute $P(H_i|b_3,a_2,a_1)$. Bayes Rule then enables computation of the posterior $P(H_i|b_3,a_2,a_1)$ for each hypothesis H_i, as a measure of how well each H_i explains the collective evidence (b_3,a_2,a_1), as follows:

$$P(H_i|b_3,a_2,a_1) = P(H_i) * P(b_3,a_2,a_1|H_i) / \Sigma_i\ P(H_i) * P(b_3,a_2,a_1|H_i)$$

where the sum is taken over the set of hypotheses $\{H_i\}$ being maintained in working memory. Per the EVE' model, the H_i with highest $P(H_i|b_3,a_2,a_1)$ would represent the best explanation, and the magnitude of resolution $R(M|s)$ would be measured by this highest $P(H_i|b_3,a_2,a_1)$.

After receiving the signal b_3 = *said devil to Eve*, a reader who "gets it" will have abducted a new meaning: $H_i = B$ = apples are bad to eat in the Garden of Eden and the devil is tempting Eve with a proverbial truism about apples being good to eat. This reader will thus recognize that the proverb has been twisted to suit the devil's intent. The effect is much like the notion of "reversal and recognition" that Aristotle argues, in *Poetics* [12], is central to art and aesthetics – also see *Mathematics and Humor* [28].

To finish the analysis, we can assume that the resolution of violation is nearly complete, because B is the only hypothesis in working memory that explains the meaning of the collective evidence (b_3,a_2,a_1), i.e., $P(B|b_3,a_2,a_1) \approx 1$ so $R(M|s) \approx 1$. Thus the measure of energy X_3 produced by this last line of the poem is as follows:

$$X_3 \approx [-lg\ \varepsilon] * [G' - G] + c.$$

Using the same values as above for G' = 3, G = 1, and c = 1, and assuming ε = 0.03 ≈ $1/2^5$, we obtain: X_3 = 5 * (3 - 1) + 1 = 11. In words, the reader has experienced X_3 = 11 bits of positive energy from the third line of the poem, on top of the X_2 = 3 bits experienced from the second line of the poem. Taken together, the reader has experienced a total of $X = X_2 + X_3$ = 3 + 11 = 14 bits of positive energy, and most of this energy arises from the effective resolution of a very large violation in the "punch line" of the poem.

3.4 The Punch Line

Notice that this large positive energy from the last line of the poem only comes after a good explanation. For example, the joke does not work when the last line is *said postman to Eve*, because a reader will wonder: Why is the postman talking to Eve about apples? Also the joke does not work when the last line is *said devil to Jill*, because a reader will wonder: Why is the devil talking about apples to Jill, of Jack and Jill fame, or any other Jill? Psychologically, these alternative endings are surprising (high entropy Y) because they appear to have little to do with the proverbial meaning A. But they are also unsatisfying (low conversion Z) because no hypothesis H_i can be found with a high likelihood of having caused the collective evidence (b_3,a_2,a_1).

For example, assume there are five hypotheses and the posteriors are all equal, such that $P(H_i|b_3,a_2,a_1)$ = 0.2 for each hypothesis. In that case $R(M|s)$ = 0.2, and the aesthetic energy at this last stage of the poem is computed as follows:

$$X_3 = [-lg\ \varepsilon] * [G' * 0.2 - G] + c.$$

Using the same values as above for G', G, c, and ε, we obtain: X_3 = 5 * (3 * 0.2 - 1) + 1 = -1. This result is consistent with the mild displeasure that one feels as "I don't get it" after a poor punch line.

Now from a psychological perspective, we must also consider the matter of individual differences [16, 22, 27] – because different people may have different values of G', G, and c, which would affect X, perhaps drastically. In particular, a person's G' would be especially important because G' scales the conversion of entropy to energy, and G' may do so negatively rather than positively. For example, if the biblical Eve herself read the poem, she might find the meaning B to be insulting as it refers directly to her failing in the Garden of Eden. In effect, her success in understanding the poem comes from acknowledging her mistake in the Garden. So assuming she "gets it" with $R(M|s) \approx 1$, she will do so with negative G' because the resolution has negative *valence* for her. For instance, if G' = -3 instead of G' = 3, then her aesthetic energy from the punch line would be X_3 = 5 * (-3 - 1) + 1 = -19 bits. This negative energy, which represents a very large displeasure, is a far cry from the X_3 = 11 bits of positive energy that would be experienced by a person with G' = 3.

On the other hand, perhaps after all these years the biblical Eve would have come to grips with her mistake in the Garden and now be able to laugh about it – especially if she knows that the intent of the artist telling the joke is to make her and others feel better about what happened. Or, if the reader is the devil, then he may feel special delight because the joke makes him look clever. Notice that in either of these cases the aesthetic response has more to do with *dominance* than with valence per se. Indeed for the biblical Eve who can now laugh about her fate, a negative valence from failure in the Garden produces a positive experience from success in making sense of the poem, thanks to her own appraisal of dominance.

Thus based on personal appraisals that affect feelings of dominance, both she and the devil may have a higher *G*' than an average reader, perhaps *G*' = 5 such that the punch line would bring them each X_3 = 5 * (5 - 1) + 1 = 21 bits of positive energy.

Figure 2 sums up the analysis in a plot of total energy *X* computed after each line of the poem. The solid line shows the case where the joke works, as designed, with large positive energy at the end. A dashed line shows the case where energy falls flat at the end because the punch line is not resolved. A dotted line shows the case where the punch line offends a listener, with large negative energy at the end.

4. Generation

The EVE' model was developed for the dual purposes of *evaluating* and *generating* artworks. An underlying assumption is that some capability for evaluation is needed, first and foremost, simply because the generation of artwork must be guided by what makes the art "work" in the mind of an audience (including the artist). The EVE' model makes at least plausible predictions of what works, and thereby suggests the following procedure for generating haiku humor.

Step 1: Choose a topic that involves some controversy or ambiguity between two *meanings*, *A* and *B*.

Step 2: Compose *signals* (a_1,a_2) for the first two lines, in one consistent and coherent phrase. This phrase should set up (a_1) and build up (a_2) to the meaning *A*, while not suggesting the meaning *B*.

Step 3: Compose a signal (b_3) for the third line, which serves as the punch line. The signal b_3 should be unexpected and inconsistent with the meaning *A* based on the first two lines (a_1,a_2), but coherent and consistent with the meaning *B* based on all three lines (a_1,a_2,b_3).

In my own application of this procedure there is no utilization of computing technology, because computers do not currently possess the human knowledge and value structures that would be needed to write good jokes [19]. For example, the procedure works best when the topic (Step 1) is socially taboo or otherwise arousing, and when the punch line (Step 3) is somewhat insulting in celebrating a human failure (of oneself or others). No computer database to my knowledge comes close to capturing these nuances of *dominance* in language and culture. Even the composition of individual signals (a_1, a_2, and b_3) and comprehension of overall meanings (*A* or *B*) go well beyond the current state of science in computing sentiment and style [1].

As described above, I consider the processes of generation and evaluation to be mutually constructive. Of course others have argued quite differently, as in the famous quote by E. B. White: *"Analyzing humor is like dissecting a frog. Few people are interested and the frog dies of it."* Provoked by this controversy, I used the above procedure to write as follows:

frog croaks in haiku
after analyzing it
frog croaks in haiku

I also used the same procedure to generate additional examples, see P.O.E.M. in Figure 3.

5. Conclusion

After reviewing previous theories of aesthetic experience, I presented my EVE' model with a Bayesian basis. The name refers to psychological processes of expectation (*E*), violation (*V*), and explanation (*E'*), which produce the mental energy (*X*) that makes art "work" – through effective conversion (*Z*) of marginal entropy (*Y*), according to $X = Y * Z$. I then applied this model to haiku humor, by computing the total aesthetic *X* after each line of a poem. The analysis demonstrated how energy *X* depends on general processes of *E*, *V*, and *E'*, as well as on individual differences in the background knowledge and value structures of an audience. Finally, turning from evaluation to generation, I offered a procedure for writing haiku humor and employed the procedure to produce further exemplars.

The EVE' model has also been used for evaluating aesthetics in sketching [9], music [10], and gaming [8] – as well as for generating artworks in these and other media [11]. This diversity suggests that the EVE' formulation does indeed capture at least some of the universals that underlie art and aesthetics. However, a detailed analysis of one art form, i.e., haiku humor, highlighted the overwhelming importance of background knowledge and human values. These things are required as input to the EVE' model, in the form of likelihoods $P(e|H)$ for evidential signals (*e*) of hypothetical meanings (*H*), and magnitudes of personal factors (*G', G, c)* that affect cognitive appraisals. The need for such input is clearly a limitation of the EVE' model when it comes to practical applications. However, the same limitation would apply to any model of aesthetics that attempts to address how humans process signals to infer meaning and achieve feelings.

An advantage of my model lies in formalizing the role of psychological processes (*E, V, E'*), and identifying individual parameters *(G', G, c)* that may affect these processes. Another advantage is that my model addresses the temporal and uncertain nature of information processing, via the Bayesian-mathematical formalism of probability. Contrary to the assertion of Stiny and Gips [37], it is *not* unreasonable to expect that an audience will have probabilistic knowledge about the likelihoods of signals and meanings. In fact it is exactly such probabilistic knowledge, gained through experiences that are common among people in a culture, that enable humans to interpret the meanings of signals communicated by artworks and thereby experience the feelings that we refer to as aesthetics.

It is also important to note that a probabilistic approach does not require complete knowledge of probabilities and thus lead to intractability. Instead the analyses here and elsewhere [8, 9, 10] show that approximate, or even order-of-magnitude, probabilities are sufficient for application of the EVE' model. Indeed the haiku analysis in this article shows that imprecise and incomplete knowledge of probabilities, formed in human minds, is actually *necessary* for an audience to achieve the aesthetic experience.

A final advantage of my approach, supported by the examples in Figure 3, is that the EVE' model can offer useful guidance to those who wish to generate artworks – even when the required knowledge and value structures remain implicit in the minds of the artist and his audience.

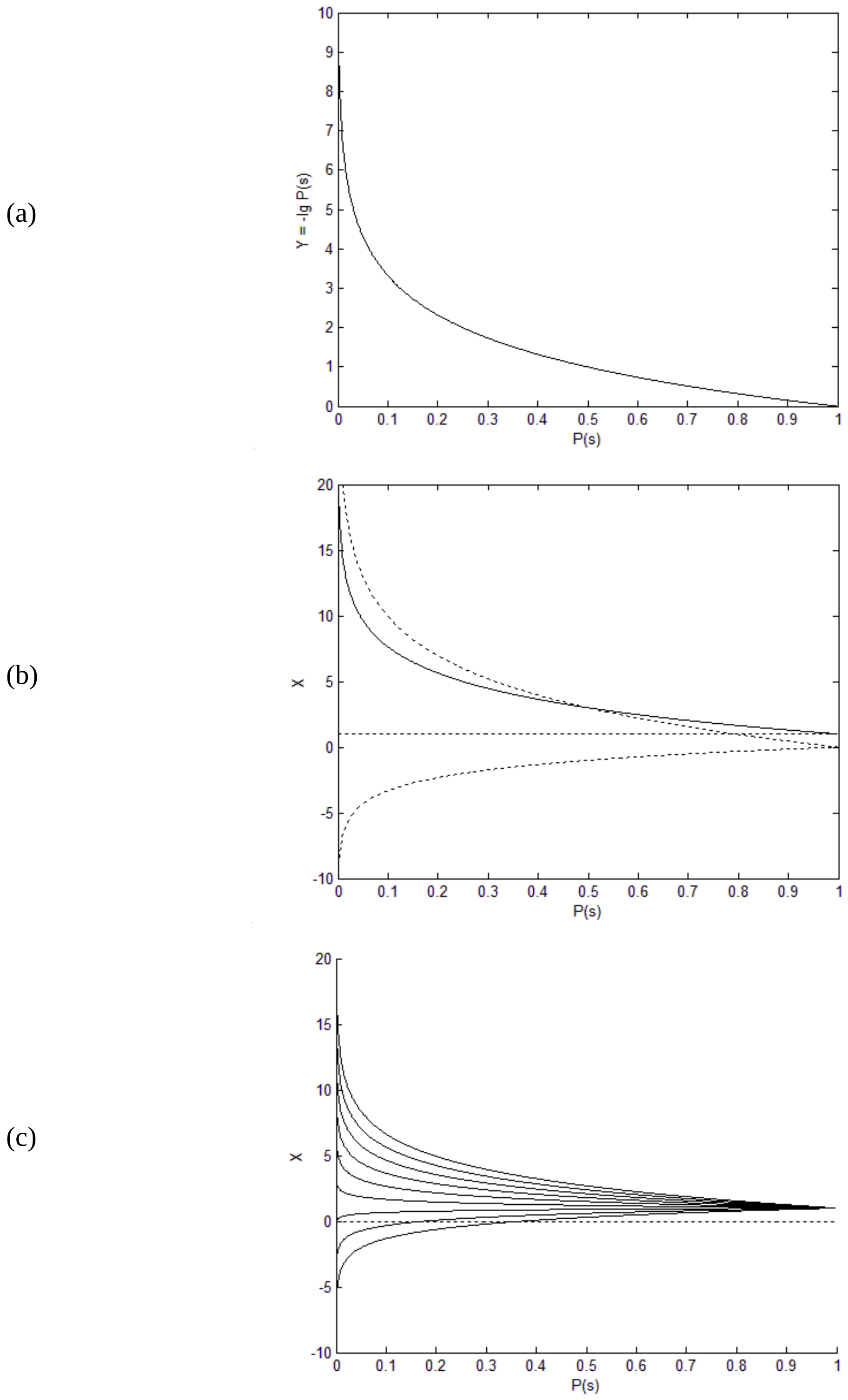

Figure 1. (a) Plot of entropy function, *Y* = *-lg P(s)* versus *P(s)*. (b) The aesthetic measure *X* versus *P(s)*. Solid line plots *X* = *(G * E)* + *c* + *(G' * E')* assuming *G' = 3*, *G = 1*, *c = 1*, and *R(M*|*s)* = 1. Dotted lines are *G * E* at bottom, *c* in middle, and *G' * E'* at top. (c) Family of *X* versus *P(s)* curves, shown for fixed values of *R(M*|*s)* in increments of 0.1, ranging from *R(M*|*s)* = 0.1 at bottom to *R(M*|*s)* = 0.9 at top.

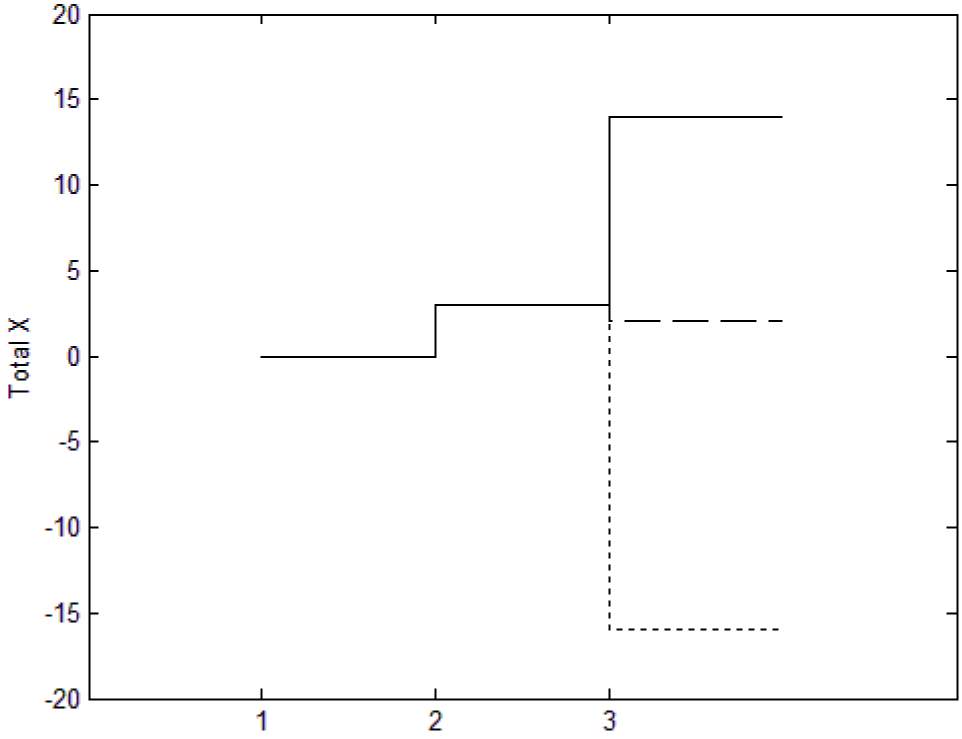

Figure 2. Plot of total (cumulative) energy *X* computed by the EVE' model, after each of the three lines (1, 2, and 3) of a haiku. At line 3, energy *X* is computed for three cases (see text for details).

Perceptions

with just seventeen
syllables a haiku's worth
a thousand pictures

artist and scientist
both seek beauty so they join
a dating service

squirrel has the nuts
in a texas hold'em game
cheeks spoil poker face

life's a win-win game
first the other guy will win
then he wins again

faster than light speed
einstein crossed a bridge before
he could get to it

Occupations

uranium splits
to start the chain reaction
then the neighbors split

statistics don't lie
they just don't tell everything
statisticians know

peter pan grew up
now he can no longer fly
so he takes the bus

i'm an old cowhand
who can't seem to find a job
now'days cows have hoofs

specialists are best
at getting the right answer
to the wrong problem

Expressions

measure twice cut once
but where do you cut when two
measurements differ

coffee tea or milk
well that is a tough question
when you want a beer

sportscaster was seen
wearing women's lingerie
pretty freudian slip

trial and error
is the jeopardy answer
why o.j. is free

object in mirror
is closer than it appears
and then it hits you

Meditations

extra virgin oil
comes from olives that never
even thought of sex

geometry proves
beyond a shadow of doubt
there has to be doubt

early man painted
primal utterance on wall
waiting for subway

lord on this easter
bless our eggs so they create
good cholesterol

mom said you can fool
some people all of the time
was she fooling me

Figure 3. An acrostic collection of haiku humor, titled *P.O.E.M.*

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] A. Baddeley, *Working Memory*, Science 255 (1992), pp. 556-569.

[3] T. Bayes, *An Essay Toward Solving a Problem in the Doctrine of Chances*, Philosophical Transactions (1763) Essay LII, pp. 370-418.

[4] M. Bense, *Aesthetica. Einfürung in die Neue Aesthetik*, Agis-Verlag, Baden-Baden, 1965.

[5] D. Berlyne, *Aesthetics and Psychobiology*, Appleton Century Crofts, New York, 1971.

[6] G. Birkhoff, *Aesthetic Measure*, Harvard University Press, Cambridge, MA, 1933.

[7] P. Bloom, *How Pleasure Works: The New Science of Why We Like What We Like*, W. W. Norton, New York, 2010.

[8] K. Burns, *EVE's Entropy: A Formal Gauge of Fun in Games*, Studies in Computational Intelligence 71 (2007), pp. 153-173.

[9] 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.

[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] K. Burns and J. Sage, *Art Tech Expo*, 2005. Available at http://www.ask-how.org/contents/on_line_exhibition.htm (Accessed 17 July, 2011).

[12] S. Butcher (trans.), *Aristotle Poetics*, Dover Books, New York, 1951.

[13] N. Cowan, *The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity*, Behavioral and Brain Sciences 24 (2001), pp. 87-114.

[14] J. Dewey, *Art as Experience*, Perigee, New York, 1980.

[15] D. Dutton, *The Art Instinct: Beauty, Pleasure, and Human Evolution*, Bloomsbury Press, New York, 2009.

[16] H. Eysenck, *The Experimental Study of the 'Good Gestalt' – A New Approach*, Psychological Review 49 (1942), pp. 344-364.

[17] P. Fishwick (ed.), *Aesthetic Computing*, MIT Press, Cambridge, MA, 2006.

[18] S. Freud, *Der Witz und Seine Beziehung zum Unbewussten*, Deuticke, Leipzig, 1905.

[19] M. Hurley, D. Dennett, and R. Adams, *Inside Jokes: Using Humor to Reverse-Engineer the Mind*, MIT Press, Cambridge, MA, 2011.

[20] L. Itti and P. Baldi, *Bayesian Surprise Attracts Human Attention*, Vision Research 49 (2009), pp. 1295-1306

[21] P. Johnson-Laird, *Freedom and Constraint in Creativity*. R. Sternberg, ed., *The Nature of Creativity: Contemporary Psychological Perspectives*, Cambridge University Press, Cambridge, 1988, pp. 202-219.

[22] H. Kreitler and S. Kreitler, *Psychology of the Arts*, Duke University Press, Durham, NC, 1972.

[23] S. McGrayne, *The Theory that Would Not Die: How Bayes Rules Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, New Haven, 2011.

[24] A. Moles, *Information Theory and Esthetic Perception*, University of Illinois Press, Urbana, IL, 1966.

[25] N. Norrick, *Repetition in Canned Jokes and Spontaneous Conversational Joking*, International Journal of Humor Research 6 (1993), pp. 385-402.

[26] K. Oatley, *Creative Expression and Communication of Emotions in the Visual and Narrative Art*s. R. Davidson, K. Scherer, and H. Goldsmith, eds., *Handbook of Affective Sciences*, Oxford University Press, Oxford, 2003, pp. 481-502.

[27] M. Parsons, How *We Understand Art: A Cognitive Development Account of Aesthetic Experience*, Cambridge University Press, Cambridge, UK, 1987.

[28] J. Paulos, *Mathematics and Humor*, The University of Chicago Press, Chicago, 1980.

[29] 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.

[30] J. Reichhold, *Writing and Enjoying Haiku: A Hands-on Guide*, Kodansha, Tokyo, 2002.

[31] B. Ross, *How to Haiku: A Writer's Guide to Haiku and Related Forms*, Tuttle, Tokyo, 2002.

[32] P. Rozin, A, Rozin, B. Appel, and C. Wachtel, *Documenting and Explaining the Common AAB Pattern in Music and Humor: Establishing and Breaking Expectations*, Emotion 6 (2006), pp. 349-355.

[33] R. Scha and R. Bod, *Computationele Esthetica*, Informatie en Informatiebeleid 11 (1993), pp. 54-63.

[34] K. Scherer, *Appraisal Theory*. T. Dalgleish and M. Power, eds., *Handbook of Cognition and Emotion*, Wiley, New York, 1999, pp. 637-663.

[35] C. Shannon, *The Mathematical Theory of Communication*. C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Chicago Press, Urbana, IL, 1949, pp. 29-125.

[36] P. Silvia, *Emotional Responses to Art: From Collation and Arousal to Cognition and Emotion*, Review of General Psychology 9 (2005), pp. 342-357.

[37] G. Stiny and J. Gips, *Algorithmic Aesthetics, Computer Models for Criticism and Design in the Arts*, University of California Press, Berkeley, CA, 1978.

[38] P. Thagard, *Abductive Inference: From Philosophical Analysis to Neural Mechanisms*. A. Feeney and E. Heit, eds., *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*, Cambridge University Press, Cambridge, UK, 2007, pp. 226-247.

[39] 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 Chicago Press, Urbana, IL, 1949, pp. 1-28.

[40] S. Wilson, *Information Arts: Intersections of Art, Science, and Technology*. MIT Press, Cambridge, MA, 2002.