Atoms of EVE′: A Bayesian basis for esthetic analysis of style in sketching

KEVIN BURNS
MITRE Corporation, Bedford, Massachusetts, USA

Abstract

At its root level, style is actually an esthetic agreement between people. The question is, how can esthetic agreements be modeled and measured in artificial intelligence? This paper offers a formal theory called EVE′ and applies it to a novel test bed of dynamic drawings that combine features of music and sketching. The theory provides mathematical measures of expectations, violations, and explanations, which are argued to be the atomic components of the esthetic experience. The approach employs Bayesian methods to extend information measures proposed in other research. In particular, it is shown that information theory is useful at an entropic level to measure expectations (E) of signals and violations (V) of expectations, but that Bayesian theory is needed at a semantic level to measure explanations (E') of meaning for the signals. The entropic and semantic measures are then combined in further measures of tension and pleasure at an esthetic level that is actually style.

Keywords: Bayesian Theory; Drawing; Flow; Fun; Information Theory; Music

1. ESTHETIC AGREEMENT

Style is a matter of substance in the arts and fashion, where distinctive and desirable styles are design objectives. However, style is also important in any effort that involves communication of information in human–cultural interaction, including modern domains involving human–computer interaction. Thus, a scientific understanding of style is needed for many reasons, not the least of which is to advance the practice of artificial intelligence in design of *information applications*.

However, what exactly is style? In arts and fashion people recognize styles when they see them, and people often agree on what is stylish or what is not. Superficially these agreements are governed by observable features, such as textures and patterns, but fundamentally they are grounded in emotional feelings, such as tension and pleasure. In between there are complex psychosocial processes that determine how features give rise to feelings and vice versa. Thus, the substance of style can be characterized as *esthetic agreement in a cultural context*.

In short, styles exist in a social culture (Sosa & Gero, 2006) where people *produce* styles and *perceive* styles in mutual interaction (see Fig. 1). People have *preferences* that produce distinctive styles in conceptual spaces, and people use *prototypes* (Rosch, 1978) to perceive distinctions in corresponding style spaces. At the highest level these are spaces of features, but at the deepest level they are spaces of feelings. Therefore, a challenge for style research is to connect these spaces computationally.

There are many levels at which people can agree or disagree, but a useful breakdown is the triad proposed by Weaver in his introduction to Shannon's theory of communication (Shannon & Weaver, 1949). Weaver defined "communication" in a broad sense to include "all the procedures by which one mind may affect another." His three levels can be characterized as follows:

A. the *entropic* level, which is concerned with *signals*;
B. the *semantic* level, which is concerned with *meaning*; and
C. the *esthetic* level, which is concerned with *feelings*.

Similarly, one can use the same levels to describe how a sample of "information" (in language, drawing, music, or
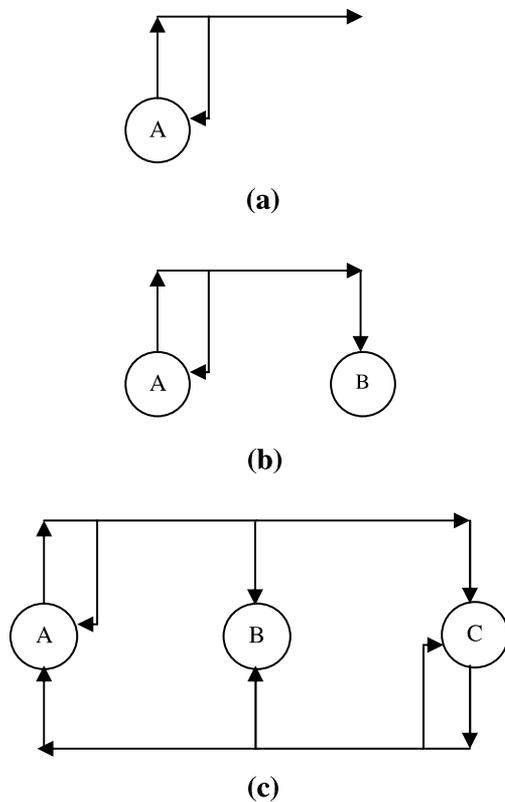
Fig. 1. (a) Production and perception by agent A, an artist; (b) production and perception by agent A, an artist, with perception by agent B, an audience; and (c) the cultural context in which people perform, where artists (A and C) are producing and perceiving while an audience (B) is perceiving.

any other medium) may be said to have a *style*. For example, if the sample of information is a motion picture, then one might distinguish between different styles at level A of signals (e.g., predictable or unpredictable), level B of meaning (e.g., fantasy or documentary), or level C of feelings (e.g., uplifting or depressing).

Shannon's theory of information was concerned only with signals at the entropic level. In fact, he wrote, "These semantic [and esthetic] aspects of communication are irrelevant to the engineering problem" (Shannon & Weaver, 1949). Thus, for Shannon the engineering problem was limited to level A. Now, to advance the theory and practice of information engineering, we must address semantics and esthetics at levels B and C. Toward that end, this paper combines information theory and Bayesian theory in a novel framework called the *Atoms of EVE'*.

In developing this framework I focus on style in sketching because sketches play a key role in the design of buildings, clothing, and many other artifacts. That is, in addition to being worthy of study as artworks themselves, sketches also serve as representations of objects and actions in many engineering and design efforts. This makes sketching a useful test bed for theoretical investigations of style and eventual applications to communication challenges in fields

ranging from architectural design (Do, 2002) to military defense (Forbus, 2004).

Human sketches are often said to exhibit certain styles, and here again we can distinguish between different levels of style. At level A, which is concerned with information in signals, one might refer to the actual drawing as jagged or curvy. At level B, which is concerned with recognition of meaning, one might refer to the depicted subject as a cat or dog. At level C, which is concerned with emotional feelings, one might refer to the audience response as pleased or tense. Clearly these levels are closely related and one must analyze signals at the highest level if one is to understand meaning and feelings at the other levels. However, I argue that the deeper levels are of special concern for style research simply because these levels are the least understood (Oatley & Johnson-Laird, 1987).

Thus, here I focus on esthetic level C that is concerned with emotional feelings and how they arise from observable features. I consider *production* of drawings as well as *perception* of drawings (see Fig. 1), but I stress perception for two reasons. First, previous research on style in artificial intelligence has been more concerned with production of styles than with perception of styles (Stiny & Mitchell, 1978; Kirsch & Kirsch, 1986; McCorduck, 1991; Boden, 2004). Second, artificial agents cannot truly capture human styles unless they perceive styles as well as produce styles, because both are part of the cultural context in which people perform (see Sosa & Gero, 2006). In short, an artificial artist who sketches but does not see can never simulate the psychosocial processes that underlie esthetic agreements.

2. DYNAMIC DRAWINGS

The act of *producing* a sketch is clearly a dynamic process, but so is the act of *perceiving* a sketch. That is, even though sketches are often treated as static objects, the cognitive processes by which people perceive them are not static at all. With conscious gazes and unconscious saccades, human perception of a sketch involves the piecing together of numerous snapshots in a temporal sequence, which depends on where one attends, which in turn depends on what one perceives. In this sense sketches are like songs or stories that are not only produced in a sequential manner by artists but are also perceived in a sequential manner by audiences.

Here, I study style in *dynamic drawings*, which are sketches in time. These dynamic drawings are akin to musical melodies that do not fade, so that each signal segment adds to the drawing without attenuation. One advantage of this medium is that it offers more control over what a viewer sees at each point in time, compared to static drawings where there is no control over how a viewer attends to various segments. Another advantage is that the dynamics of these drawings are similar to music and stories and the motion pictures seen in video media. I begin by reviewing some research on sketching and music to motivate the study of dynamic drawings.

2.1. Sketching silhouettes

Pilot studies of style in sketching (Burns, 2004) explored the art of caricature, which some have suggested is the essence of all art (Ramachandran & Hirstein, 1999). In these studies I focused on contour drawings of animal silhouettes, which are much like the first forms of wall art (Wachtel, 1993; Mithen, 1996).

I studied both perception and production of animal caricatures by human beings and machine programs. In perception the problem involved judging similarity relative to a prototype, and I found that a piecewise matching method could be used to compute a measure that corresponded to human judgments of similarity. In production my approach used this matching method to compute the caricature of a subject drawing relative to a reference drawing, by making "controlled" exaggerations between "matched" segments of the two drawings. This automated the whole process of drawing a caricature, which was an improvement over previous work in which the piecewise matching was done manually and the matched segments were input to an exaggeration algorithm (Brennan, 1985).

In production, the "control" comes from an artist's preference for an exaggeration magnitude and the "matching" is to a prototype in his mind. In perception, the audience presumably has a similar prototype in their minds as well as their own preferences for an exaggeration magnitude that looks best. My main finding was that computer-drawn caricatures did indeed appear to capture and enhance the style of human drawings (see also Brennan, 1985), and exaggeration in moderation did indeed seem to produce more distinctly recognizable animal silhouettes (see also Rhodes et al., 1987; Rhodes & McLean, 1990). In short, these studies suggested that the art of caricature can be characterized as a computational process that employs preferences and prototypes in producing and perceiving *exaggerations of a thing*, much like the musical device of *variations on a theme* (see also Matisse, 1995).

2.2. Musical measures

Dubnov et al.'s (2004, in press) recent research on music also studied *similarity* and other esthetic aspects of style. In particular, they developed two measures, called *signal recurrence* (SR) and *information rate* (IR), and discovered a correlation between these computational measures and psychological judgments of *familiarity rating* (FR) and *emotional force* (EF), respectively. Listeners made the psychological judgments continuously in time as they heard a musical piece. This research is reviewed here because it relates to my studies of sketching and because it suggests how one might derive mathematical measures of esthetic experience in any medium.

2.2.1. Similarity

The research by Dubnov et al. (2004, in press) focused on timbre, which refers to the tonal quality or "texture" of the music. Compared to contour segments in outline drawings, the analysis of timbre in music involves a much more complex signal of numerous dimensions at each time segment. Nevertheless, the same basic notion of similarity applies and in fact Dubnov's measure of similarity in music was analogous to my measure of similarity in drawings. In both cases, the signal (music or contour) was approximated by a sequence of vectors at discrete times or points, each with a computed orientation in a dimensional space where piecewise similarity is measured by vector differences. However, I computed similarity between two drawings, where one served as the prototype (baseline) for comparison, whereas Dubnov et al. computed the similarity of each signal segment's orientation relative to the average orientation of all signal segments in the same song. This may help explain why their data show large differences between the computational measure of similarity (SR) and the psychological judgment of familiarity (FR) near the beginning of the song. That is, listeners were asked to rate "how familiar what they were currently hearing was to anything they had heard from the beginning of the piece," and it is not clear what prototype the listeners were actually using as a baseline for comparison. However, it was probably not the average of all signals for the piece simply because they had not yet heard the whole piece.

Nevertheless, it is interesting that Dubnov et al. (2004, in press) found significant correlation between a computational measure (SR) and a psychological judgment (FR) for timbre in music (mostly later in the piece), much as I did for contour in drawings. However, it is also important to acknowledge that the prototypes for these studies were either assumed from a computed average (in the music) or given as reference images (in the drawings). A major challenge for style research is to better characterize the details of the prototypes that exist in the minds of human perceivers because a machine cannot make matches to a model it does not have.

2.2.2. Entropy

In addition to using SR to measure signal similarity, Dubnov et al. (2004, in press) derived another measure called IR that uses *entropy* (Shannon & Weaver, 1949) to measure signal *unpredictability*. More formally, the *marginal entropy* of a single signal (m_i) is measured as $-P(m_i|M)\log P(m_i|M)$, where M is a model of the underlying process that generates a set of signals $\{m_i\}$ and $P(m_i|M)$ is the probability of a single signal m_i being selected from the set. Mathematically, the entropy for a set of signals $\{m_i\}$ is the sum of marginal entropies for all m_i in the set, that is, $-\Sigma_i P(m_i|M)\log P(m_i|M)$. Conceptually, entropy is highest when there are numerous messages in the set and when all messages are equally likely to be selected, whereas entropy is lower when there are fewer messages in the set and/or some messages are more likely than others to be selected.

An elegant aspect of information theory is the correspondence between entropy as defined above and entropy as derived in statistical mechanics. However, this correspondence arises in part because Shannon chose to measure expectation as log P rather than raw P, for which he offers three reasons (Shannon & Weaver, 1949):

1. "It is practically more useful," for example, in hardware like bistable relays where the log number of messages is what determines the number of relays needed to encode the set of messages.
2. "It is nearer to our intuitive feeling as the proper measure . . . since we intuitively measure entities by linear comparison with common standards."
3. "It is mathematically more suitable," in that it simplifies some limiting operations in the governing equations.

For Shannon the first and third reasons were the most important because his theory was concerned only with signals at entropic level A. However, in extending this theory to semantic level B (meaning) and esthetic level C (feelings), the second reason is actually the most important because it is not clear exactly what *intuitive* means.

Here I suggest that intuitive means *consistent with the cognitive structure by which people naturally reason*, which is known to involve a mental "number line" (Dehaene, 1997). Thus, assuming that psychological estimation is basically linear, a condition that should be satisfied by an intuitive measure of expectation is as follows: the measure and its inverse should sum to zero, which is the mental anchor about which linear inverses are symmetric. At first examination, a measure of raw P and its additive inverse $1 - P$ might appear to be the most intuitive. However, P and $1 - P$ sum to 1.0 (not zero) and they are symmetric about 0.5 (not zero). Similarly, a measure of raw P and its multiplicative inverse $1/P$ do not satisfy the condition. However, $\log P$ and the log of P's multiplicative inverse $\log 1/P = -\log P$ do satisfy the condition, and in this sense $\log P$ can be considered as an intuitive measure of expectation. Note that this "intuitive" argument (above) makes no commitment as to the base of the log, which can be converted from any base "a" to another base "b" by a multiplicative constant $\log_b a$.

2.2.3. Predictability

Founded on entropy, Dubnov et al. (2004, in press) propose IR to measure the *predictability* of a signal segment m_i given a model M of the process that is generating the signal segments. In the case of music, a listener's model is a model of the composer/performer's preferences, as measured relative to some prototype, because these preferences determine the likelihood that a segment m_i will be generated and hence heard by the listener.

Formally, the IR at a segment (snapshot) m_i of music is defined as the difference in multiple information between two sets of signals: {all signals heard up to time i} and {all signals heard up to the previous time $i - 1$}. This difference can be written as the mutual information between two sets of signals, a single-item set $\{m_i\}$ and a multiple-item set $\{m_{i-1}, \ldots, m_1\}$, as follows:

$$\mathrm{IR}_i = I(\{m_i\}; \{m_{i-1}, \ldots, m_1\}).$$

Using H to denote entropy, the mutual information between two sets X and Y is as follows:

$$I(X; Y) = H(X) + H(Y) - H(X, Y),$$

where the joint entropy $H(X, Y)$ is as follows:

$$H(X, Y) = H(X|Y) + H(Y).$$

Thus, as shown by Dubnov et al. (2004, in press), the definition of IR reduces to the following expression:

$$\mathrm{IR}_i = H(m_i) - H(m_i|m_{i-1}, \ldots, m_1) = H(m_i|M_0) - H(m_i|M_n),$$

where M denotes a model of the signals that captures the conditional dependency of a given signal m_i on previous signals, and the subscript n denotes the Markov order of this dependency. For example, if the model of the composer/performer is such that each note in a song depends only on the previous two notes of the song (m_{i-1}, m_{i-2}), then the subscript n would be 2. Here, M_0 denotes a "baseline" model, which Dubnov et al. (2004, in press) assume is the same model M but of Markov order zero, such that the probability of m_i in M_0 is the independent probability of m_i computed from M, ignoring the dependencies of m_i on $m_{i-1}, \ldots,$ and so forth, that are modeled in M_n. Therefore, one might say that $H(m_i|M_n)$ reflects the preference of a producer to generate a signal m_i as modeled by the perceiver's model M_n of the producer. Likewise, one might say that $H(m_i|M_0)$ reflects a prototype that serves as a baseline from which preferences can be measured by perceivers and producers in their common culture.

Fundamentally, IR measures the *amount of information* provided to a listener by a signal m_i in light of the *mental models* (see Burns, 2001) M_n and M_0 that the listener is assumed to have in his head. Mathematically, IR is given by the entropy of m_i computed from the reference model M_0 minus the entropy of m_i computed from the preference model M_n. Conceptually, IR is large when the unpredictability of m_i is much lower (i.e., the predictability of m_i is much higher) in model M_n than in model M_0, which means that having model M_n gives the observer a lot of information about m_i. Thus, large IR means that having model M_n allows the observer to make a much better *prediction* of m_i than he could otherwise make from model M_0.

Note that the entropy difference computed by IR measures the predictability of signals *within* a model M, using the *joint entropy* of a single-item set $\{m_i\}$ and a multiple-

item set $\{m_{i-1}, \ldots, m_1\}$ to measure how well M_n predicts m_i (relative to how well M_0 predicts m_i). This can be contrasted to the *cross entropy* that might be computed as a measure of similarity taken *between* two models. For example, others (Jupp & Gero, 2003) use the cross entropy of two multiple-item sets as an overall measure of similarity between these two sets, which is more a measure of *model similarity* than it is a measure of *signal predictability*.

As a practical matter, the psychological plausibility of IR is critically dependent on the correspondence between actual *mental models* and assumed mathematical models, in this case M_n and M_0. In their study of music, Dubnov et al. (2004, in press) assumed that M_n was of low order (e.g., $n = 1$ or 2), and they used the whole data set (song) to construct the models M_n and M_0. In computing IR, these M_n and M_0 were then assumed to apply at each segment i, even early in the song when the listener's models may be much different from M_n and M_0 because they have not yet heard much of the song (signals). These assumptions appear to be reasonable for the case of timbre in music, as evidenced by the experimental finding of a correlation between the computational measure of IR and the psychological judgment of EF. However, the same assumptions may not be so reasonable for other aspects of music, such as melody, which is analogous to contour in drawings.

Moreover, I posit that IR, which concerns the *prediction of signals*, is missing an important component of esthetic experience, which concerns the *perception of meaning*. Later I discuss the difference in the context of a dynamic drawing that combines features of music and sketching. However, I first discuss the notion of Bayesian belief as a basis for perception of meaning as well as prediction of signals.

2.3. Bayesian belief

In Bayesian belief there is an important distinction between probabilities of the form $P(m|M)$ and probabilities of the form $P(M|m)$, where M is a model (e.g., of a song or sketch); m is a datum (e.g., a signal or segment); $P(m|M)$ is called a *likelihood*, which is the probability of m in light of M *before* m is observed; and $P(M|m)$ is called a *posterior*, which is the probability of M in light of m *after* m is observed. The likelihood and posterior are related by a *prior* $P(M)$, which is the probability of M before m is observed. The relation is based on the following axiom of probability theory:

$$P(M, m) = P(m) \times P(M|m) = P(M) \times P(m|M) = P(m, M),$$

where the middle equality can be rewritten as follows:

$$P(M|m) = P(M) \times P(m|M)/P(m).$$

Now assuming that there are K models $\{M_1, M_2, \ldots, M_K\}$ in a perceiver's *frame of discernment*, where each model M_k has a prior probability $P(M_k)$ and a likelihood $P(m|M_k)$ of causing m, the term $P(m)$ can be written as the sum of $P(M_k) \times P(m|M_k)$ over all k and Bayes rule can be rewritten as follows:

$$P(M_k|m) = P(M_k) \times P(m|M_k)$$

$$\div [P(M_1) \times P(m|M_1) + P(M_2) \times P(m|M_2) \cdots$$

$$+ P(M_K) \times P(m|M_K)].$$

In other words, the probability that an observed signal m was caused by a modeled process M_k is the product of the prior $P(M_k)$ and the likelihood $P(m|M_k)$, normalized by the sum of such products for all modeled processes that might have caused the observed m.

The advantage of Bayes rule is that it allows one to infer cause (M) from effect (m). That is, the world works from cause to effect where a cause modeled by M gives rise to an effect observed as m. This causal knowledge can be captured in a likelihood of the form $P(m|M)$, which is useful for predicting the most likely effect (m) given a causal model (M). However, these likelihoods alone are not enough for perceiving (Knill & Richards, 1996) the most likely model (M) for the cause of the effect (m), which is the reverse inference. Bayes rule provides the mathematical machinery for making this reverse inference of meaning (M) from signal (m), using likelihoods along with priors to compute posteriors.

As such, a Bayesian approach is useful for two kinds of model-based reasoning: forming *expectations* (i.e., for prediction of signals *before* they are received) and forming *explanations* (i.e., for perception of meaning *after* signals are received). The Bayesian approach is especially important for forming explanations of meaning, but a Bayesian approach is also important for forming expectations of signals because it considers more than one model in its prior-weighted computations. That is, the above expression for $P(m)$ gives a measure of expectation for the signal m over all models $\{M_k\}$ that might cause that m. This expectation will be equal to the likelihood $P(m|M_k)$ given by a single model M_k, as employed by IR, only if M_k is the only model in a perceiver's frame of discernment.

Now I present an example of esthetic experience using a dynamic drawing to highlight the advantage of a Bayesian approach. The example is presented as a thought experiment. Imagine that the dog in Figure 2c was drawn in time from start to end, starting at the ear and moving counterclockwise. For example, Figures 2a and 2b show snapshots in the sequence.

At a given point in this dynamic drawing, the questions are how a viewer would form expectations about what she will see next in the sequence, as well as explanations about what she has seen thus far in the sequence, and how these expectations and explanations would affect the viewer's esthetic experience. More specifically, in the language of information theory, how would one compute the "amount of information" that a perceiver receives from each seg-
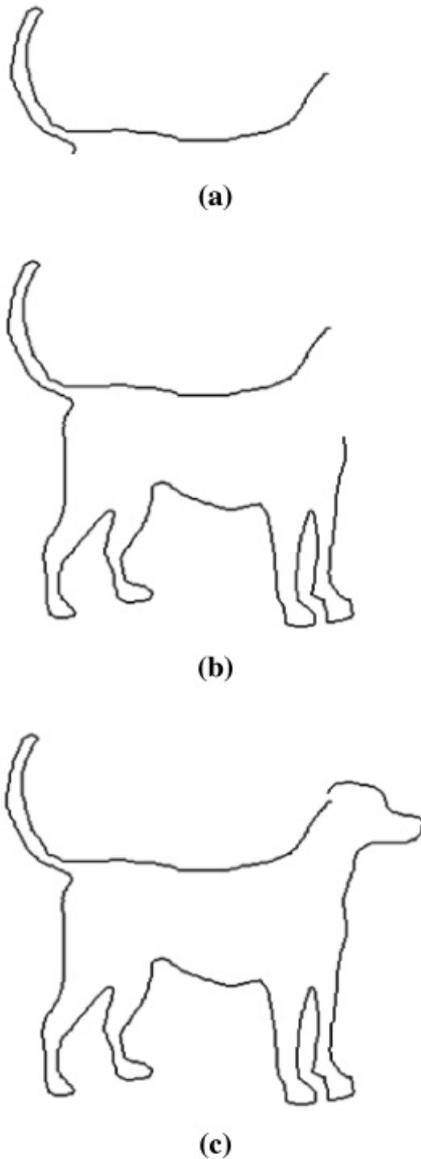
(a)

(b)

(c)

Fig. 2. The dynamic drawing used in a thought experiment.

ment of the signal (drawing or music), and how might this measure of *information* at level A (see Section 1) be related to the perceiver's *recognition* at level B or *emotion* at level C?

2.3.1. Discretizing the curve

To begin a perceiver must have a mental model (Burns, 2001) of the animal/artist that is producing the signals, as a basis for forming expectations and explanations. In fact, a perceiver probably has many such models M_k in mind, but for now I assume just one model in which a perceiver has internalized the probabilistic structure of the dog drawing based on having seen this drawing or other dogs. A simple version of such a model can be computed as follows: first we select 200 equally spaced points along the dog contour (Fig. 2c) and connect them by straight lines (d_i). Using the

vector orientations of these segments (see Fig. 3a), we compute curvature at each segment using the method of Mokhtarian and Mackworth (1986; see Fig. 3b). We then take curvature as a feature basis for modeling the process that generated the contour and compute the probability density of discrete curvature values ranging from very concave (large negative) to very convex (large positive).

The result is a Gaussian-like distribution (Fig. 3c), which is skewed toward positive because the silhouette is a closed contour (i.e., overall it is more convex than concave). This discrete distribution of curvatures can now be treated as a model M of the process by which the drawing was generated. That is, Figure 3c models the probability $P(c_d|M)$ on the y axis of each discrete curvature value (c_d) on the x axis, and this gives the probability that a c_d will actually appear as the curvature (c_i) at time i in the dynamic drawing. Here, I denote this model as M_0 because it assumes that the curvature of each segment is independent of other segments, that is, the animal/artist who produced the dog drawing is modeled as a Markov process of order zero.

Now with such a model M_0, a perceiver could form expectations about the curvature that will be seen at each time i as the dog is drawn. Likewise, a perceiver could measure the amount of information given by a segment at time i based on the *unexpectation* of that segment's curvature c_i in the model M_0, that is, $-\log P(c_i|M_0)$, which is high when P is low. In short, when $P(c_i|M_0)$ is low, then seeing the segment c_i gives the perceiver a lot of information in light of model M_0 because M_0 had little expectation for a segment with curvature $c_i = c_d$. Thus, the *amount of information* in a signal is a *measure of unexpectation* for that signal.

2.3.2. Recognizing a cat

In fact, the above approach has recently been used by several researchers to compute a measure of information along contours (Feldman & Singh, 2005). The idea was first proposed informally by Attneave (1954) in his famous drawing known as "Attneave's cat" (see Fig. 4a), which Attneave made to demonstrate that relatively few points of high curvature could be connected by straight lines to make a recognizable drawing. This suggested to him that these points held most of the information in the drawing. Subsequent researchers (Resnikoff, 1985; Feldman & Singh, 2005) have formalized this claim assuming a Markov model of order zero, arguing that because segments of high curvature contain the most information, they must be the most important segments for making a recognizable drawing, that is, the amount of information and the *goodness of recognition* are directly related.

To test this claim, I took Figure 4a and deleted some portions of straight segments to get Figure 4b. Then, I deleted *all* segments of high curvature (i.e., intersections) in Figure 4b to get Figure 4c. Thus, Figure 4b should be almost as recognizable as Figure 4a because I deleted just a small amount of information (also see Biederman, 1987), and Figure 4c should be nearly *unrecognizable* because I
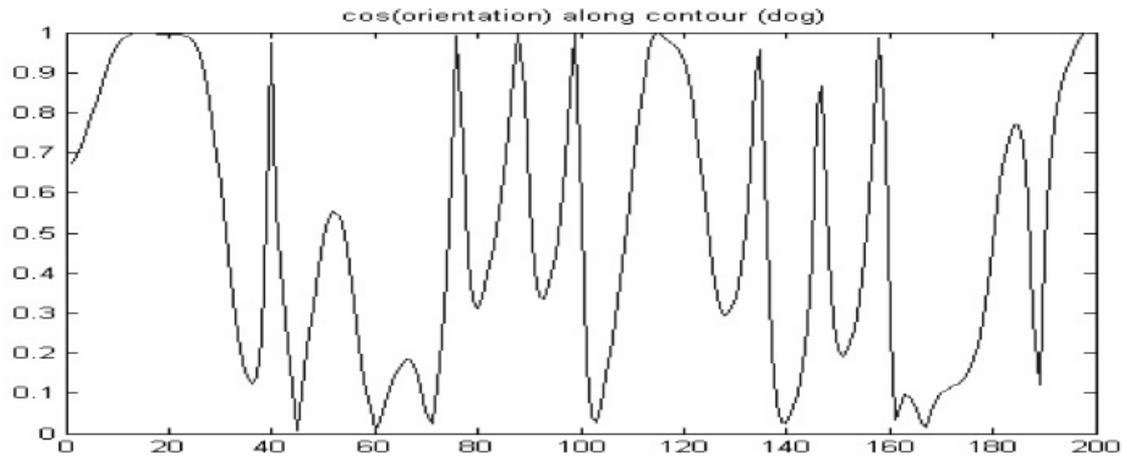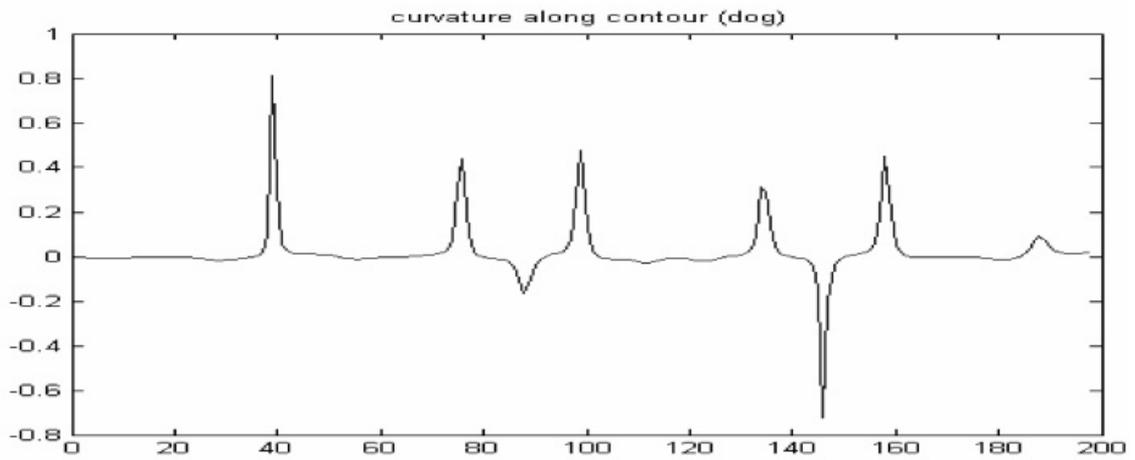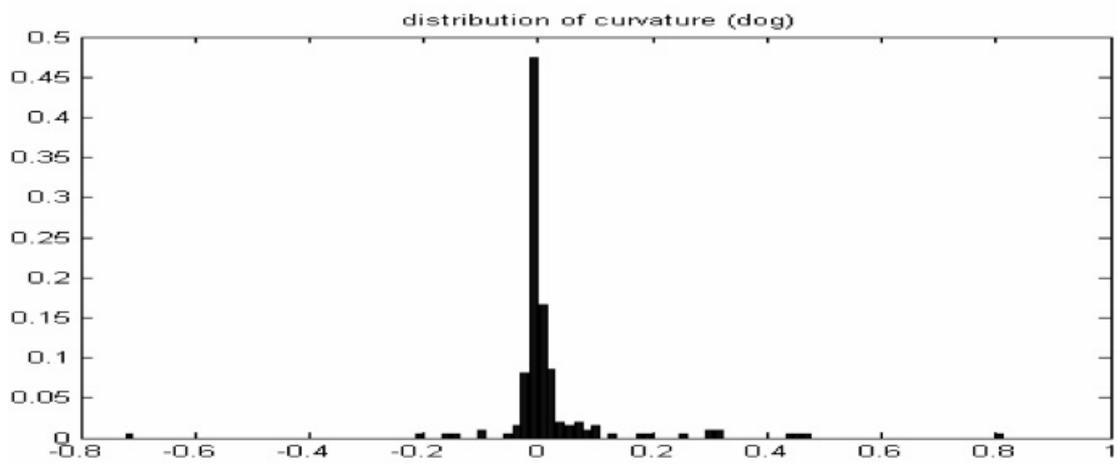
Fig. 3. (a) The orientation of contour segments computed along the dog of Figure 2, starting at the ear and moving counterclockwise; (b) the curvature of contour segments computed along the dog of Figure 2, starting at the ear and moving counterclockwise; and (c) EVE' the probability density for discrete curvature of contour segments in the dog of Figure 2.
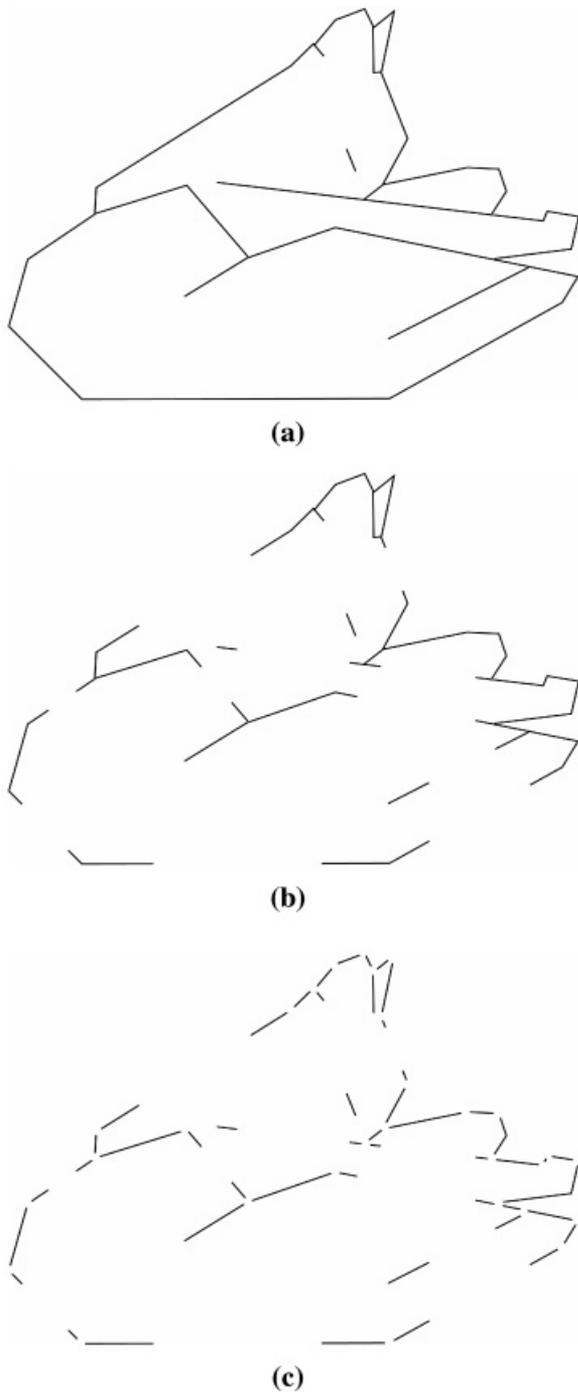
(a)

(b)

(c)

Fig. 4. Attneave's (a) original cat and (b, c) variations.

deleted nearly *all* the information. In other words, the relative amount of information, hence, the relative goodness of recognition, is predicted to be in relation Figure 4a ≈ 4b ≫ 4c. However, examination of these three cats indicates that the goodness of recognition is Figure 4a ≈ 4b ≈ 4c, perhaps even Figure 4a ≈ 4b < 4c. That is, Figure 4c is still recognizable; in fact, many viewers report that Fig-

ure 4c is even more recognizable than Figure 4b (which presumably has much more information) and even more recognizable than Figure 4a (which presumably has the most information). They said that one reason is that Figure 4c has a *softer style* that is more characteristic of a cat. Moreover, many viewers report that the *most important* segment in all three figure parts (a, b, c) as far as recognizing a sleeping cat, which is the object that Attneave (1954) intended to draw, is the small straight segment that depicts the cat's eye; and yet this segment, because it is straight, carries almost *no information*!

Thus, although curvature extrema are certainly important in the human perception of shape from contour (Richards et al., 1988), there is something missing from the above information-theoretic argument about how the goodness of recognition depends on the amount of information. I argue here that the problem is the *model*. That is, any *measure* of information by a perceiver is critically dependent on the perceiver's *models* because information must always be measured in the context of models. In that vein I argue that there are at least three flaws in the information-theoretic analysis of Attneave's cat (1954). The first flaw is that the analysis did not consider assumptions that the viewer makes in mentally modeling the data itself, for example, the way the viewer might fill in the gaps with lines of her own. The second flaw is that the analysis considered only one causal model; that is, it did not address the fact that the viewer is also considering other possibilities about what the drawing may depict other than a sleeping cat, perhaps a dog. Here the cat is recognized when $P(\text{cat}|c) > P(\text{dog}|c)$, so recognition is governed by *multiple posteriors* $P(\text{cat}|c)$ and $P(\text{dog}|c)$ rather than a *single likelihood* $P(c|\text{cat})$.

The third flaw in the information-theoretic argument is that the assumed model is of Markov order zero, which means that all segments are independent. This is not a good model of the cats depicted in mental models. To see why, consider another thought experiment in production rather than perception and imagine that the same Markov model was sampled in a sequence to make a dynamic drawing. With this model of zero order, low curvature segments would be drawn much more often than high curvature segments and positive curvature would be drawn more often than negative curvature. However, the chances are slim that the result would look anything like a cat or dog. This is because cats and dogs and other animals are characterized by distinctive features like heads and tails and legs, and the modeled dependencies between contour segments must be much higher than zero order to generate these naturalistic regularities (Richards, 1988) in random sampling. For example, in my mental model of a cat, the chances of seeing a leg are fairly high if none has been seen yet because cats have legs; and the chances of seeing a leg are very high in the neighborhood where one leg has already been seen because cats legs are usually close together, but the chances of seeing a leg drop to almost zero after four legs have been seen because most cats have just four legs.

2.3.3. Recognizing the dog

In light of this critique, I now return to the dog in Figure 2 and consider IR. Using two models (M_2 and M_0), IR would compute the amount of information for a segment m_i as the *difference* between two *unpredictabilties* (entropies) for m_i in these two models, rather than the unexpectation for m_i in a single model M_0 as discussed above.

However, IR really uses only one model, M. Thus, IR is actually measuring the amount of information in M_2 relative to M_0, and realistically M_2 is probably not much better than M_0 for modeling the probabilistic dependencies in mental models of dogs and cats (see above). In contrast, if one were interested in recognizing the texture or character of the contour (e.g., jagged or curvy) rather than recognizing the meaning or animal depicted in the drawing (e.g., dog or cat), then a low order model like M_2 may be enough. Moreover, because timbre in music is analogous to texture in drawing, a low order may also be enough for the case of timbre in music. However, for other features like melody, which is analogous to contour, the comparison of music to drawings suggests that a much higher order will be needed if a measure like IR is to predict EF or any other esthetic feeling.

Nevertheless, the Markov order is an issue for the model used by IR and is not a fundamental limitation of the measure IR itself. A more critical limitation of the measure itself is that IR considers only one model M so it can only measure things *within* that single model, not *between* different models M and N. In addition, although current research is underway to extend IR to the case of multiple models (Dubnov, 2006), a fundamental limitation is that even a multimodel IR would still only measure expectations for *signals among* these models; that is, it would not measure explanations of *meaning between* the models. To see the difference, I return to my thought experiment on the dog drawing.

To begin, I have a high prior for cats even before the dynamic drawing begins, perhaps because we have just discussed Attneave's cat (1954). Thus, assuming that I have only two models (cat and dog) in my frame of discernment, my belief would be $P(\text{cat}) > P(\text{dog})$. At some time when I have seen a fraction D of the complete contour, perhaps the back and tail (Fig. 2a), which look rather catlike, I might think "Yes, it's a cat." Mathematically, $P(\text{cat}|D) > P(\text{dog}|D)$.

Now my mental models (Burns, 2001) for cats and dogs also include likelihoods for $P(D_j|\text{cat})$ and $P(D_j|\text{dog})$, which reflect my knowledge of how some sample D_j of the contour will look at a later time if indeed the drawing is one of a cat or a dog. Weighted by my priors for $P(\text{cat}|D)$ and $P(\text{dog}|D)$ above, these likelihoods govern my expectations for the future D_j. For example, because cats and dogs both have legs my likelihoods would be $P(\text{leg}|\text{cat}) \approx P(\text{leg}|\text{dog}) \approx 1$, so I would expect a leg regardless. Notice that these high likelihoods mean that a leg carries only a small amount of information in each model (cat or dog), and yet legs are clearly important for distinguishing cats or dogs from other objects like snakes. Notice also that the leg likelihood is about the same in each cat and dog model, which means that although a leg may be a good feature for distinguishing cats or dogs from snakes, it is not good for distinguishing cats from dogs.

Eventually I will see D_j that includes the head (Fig. 2c), which shows a snout. This was not my expectation, because my prior was $P(\text{cat}) > P(\text{dog})$ and my likelihoods for snouts are $P(\text{snout}|\text{cat}) \approx 0$ and $P(\text{snout}|\text{dog}) \approx 1$, that is, $P(\text{snout}|\text{cat}) \ll P(\text{snout}|\text{dog})$ because cats do not have snouts and dogs do. Thus, I am surprised and aroused by this *violation* of my expectation. To update my beliefs in light of this new datum, I then take the normalized product of the snout likelihoods and my priors, per Bayes rule, and form an explanation of the snout that I have just seen. This normalized product of prior × likelihood (see Section 2.3) makes my posterior belief $P(\text{cat}|D_j) < P(\text{dog}|D_j)$, which is a reversal of my prior belief $P(\text{cat}|D) > P(\text{dog}|D)$. Thus, I think "Hey, it's not a cat . . . It's a dog!" This explanation helps to resolve the tension caused by the earlier violation of my no-snout expectation, which gives me pleasure, much like the feeling of satisfaction I get in solving a crime in mystery or "getting" a joke in comedy. Therefore, based on this example I would argue that IR or any other measure of information, which is really a measure of E, is missing an important component of esthetic experience, namely, a measure of E'.

In short, expectations occur before signals are observed and they are concerned with predicting the signals, whereas explanations occur after signals are observed and they are concerned with perceiving the meaning. The two are at different levels of Weaver's triad (see Section 1) because E are concerned with signals at level A while E' are concerned with meaning at level B. However, the two are connected because meaning is derived from signals and because *expectation* (E) and *explanation* (E') can both affect feelings at level C.

3. MATHEMATICAL MEASURES

3.1. Atoms of EVE′

Here I propose that the atomic components of esthetic experience are E, V, and E' (EVE′). The theory is that a feeling of pleasure arises in two ways, both involving some level of success in cognitive processing during a media experience. One kind of pleasure (p) arises from success in forming E and *avoiding* V, whereas another kind of pleasure (p') arises from success in forming E' and *resolving* V.

3.1.1. Expectation

Assume that a perceiver has a set of posterior explanations $\{P(H_k|\{D_{i-1}\})\}$ for a set of observations $\{D_{i-1}\}$ up to some time $i - 1$. Here, P denotes probability and H_k is a

hypothesized explanation for the observations $\{D_{i-1}\}$. Each posterior probability $P(H_k|\{D_{i-1}\})$ then becomes a prior probability for the next i, and these priors along with likelihoods of the form $P(D_{j,i}|H_k)$ can be used to establish the expectation for each possible datum $D_{j,i}$ of type j that might be observed at i.

For example, near the end of the thought experiment in Section 2.3.3, I had two possible explanations, H_c = cat and H_d = dog, and I considered two possible observations at i, $D_{s,i}$ = snout or $D_{n,i}$ = no snout. In this case there would be four likelihoods and each probability $P(D_{j,i})$ would be a prior-weighted sum of likelihoods as follows:

$$P(D_{s,i}) = P(H_c|\{D_{i-1}\}) \times P(D_{s,i}|H_c) + P(H_d|\{D_{i-1}\})$$

$$\times P(D_{s,i}|H_d),$$

$$P(D_{n,i}) = P(H_c|\{D_{i-1}\}) \times P(D_{n,i}|H_c) + P(H_d|\{D_{i-1}\})$$

$$\times P(D_{n,i}|H_d).$$

Here, $\log P(D_{s,i})$ and $\log P(D_{n,i})$ can be taken as measures of expectation for each datum, thus:

$$E_{j,i} = \log P(D_{j,i}).$$

3.1.2. Violation

Now when time i comes and a particular datum $D_{j,i}$ is actually observed, the inverse of E or $\log(1/P(D_{j,i})) = -\log P(D_{j,i})$ can be taken as a measure violation at that time. Thus, $V = -E$; hence,

$$V_{j,i} = -\log P(D_{j,i}).$$

Note that this measure of V is not a measure of information (entropy) per se, because entropy would be $-P(D_{j,i}) \times \log P(D_{j,i})$. However, before the datum is actually observed at time i, a measure of *apprehension* (A) about which data will actually be observed is in fact equal to entropy. That is, taking the sum of all possible V weighted by their respective probabilities, and using this as a measure of A about which data will be observed, we have

$$A_i = -\sum_j P(D_{j,i}) \times \log P(D_{j,i}).$$

The difference between V and A is as follows: V measures unexpectation for the single datum $D_{j,i}$ that was actually observed with probability one at time i whereas A measures unpredictability for the set of possible data $\{D_{j,i}\}$, computed by summing and weighting each unexpectation by its probability.

In comparing the measure of entropy given by A to that of IR discussed in Section 2.2.3, I highlight two important differences. First, in A each $P(D_{j,i})$ is a prior-weighted likelihood computed over all models $\{H_k\}$ in a frame of dis-

cernment, whereas IR uses just one model (but see Dubnov, 2006). Second, in A the measure is one of *absolute* entropy for a set of signals $\{D_{j,i}\}$, whereas in IR it is the *relative* entropy for a single signal m_i in a model M_n compared to its "baseline" entropy in a model M_0, which is the same model M but of Markov order zero. In short, A measures absolute entropy for a set of signals in a set of models whereas IR measures relative entropy for a single signal between two versions of a single model.

3.1.3. Explanation

Now after a violation of a prior expectation, the modeled hypotheses $\{H_k\}$ in a frame of discernment serve as possible explanations (E'), and the after-datum posteriors $\{P(H_k|D_{j,i})\}$ computed by Bayes rule can be used to measure how well the datum has been explained, that is, how well the violation has been resolved. Here I assume that a posterior probability of one would completely resolve the violation; hence, $P(H_k|D_{j,i})$ measures the fraction of $V = -\log P(D_{j,i})$ that has been successfully resolved by an explanation H_k. Thus,

$$E'_{k,j,i} = -P(H_k|D_{j,i}) \times \log P(D_{j,i}).$$

For example, in Section 2.3.3 I had a violation of my expectation when I saw the snout $D_{s,i}$ because $P(D_{s,i})$ was low in light of my prior-weighted likelihoods. The magnitude of this violation can be measured as $-\log P(D_{s,i})$. Then, in forming an explanation, my Bayesian belief changed from prior $P(H_d) < P(H_c)$ to posterior $P(H_d|D_{s,i}) > P(H_c|D_{s,i})$. The upshot was some resolution of the V via this E', which can be measured as $-P(H_d|D_{s,i}) \times \log P(D_{s,i})$. Notice that this formula is like entropy, that is, $-P \times \log P$, except that it is $-P_2 \times \log P_1$, where P_2 is the *posterior probability of a model* after a datum is observed whereas P_1 is the *prior-weighted likelihood of the datum* before that datum was observed. Notice also that when there is only one model H, then $P(H|D) = 1$; for that special case (like IR) we have $E' = V = -E$.

3.1.4. Pleasure and tension

Finally, the expression for total pleasure (p_{tot}) can be written as follows for a single instance (i, j, k) of EVE':

$$p_{tot} = p + p' = G \times E + G' \times E',$$

where G and G' are scaling factors that reflect the relative pleasure that a perceiver gets from a unit of success at E and E', respectively. These scaling factors must be established empirically (see Goldilocks functions in Section 3.2). Similarly, total *tension* is the sum of A and V, where these two sources of tension are weighted by factors W_A and W_V, which must be established empirically. Thus, the expression for total tension can be written as follows:

$$t_{tot} = t_A + t_V = W_A \times A + W_V \times V.$$

3.2. Goldilocks functions

Here it is useful to give an example of the trade-off between E and E' that is central to EVE'. The example is illustrated in plots of Goldilocks functions. As the story goes, Goldilocks is a young girl who is walking in the forest when she happens upon a bear family's cottage. The bears are away so Goldilocks enters. She then proceeds to find and taste their bowls of porridge, sit on their chairs, and lie in their beds. The climax of the story comes when the bears return to find Goldilocks still sleeping in Baby Bear's bed, and the reader's tension is resolved when Goldilocks wakes up and runs off.

This story is typical in that its main plot has a beginning (Goldilocks in the forest), middle (Goldilocks in the cottage), and end (Goldilocks running away), as Aristotle wrote that all stories should (Butcher, 1955). In fact, the same structure is also repeated at smaller scales in subplots and subsubplots down to the esthetic atoms of the story, much like the basic sequence of beginning–middle–end in E-V-E' itself. For example, in three subplots of the main plot's middle we find Goldilocks tasting porridge, sitting on chairs, and lying in beds. In each of these subplots we find three subsubplots, for example, with porridge where Goldilocks finds Papa Bear's "too hot," Mama Bear's "too cold," and Baby Bear's "just right,"

The story is useful in exposing the fact that personal preferences must be established empirically, as Goldilocks did, and in expressing the fact that "not too big or too small but just right" applies to almost all esthetic experiences, because of trade-offs that give rise to styles. For example, $G > G'$ models a listener whose style is such that they prefer E to E', so they will enjoy an esthetic experience that is all or mostly E, such as hearing Goldilocks' story repeatedly. These listeners prefer not giving up a unit of E to V even if they get it all back in a unit of E'. Conversely, $G' > G$ models a listener whose style is such that they prefer E' to E, so they will enjoy an esthetic experience that has more E', such as hearing a new story about the Three Bears. These listeners prefer to give up several units of E just to get one unit of E'.

As a concrete example, consider a perceiver listening to a performer who plays only two notes, A or B, where $P(\text{A}) = F$ and $P(\text{B}) = 1 - F$ are the average frequencies at which the notes are played. Assume also that there is only one model M of the performer's preferences in the perceiver's frame of discernment, and this model has $P(\text{A}|M) = P$ and $P(\text{B}|M) = 1 - P$. Then, by the above measures of EVE', the listener has a measure of expectation for note A that is $E_\text{A} = \log P$ and a measure of expectation for note B that is $E_\text{B} = \log(1 - P)$. The notes are actually played with frequencies F and $1 - F$, so the average pleasure p, which weighs E_A and E_B by their relative frequency of occurrence and scaling factors G_A and G_B, is as follows:

$$p = G_\text{A} \times F \times \log P + G_\text{B} \times (1 - F) \times \log(1 - P).$$

Thus, if $P = F$ such that the perceiver's M of the performer's preferences actually matches the performer's preferences, then p is simply a G-weighted negative entropy. Assuming $G_\text{A} = G_\text{B}$, Figure 5a shows that this is a bowl-shaped function with peaks at $P = 0$ and 1. When $G_\text{A} \neq G_\text{B}$, the curve is not symmetric but it is still bowl shaped. Therefore, assuming that E is the only source of pleasure for a listener, she would most enjoy listening to a performer who always plays the same note, A or B. However, according to EVE' there is more to the story, namely, violations of E that provide the listener with opportunities for E', which in turn can bring pleasure-prime. Here, because the listener is assumed to have only one model M, I assume that all violations are completely resolved so the measure of explanation $E' = V = -E$. Thus, pleasure-prime (p') is as follows:

$$p' = -G'_\text{A} \times F \times \log P - G'_\text{B} \times (1 - F) \times \log(1 - P).$$

Now with these two expressions for p and p', if $G_\text{A} = G_\text{B} = G'_\text{A} = G'_\text{B}$, then p' from E' would cancel p from E and the total pleasure $p_\text{tot} = p + p'$ would be a flat line of zero; that is, the listener would not prefer any performer or even any range of performers. In short, if $G_\text{A} = G_\text{B} = G'_\text{A} = G'_\text{B}$, then there would be no style. If instead $G_\text{A} = G_\text{B} > G'_\text{A} = G'_\text{B}$, then the total pleasure would be a bowl shape like Figure 5a, similar to the case of E with no E' (discussed earlier), where the listener would prefer a performer who always plays the same note A or B. However, if $G_\text{A} = G_\text{B} < G'_\text{A} = G'_\text{B}$, then $p_\text{tot} = p + p'$ is an inverted bowl, as shown in Figure 5b for several different G/G' fractions. As we know from Goldilocks' story as well as everyday observations, this is typically the case in esthetic experiences. Therefore, I argue that $G < G'$ for almost any esthetic experience, which means that E' is an important addition to E in EVE'. Moreover, because E' arises from semantic (level B) considerations in Bayesian-theoretic analysis, it is an important addition to entropic (level A) considerations in information-theoretic analysis; and both are needed for esthetic (level C) analysis of styles in media experiences.

However, even this is still not the whole story of Goldilocks-EVE' because different perceivers also have different preferences for different percepts (i.e., notes or songs). For example, a type A listener like Papa Bear may prefer to hear note A whereas a type B listener like Mama Bear may prefer to hear note B, and these preferences will affect the location of peak pleasure on their Goldilocks curves. To illustrate this, I assume two such listeners, one for whom only violations $V_\text{A} = -\log P$ are resolved with pleasure (because he likes note A) and the other for whom only violations $V_\text{B} = -\log(1 - P)$ are resolved with pleasure (because she likes note B). Both listeners are assumed to have $G < G'$, as in Figure 5b. The result, plotted in Figure 5c, shows how these preferences produce Goldilocks functions that are skewed toward low P or high P; that is, there are two different "sweet spots" of the pleasure function for the two different listeners. Notice that the functions
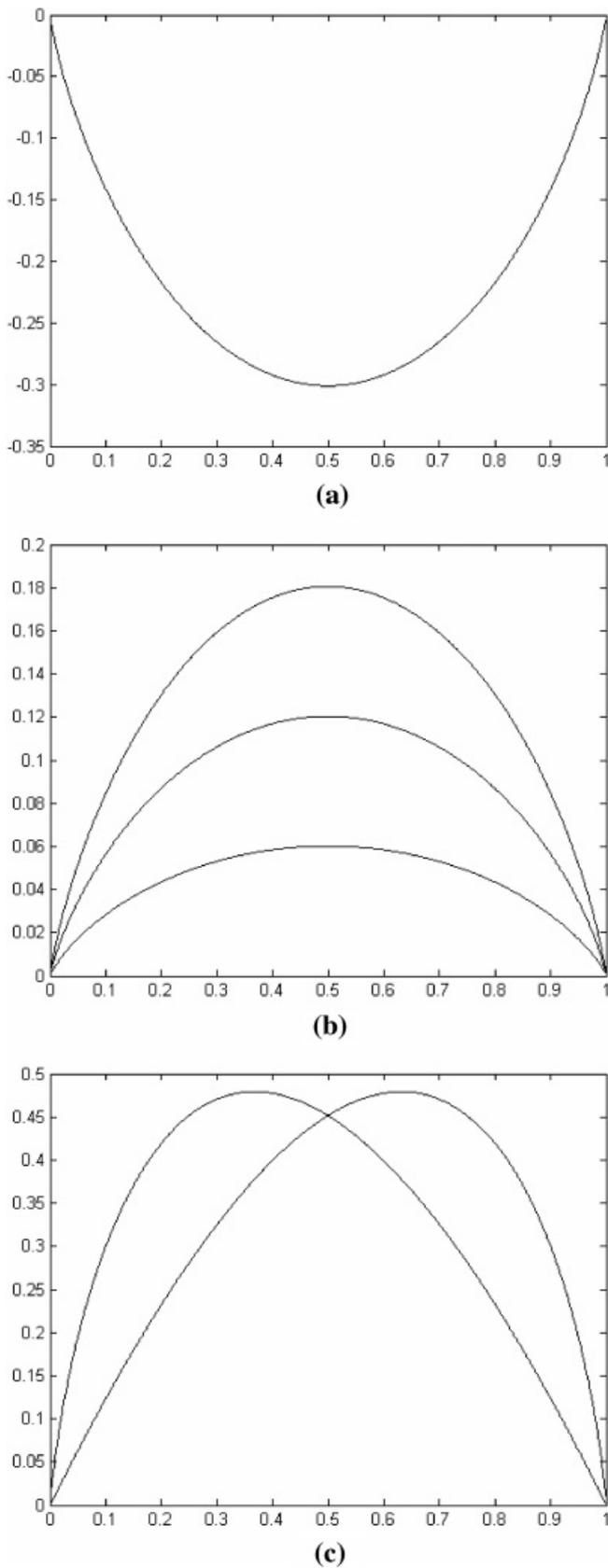
are not peaked at $P = 0$ and 1, because if they were, there would be no possible violations and hence no pleasurable explanations.

This shows how a *perceiver's preferences* for hearing specific notes along with his/her mental models of a *producer's preferences* for playing specific notes combine to create Goldilocks functions in a cultural context where the perceivers and producers interact. Moreover, although the preferences are sometimes different *between people*, they are often similar *among people*, too, such as all Papa Bears or all Mama Bears who have A or B styles.

Thus, the Atoms of EVE′ can be used to develop *theoretical* Goldilocks functions as just described, using the mathematical measures outlined in Section 3.1. In addition, data from field observations or lab experiments can be used to develop *empirical* Goldilocks functions in the same format. The practical value of Goldilocks-EVE′ is that the theoretical functions can help engineers analyze and synthesize styles in design, whereas the empirical functions can help scientists flesh out and test their theories of style. This makes Goldilocks-EVE′ a formal and useful framework for linking theory to practice. In practice, Goldilocks functions might be used to program artificial agents that produce esthetic styles as well as perceive style esthetics or at least act as if they produce and perceive styles like people do.

3.3. Other fun functions

At first glance the Goldilocks functions in Figure 5 may appear similar to the famous Yerkes–Dodson (1908) function shown in Figure 6, which is often cited in human factors engineering (Wickens & Hollands, 2000).

However, there is an important difference: the Yerkes–Dodson (1908) function plots *objective performance* (e.g., task score) versus *objective arousal* (e.g., pulse rate), which is basically the opposite of a Goldilocks function that plots *subjective pleasure* (similar to arousal) versus *subjective success* (similar to performance) measured entropically (probabilistically). Here, I argue that a Goldilocks function is a better format for esthetic analysis, because it treats pleasure as the dependent variable and because it shows that subjective measures of success rather than objective
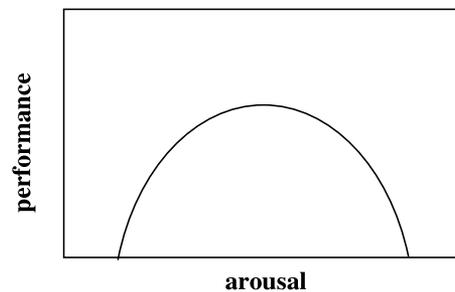
Fig. 5. A Goldilocks functions plot of total pleasure (p_{tot}) versus probability (P).

Fig. 6. The Yerkes–Dodson function for performance versus arousal.

measures of success are what give rise to pleasure (although the objective can affect the subjective, as in game play where the score affects fun). I also posit that pleasure and tension are two different but related feelings, as modeled and measured in the Atoms of EVE′, and that these two feelings are confounded in a single measure of "arousal."

The basic notion of a Goldilocks function for *fun in perception* can also be applied to *flow in production*. That is, producers are also perceiving the products of their efforts (see Fig. 1), and their feelings of *flow* (engagement) are feelings of *fun* (enjoyment) about their performance. Thus, the Atoms of EVE′ is also related to a theory of flow offered by Csikszentmihalyi (1991; see Fig. 7, which is adapted from Salen & Zimmerman, 2004), which plots the *challenge (task)* versus the *person (skill)*, showing the diagonal as flow. This theory of flow (Fig. 7) is basically the same idea as that expressed informally by the Goldilocks story and captured more formally in Goldilocks functions. That is, the Goldilocks functions shown in Figure 5b and 5c show an optimal level of pleasure (fun or flow) on the *y* axis that is not too high or too low on the *x* axis. These Goldilocks functions are more useful in illustrating the shape of the pleasure function, which is not seen in Figure 7 but may be very important as a design guideline. For example, designers would benefit from knowing not only the location of the peak but also how fast fun or flow drops off in each direction away from the peak, because small increases in consumer fun or pleasure would not be worth large increases in designer cost or effort.

Referring to the theory of flow illustrated in Figure 7, it is not clear how one would go about modeling or measuring parameters like task or skill or even flow itself; yet, models and measures are needed for a theory to be formally tested and practically useful. By comparison, Goldilocks-EVE′ does offer formal models and measures to promote scientific testing and engineering uses. Thus, I believe that the Atoms of EVE′ and its Goldilocks functions have important differences from previous theories of human perfor-

mance in the fields of cognitive psychology and human factors engineering.

4. EVALUATING EVE′

The contribution of this paper is to outline a basic theory of esthetic experience and relate it to previous research as well as potential uses. EVE′ is evaluated by addressing its novelty, validity, and applicability.

4.1. Novelty of EVE′

With respect to novelty, the question is whether my thesis about esthetic agreement and my theory about the Atoms of EVE′ are substantially different from previous authors. In reference to earlier sections, Section 1 discussed why esthetics are important to style and how esthetics have been largely ignored in the field of information engineering. Section 2 discussed how my research on sketching can be related to other research on music and how a Bayesian-theoretic approach can extend and improve information-theoretic measures that have been proposed previously. Section 3 discussed how the Atoms of EVE′ can model and measure the atomic components of esthetic experience and thereby integrate and generalize other theories in artificial intelligence, cognitive psychology, and human factors engineering.

4.2. Validity of EVE′

With respect to validity, the question is whether the Atoms of EVE′ can be empirically tested. Three planned experiments are outlined and three hypotheses are offered. The experiments employ dynamic drawings of the type discussed in Section 2, collecting human judgments and comparing these empirical data to theoretical measures of EVE′, much like Dubnov et al. (2004, in press) have done for their computational measure of IR and psychological judgments of EF in music.

Experiment 1 measures a viewer's response continuously, as the dynamic drawing is seen in time, by having the viewer adjust a sliding scale to provide data on feelings that are modeled by EVE′. This is similar to the measurement of EF in music by Dubnov et al. (2004, in press). However, unlike the judgment of EF, which may confound feelings of pleasure and tension, this experiment measures pleasure and tension separately as they are modeled separately in EVE′.

Experiment 2 collects more detailed data on tension in the same dynamic drawings at key points in time. Here, the sliding scale is used to measure when the viewer feels apprehension about what signal he expects next in the sequence or to measure when the signal he has just observed feels like a violation of what he had expected. Afterward the viewer is shown snapshots of the dynamic drawing taken at key times in the esthetic experience, such as the measured peaks of *A* and *V*, and asked to provide retrospective comments. Alternatively, these comments might be collected at
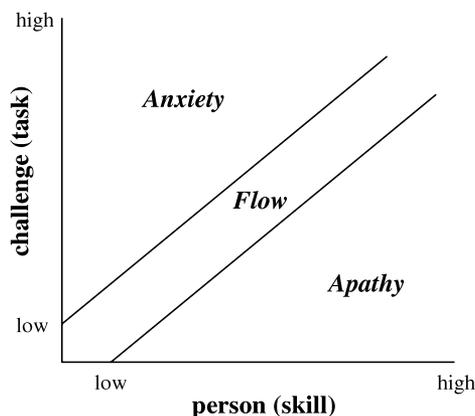

Fig. 7. Csikszentmihalyi's theory of flow as adapted from Salen and Zimmerman (2004).

the time the apprehension or violation occurs in the dynamic drawing, via a think aloud protocol, but this may interfere with the esthetic experience.

Experiment 3 is similar to experiment 2 except it collects data related to pleasure rather than tension. The sliding scale is used to measure when the viewer feels they have succeeded in forming expectations or explanations. Afterward the viewer is shown snapshots of the dynamic drawing taken at key times in the esthetic experience, such as the measured peaks of E and E', and asked to provide retrospective comments. These comments are especially valuable in correlating the explanations measured in experiment 3 with the violations measured in experiment 2, because an explanation may come some time after the violation.

Hypothesis 1 is that pleasure and tension are more than just simple inverses of one another, because pleasure arises from avoiding tension at E as well as from resolving tension at E' and because all tension is not necessarily resolved as pleasure at E'. This is different from a measure like IR that only considers the equivalent of E and hence implicitly assumes $E' = V = -E$. Hypothesis 2 is that the peaks of A and V in experiment 2, which are the theorized causes of tension in experiment 1, will in fact correspond to the indicated peaks of tension in experiment 1. Experiment 2 could also help establish the relative magnitudes of the weighting factors W_A and W_V that determine how A and V combine to cause tension. Hypothesis 3 is that the peaks of E and E' in experiment 3, which are the theorized causes of pleasure in experiment 1, will in fact correspond to the indicated peaks of pleasure in experiment 1. Experiment 3 could also help establish the relative magnitudes of Goldilocks factors G and G' that determine how E and E' combine to cause pleasure. In particular, a strong test of the theory is that $G < G'$ (see Section 3.2).

4.3. Applicability of EVE′

With respect to applicability, the question is whether the Atoms of EVE′ can be applied in engineering and design practice. The answer here is speculative, because the focus of this article is on the theory itself and because the theory has yet to be validated (see earlier). Nevertheless, I believe that a basic theory of esthetic experience can have widespread impact on the design of systems for *information applications*, which involve processing and presenting information to people engaged in activities ranging from game play to business to warfare.

Recent research in the emergent discipline of affective computing (Picard, 2000) has been concerned with developing systems that can recognize the emotional states of their users, which might then allow systems to adapt to the wants and needs of their users. However, still lacking are formal theories of how a media experience gives rise to emotional states in the first place, and such theories are needed if systems are to know *how* to adapt themselves to *what* they are sensing. Other research in the more established disciplines of human factors and human–computer interaction have also been concerned with optimizing the engagement of humans with systems. However, computationally speaking, the focus of these efforts has been on *usability* from an ergonomic perspective rather than *likeability* from an esthetic perspective.

A specific application of EVE′ is in the field of computer game systems (see Salen & Zimmerman, 2004), which are a popular form of entertainment and a practical means of training people in various occupations. To be a good game, a game must be engaging and yet existing theories of fun (Koster, 2005) are lacking specificity and testability. More formal models and measures, like the Atoms of EVE′, are needed to advance the science and practice of game design in particular and information engineering in general. In short, machine systems must be engaging to their human users if they are to be much fun or much use, and effective engagement is largely a matter of *esthetic agreement* between the human and the machine.

As an immediate application to design practice, perhaps the most significant contribution of EVE′ is that it combines information theory and Bayesian theory to model trade-offs between (and within) E and E' in esthetic experiences. Specific measures of E and E', which lead to pleasure, along with related measures of A and V, which lead to tension, are formalized and integrated as the Atoms of EVE′ and illustrated in the form of Goldilocks functions. These contributions make the theory both testable and usable.

Ultimately the Atoms of EVE′ might help foster esthetic agreements between artificially intelligent designers who produce styles and naturally intelligent consumers who perceive styles, making machines *more like* humans and making humans *like more* machines.

REFERENCES

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61(2)*, 183–193.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, *94(2)*, 115–147.

Boden, M. (2004). *The Creative Mind: Myths and Mechanisms*. New York: Routledge.

Brennan, S.E. (1985). Caricature generator: the dynamic exaggeration of faces by computer. *Leonardo*, *18(3)*, 170–178.

Burns, K. (2001). Mental models of line drawings. *Perception*, *30(6)*, 1249–1261.

Burns, K. (2004). Creature double feature: on style and subject in the art of caricature. *Proc. AAAI Fall Symp. Style and Meaning in Language, Music, Art and Design*, pp. 7–14. AAAI Technical Report FS-04-07.

Butcher, S. (1955). *Aristotle Poetics*. New York: Dover.

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*. New York: Harper Collins.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

Do, E.Y.-L. (2002). Drawing marks, acts, and reacts: toward a computational sketching interface for architectural design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *16(2)*, 149–171.

Dubnov, S. (2006). *Musical structure by information rate*. Unpublished manuscript.

Dubnov, S., McAdams, S., & Reynolds, R. (2004). Predicting human reactions to music on the basis of similarity structure and information theoretic measures of sound signals. *Proc. AAAI Fall Symp. Style and Meaning in Language, Music, Art and Design*, pp. 37–40. AAAI Technical Report FS-04-07.

Dubnov, S., McAdams, S., & Reynolds, R. (in press). Structural and affective aspects of music from statistical audio signal analysis. *Journal of the American Society for Information Science and Technology*.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review*, *112(1)*, 243–252.

Forbus, K.D. (2004). Qualitative spatial reasoning about sketch maps. *AI Magazine*, *Fall*, 61–72.

Jupp, J., & Gero, J. (2003). Towards a computational analysis of style in architectural design. *Proc. IJCAI Workshop on Computational Approaches to Style Analysis and Synthesis*, pp. 1–10, Acapulco, Mexico.

Kirsch, J.L., & Kirsch, R.A. (1986). The structure of paintings: formal grammar and design. *Environment and Planning B*, *13(1)*, 163–176.

Knill, D., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Koster, R. (2005). *A Theory of Fun for Game Design*. Scottsdale, AZ: Paraglyph Press.

Matisse, H. (1995). *Drawings: Themes and Variations*. New York: Dover.

McCorduck, P. (1991). *Aaron's Code: Meta-Art, Artificial Intelligence and the Work of Harold Cohen*. New York: W.H. Freeman.

Mithen, S. (1996). *The Prehistory of the Mind: The Cognitive Origins of Art and Science*. New York: Thames and Hudson.

Mokhtarian, F., & Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8(1)*, 34–43.

Oatley, K., & Johnson-Laird, P. (1987). Toward a cognitive theory of emotions. *Cognition and Emotion*, *1(1)*, 29–50.

Picard, R. (2000). *Affective Computing*. Cambridge, MA: MIT Press.

Ramachandran, V.S., & Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, *6(1)*, 15–51.

Resnikoff, H.L. (1985). *The Illusion of Reality: Topics in Information Science*. New York: Springer.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology*, *19(3)*, 473–497.

Rhodes, G., & McLean, I.A. (1990). Distinctiveness and expertise effects with homogeneous stimuli: towards a model of configural coding. *Perception*, *19(4)*, 773–794.

Richards, W. (1988). *Natural Computation*. Cambridge, MA: MIT Press.

Richards, W., Dawson, B., & Whittington, D. (1988). Encoding shape by curvature extrema. In *Natural Computation* (Richards, W., Ed.), pp. 83–98. Cambridge, MA: MIT Press.

Rosch, E. (1978). Principles of categorization. In *Cognition and Categorization* (Rosch, E., & Lloyd, B.B., Eds.). Hillsdale, NJ: Erlbaum.

Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press.

Shannon, C., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Sosa, R., & Gero, J. (2006). *A computational framework to investigate creativity and innovation in design*. Manuscript submitted for publication.

Stiny, G., & Mitchell, W. (1978). The Palladian grammar. *Environment and Planning B*, *5(1)*, 5–18.

Wachtel, E. (1993). The first picture show: cinematic aspects of cave art. *Leonardo*, *26(2)*, 135–140.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance*. New York: Prentice–Hall.

Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18(3)*, 459–482.

Kevin Burns is a Principal Scientist at the MITRE Corporation. He holds engineering degrees from the Massachusetts Institute of Technology where he also studied cognitive science and media arts. His research interest is in computational modeling of cognitive processing, including strategic decisions and visual perception, to improve the design of decision support systems.