



Free-Energy Model of Emotion Potential: Modeling Arousal Potential as Information Content Induced by Complexity and Novelty

Hideyoshi Yanagisawa*

Design Engineering Laboratory, Department of Mechanical Engineering, The University of Tokyo, Tokyo, Japan

Appropriate levels of arousal potential induce hedonic responses (i.e., emotional valence). However, the relationship between arousal potential and its factors (e.g., novelty, complexity, and uncertainty) have not been formalized. This paper proposes a mathematical model that explains emotional arousal using minimized free energy to represent information content processed in the brain after sensory stimuli are perceived and recognized (i.e., sensory surprisal). This work mathematically demonstrates that sensory surprisal represents the summation of information from novelty and uncertainty, and that the uncertainty converges to perceived complexity with sufficient sampling from a stimulus source. Novelty, uncertainty, and complexity all act as collative properties that form arousal potential. Analysis using a Gaussian generative model shows that the free energy is formed as a quadratic function of prediction errors based on the difference between prior expectation and peak of likelihood. The model predicts two interaction effects on free energy: that between prediction error and prior uncertainty (i.e., prior variance) and that between prediction error and sensory variance. A discussion on the potential of free energy as a mathematical principle is presented to explain emotion initiators. The model provides a general mathematical framework for understanding and predicting the emotions caused by novelty, uncertainty, and complexity. The mathematical model of arousal can help predict acceptable novelty and complexity based on a target population under different uncertainty levels mitigated by prior knowledge and experience.

Keywords: emotion, free energy, bayes, gaussian generative models, arousal, uncertainty

OPEN ACCESS

Edited by:

Emili Balaguer-Ballester,
Bournemouth University,
United Kingdom

Reviewed by:

Casper Hesp,
Amsterdam Brain and Cognition
(ABC), Netherlands
Dimitris Pinotsis,
City University of London,
United Kingdom

*Correspondence:

Hideyoshi Yanagisawa
hide@mech.t.u-tokyo.ac.jp

Received: 21 April 2021

Accepted: 26 October 2021

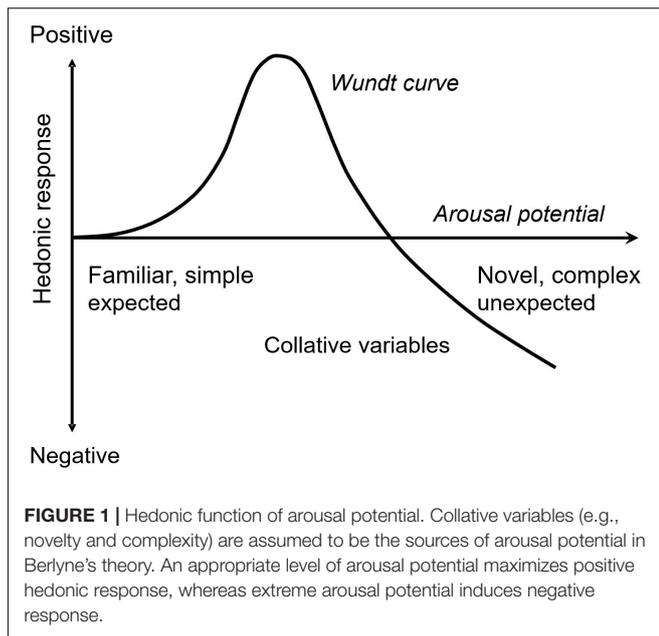
Published: 19 November 2021

Citation:

Yanagisawa H (2021) Free-Energy Model of Emotion Potential: Modeling Arousal Potential as Information Content Induced by Complexity and Novelty. *Front. Comput. Neurosci.* 15:698252. doi: 10.3389/fncom.2021.698252

INTRODUCTION

Acceptance of novelty and complexity of incoming information depend on receivers' emotions. Berlyne considered both novelty and complexity to be *collative properties* that are sources of arousal potential (Berlyne, 1970). He suggested that an appropriate level of arousal potential would induce a positive hedonic response, but extreme arousal potentials induce negative responses. The hedonic function of the arousal potential shapes an inverse U, the so-called *Wundt curve* (**Figure 1**). Arousal (i.e., intensity) and valence (i.e., positivity or negativity) are known to be dominant emotional dimensions (Russell, 1980; Lang, 1995). They comprise the *core affect* and are correlated with neural



activity in specific brain regions (i.e., orbitofrontal cortex and amygdala, respectively) (Wilson-Mendenhall et al., 2013). The Wundt curve formulates a relationship between the core affect when the hedonic response is considered to be a valence component. In this sense, the Wundt curve is a valence function of arousal explaining the condition of positive acceptance of collative properties, such as novelty and complexity.

The mathematical formulation of arousal provides a precise prediction of the acceptance condition and the effect of certain factors of arousal in a general manner. Yanagisawa et al. formalized arousal using the Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) of the Bayesian posterior, termed *information gain*, and confirmed that it corresponds to human surprise via participants' responses using event-related potential, P300, and subjective reports of surprise to novel stimuli (Yanagisawa et al., 2019; Ueda et al., 2021). The information gain is equivalent to Bayesian surprise representing visual salience (Itti and Baldi, 2009). Yanagisawa et al. formalized the Wundt curve as a summation of reward and aversion functions, representing sigmoidal functions of the information gain (Yanagisawa et al., 2019; Miyamoto and Yanagisawa, 2021). Novelty, however, is only one of several sources of arousal potential (i.e., collative properties). Berlyne exemplified other collative variables, namely complexity and uncertainty. Experimental studies have shown that both complexity and novelty affect hedonic responses (valence) in art and music (Marin et al., 2016; Miyamoto and Yanagisawa, 2021), design aesthetics (Hekkert et al., 2003; Hung and Chen, 2012), and food preference (Giacalone et al., 2014). Information gain is content acquired from novelty; however, it does not involve complexity and uncertainty. To the best of my knowledge, a mathematical formulation of arousal potential caused by all these factors is yet to be established.

The present study proposes an arousal model that includes complexity and uncertainty as well as novelty. This model

provides a general mathematical framework to explain arousal, the primary emotion dimension. Information content formulated as negative log probability of sensory stimuli has been hypothesized as being processed when one perceives sensory stimuli. In Section “Surprisal and Free Energy in Perception,” I mathematically demonstrate that information content is equivalent to minimized (variational) free energy or surprisal in perception. Free energy, which is originally a thermodynamic quantity in statistical physics, has recently been applied to the neuroscience field as a significant information quantity. Friston et al. proposed the free-energy principle that unifies brain theories (Friston, 2010). The principle of free-energy minimization mathematically explains cognitive components, such as perception and action (Friston et al., 2006). In this context, free energy represents a prediction error and uncertainties of signals in a Bayesian brain model (Knill and Pouget, 2004).

Recently, the free energy has been applied as a key information theoretic quantity to explain emotional states. Joffily and Coricelli (2013) formalized emotional valence as the rate of change of free-energy over time; they suggest that decreasing and increasing free energy corresponds to positive and negative emotion, respectively. The idea behind such a formulation is that emotional states reflect changes in the uncertainty about the somatic consequences of action (Wager et al., 2015; Seth and Friston, 2016; Clark et al., 2018). Hesp et al. (2021) formalized emotional valence as increased and decreased confidence (or precision) in one's action model, termed *affective charge*. It is estimated as a higher metacognitive component of a deep hierarchical Bayesian model based on active inference that selects an action policy minimizing (expected) free energy. Although these studies formalize emotional valence and its dynamics, a formal representation of emotional arousal in relation to the free energy has not yet been proposed. Furthermore, the relationship between the free energy and emotional arousal characterized by collative properties (e.g., novelty and complexity) of sensory stimuli have not been elucidated.

In Section “Free Energy as Emotional Arousal Potential,” I mathematically demonstrate that free energy comprises the summation of information from both perceived novelty and perceived complexity of external stimuli and that this complexity is mathematically equivalent to minimized uncertainty. In Section “Analysis of Free Energy using a Gaussian Generative Model,” I derive a functional model of free energy using a Gaussian generative model. Using the function model, I predict how predictability (i.e., prior uncertainty and prediction error) and sensory-data variation (inverse precision) affect arousal potential. The model prediction of the predictability effect explains how personal knowledge and experience as factors of prior uncertainty change the optimal arousal potential (a peak of the Wundt curve, **Figure 1**). This helps us understand the personal differences caused by prior knowledge and experience, such as between novices and experts, on the preference of novelty and complexity. I demonstrate that the prior uncertainty effect is equivalent to the effect of sensory variance on arousal.

SURPRISAL AND FREE ENERGY IN PERCEPTION

The idea at the core of our model is that the total information content to be processed in the brain after perceiving sensory stimuli represents the potential cognitive load. This load functions as a source of emotional arousal (i.e., Berlyne's arousal potential) that operates as an initiator of subsequent emotions. According to information theory (Shannon et al., 1949), information content is defined by the negative of the logarithm of probability ($-\log p$), where p is an event's probability. Therefore, I consider belief probability distributions in a situation where one obtains information content by perceiving the external world (e.g., physical properties like shape and color). Here, perception is defined as the estimation of the causes of sensory stimuli. Sensory stimuli are coded to neural activities in the brain (e.g., the firing rate of certain neuronal populations) *via* sensory organs (Yanagisawa, 2016). I term the neural codes as *sensory data* and assume the data to be a random variable, x , that follows a certain probability distribution. In our study, instead of true distributions, I assume that the brain has belief distributions, $p(x)$, where $p(x)$ is marginalized by the possible causes of sensory data represented as continuous random variables, $\theta \in R$ (e.g., physical properties (θ) like shape features and color that cause visual sensory data, x , *via* light).

I assume that a joint probability distribution between sensory data and its causes, $p(x, \theta)$, is learned based on past experiences of perceiving varied sensory data throughout one's life. The statistical model is a generative model (Friston et al., 2006; Buckley et al., 2017). I assume that the generative model comprises observation model $p(x|\theta)$ and the prior $p(\theta)$.

$$p(x, \theta) = p(x|\theta)p(\theta). \quad (1)$$

With the generative model, one can predict sensory-data distributions under the assumption of a cause, θ . For example, one knows that the color "red" causes "this" sensory data from previous experience; therefore, one can imagine (generate) "red." The exact Bayesian posterior, $p(x|\theta)$, represents perception of the cause, given data x (Knill and Pouget, 2004). This is usually difficult to compute. The variational Bayesian method then aims to find sufficient approximations of the exact posterior (Fox and Roberts, 2012). I assume an approximate posterior, $q(\theta)$, representing recognition density of a cause (Buckley et al., 2017). The deviation of $q(\theta)$ from the posterior $p(\theta|x)$ is evaluated using the KL divergence:

$$D_{KL}(q(\theta)||p(x|\theta)) = \langle \ln q(\theta) - \ln p(x|\theta) \rangle_{q(\theta)}. \quad (2)$$

I want to minimize the KL divergence to approximate the exact Bayesian posterior. However, the exact posterior is unknown. Here, I adopt the free-energy principle (Friston et al., 2006; Friston, 2010) to approximate the posterior by minimizing variational free energy in perception. The free-energy principle suggests that any self-organizing system (i.e., brain) that is at equilibrium with its environment must minimize its free energy. It explains perception as the minimization of free energy, which

comprises a joint occurrence of the sensory input and its cause (i.e., internal energy) and the negative entropy of the approximate distribution. It represents the difference between the recognition density and the generative model averaged over the recognition density in term of information (i.e., negative log probability).

$$F = -\langle \ln p(\theta, x) \rangle_{q(\theta)} + \langle \ln q(\theta) \rangle_{q(\theta)}. \quad (3)$$

With decomposition $p(x, \theta) = p(\theta|x)p(x)$, the free energy is expressed as a summation of KL divergence and surprisal (Friston et al., 2006). The surprisal is the negative log of the marginalized likelihood representing model evidence, $p(x) = \int p(\theta, x)d\theta$.

$$F = D_{KL}(q(\theta)||p(x|\theta)) - \ln p(x). \quad (4)$$

Because surprisal is independent from the recognition density, and the KL divergence is greater than zero, the lower bound of free energy is surprisal: $F \geq -\ln p(x)$. When recognition density is a sufficient approximation of the exact Bayesian posterior, the KL divergence becomes zero, and the free energy is equivalent to the surprisal. This approximation process is connected with perception using the recognition density according to the free-energy principle (Friston et al., 2006). Therefore, the surprisal corresponds to the free energy minimized by perceiving a sensory stimulus. Free energy is considered to represent information content that the brain potentially processes using recognition density to perceive sensory stimuli.

FREE ENERGY AS EMOTIONAL AROUSAL POTENTIAL

Novelty and Uncertainty

With another decomposition $p(x, \theta) = p(x|\theta)p(\theta)$, free energy is expressed as a summation of another two terms:

$$F = D_{KL}(q(\theta)||p(\theta)) + \langle -\ln p(x|\theta) \rangle_{q(\theta)}. \quad (5)$$

When the recognition distribution is a sufficient approximation of the exact Bayesian posterior, $q(\theta) \simeq p(\theta|x)$, the first term is equivalent to *Bayesian surprise*, and the KL diverges from posterior to prior. Bayesian surprise refers to the difference of prior information content from posterior information content averaged over posterior information (Itti and Baldi, 2009). It represents information content gained by perceiving sensory data x . In our previous study (Yanagisawa et al., 2019), we defined Bayesian surprise as information gain because it represents information content gained from incoming sensory data. I considered information gain to be an index of novelty and experimentally verified that it corresponds to human surprise (i.e., high arousal emotion) induced by unexpected and novel stimuli (Sekoguchi et al., 2019; Yanagisawa et al., 2019; Miyamoto and Yanagisawa, 2021; Ueda et al., 2021). Information gain refers to information content elicited by incongruity between prior expectation and posterior perception (Friston et al., 2015). Prior expectations are formed based on past experiences. One expects an event if it has been frequently observed in one's life. Therefore,

an unfamiliar event, which has rarely been observed in the past, causes incongruity between prior expectation and posterior perception and surprises the person.

The second term is the *inverse accuracy* (Penny et al., 2010; Friston et al., 2017). From the context of human cognition, the accuracy is how accurately one's generative model prediction fits incoming sensory data, x . Thus, the term represents the extent to which the sensory data support the hypothetical estimates of the generative model in one's brain. Lower accuracy implies greater uncertainty in the estimate. Someone who believes their own generative model perceives uncertainty of the cause of incoming sensory data when accuracy is low. Therefore, from a first-person perspective, the inverse accuracy is interpreted as the perceived *uncertainty* of the cause of the sensory data.

Prior/Posterior Uncertainties and Unfamiliarity

The second term of free energy in Eq. (5), the inverse accuracy $\langle -\ln p(x|\theta) \rangle_{q(\theta)}$, represents the uncertainty of a cause. However, this uncertainty is distinguished from other types of uncertainties represented by variances (or inverse precisions) of signals. In Bayesian brain hypothesis (Knill and Pouget, 2004) and free energy principle (Feldman and Friston, 2010), the term "uncertainty" means the variances of neural signals. From that meaning of uncertainty, I define other types (i.e., *prior* and *posterior uncertainties*), which are represented by prior and posterior entropies, respectively, as uncertainties generated by precision of expectation or estimation:

$$\text{Prior uncertainty} : < -\ln p(\theta) >_{p(\theta)}, \quad (6)$$

$$\text{Posterior uncertainty} : < -\ln p(\theta|x) >_{p(\theta|x)}. \quad (7)$$

Prior uncertainty refers to the extent to which one's belief of cause is uncertain before observing sensory data. For instance, if one is not familiar with paintings, one's prior uncertainty about paintings is high. Conversely, if one learns about paintings and becomes familiar with them, one's prior uncertainty decreases. During this learning process, the prior updates to the posterior, and the decrease of uncertainty averaged over the posterior is equivalent to the information gain (Bayesian surprise):

$$D_{KL}(p(\theta|x)||p(\theta)) = \langle -\ln p(\theta) - (-\ln p(\theta|x)) \rangle_{p(\theta|x)}. \quad (8)$$

Imagine that one is in an art gallery, but one does not know which exhibition one is in. One could form a generative model where θ corresponds to alternative categories of art, and x are the alternative paintings one could encounter. Upon observing a painting, x , one can update one's prior beliefs about which exhibition one is in to one's posterior beliefs about the exhibition (θ) one has found. In this situation, the more unfamiliar painting one observes, the more one is surprised and learned. Here, there are two kinds of unfamiliarity: *unfamiliarity of a category* (e.g., paintings as a category of art) and *unfamiliarity of an individual* (e.g., a specific painting). The former corresponds to

prior/posterior uncertainty and the latter to information gain elicited by novelty or Bayesian surprise.

Uncertainty Reduction in Bayesian Updating

The growing body of psychophysical evidence has shown that computations of perception, sensory motor control, and learning are Bayes' optima that lead to the Bayesian coding hypothesis (Knill and Pouget, 2004; Körding and Wolpert, 2004). The free-energy principle explains the Bayesian coding hypothesis *via* free-energy minimization (Friston, 2010). Furthermore, a variety of animals in different ecological contexts behave in manners consistent with predictions of Bayesian updating models (Valone, 2006). Here, I demonstrate how an uncertainty (the second term of Eq. 5) minimizes in a Bayesian updating process.

Using the definition of conditional probability, $p(x|\theta)p(\theta) = p(x)p(\theta|x)$, the free-energy representation takes two forms: summation of Bayesian surprise and the uncertainty and a summation of KL divergence and surprisal.

$$\begin{aligned} F(x) &= D_{KL}(q(\theta)||p(\theta)) + \langle -\ln p(x|\theta) \rangle_{q(\theta)} \\ &= D_{KL}(q(\theta)||p(\theta|x)) - \ln p(x). \end{aligned} \quad (9)$$

Here, the uncertainty regards an expected surprise (i.e., the information content) of sensory data x given a cause, θ , $-\ln p(x|\theta)$, averaged over a recognition distribution, $q(\theta)$. From Eq. 9, the uncertainty is a summation of three terms: KL divergence, (negative) Bayesian surprise, and surprisal:

$$\begin{aligned} U(x) &:= \langle -\ln p(x|\theta) \rangle_{q(\theta)} \\ &= D_{KL}(q(\theta)||p(\theta|x)) - D_{KL}(q(\theta)||p(\theta)) - \ln p(x). \end{aligned} \quad (10)$$

I assume that a recognition is set to a prior, $q(\theta) = p(\theta)$, before obtaining data x . The uncertainty is equivalent to a summation of KL divergence and surprisal because the second term (i.e., Bayesian surprise) is zero.

$$\begin{aligned} U_{pri} &:= \langle -\ln p(x|\theta) \rangle_{p(\theta)} \\ &= D_{KL}(p(\theta)||p(\theta|x)) - \ln p(x). \end{aligned} \quad (11)$$

In Bayesian estimation or recognition of θ given x , the recognition distribution is approximated to the posterior, $q(\theta) \simeq p(\theta|x)$. As a result, the uncertainty is approximated to the summation of negative Bayesian surprise and surprisal because the first term, KL divergence, is approximated to be zero.

$$\begin{aligned} U_{post} &:= \langle -\ln p(x|\theta) \rangle_{p(\theta|x)} \\ &= -D_{KL}(p(\theta|x)||p(\theta)) - \ln p(x). \end{aligned} \quad (12)$$

The difference in uncertainty before and after the recognition forms a summation of the KL divergence and the Bayesian

surprise. The difference is always greater than zero because both KL divergence and Bayesian surprise are greater than zero.

$$U_{pri} - U_{post} = D_{KL}(p(\theta)||p(\theta|x)) + D_{KL}(p(\theta|x)||p(\theta)) \geq 0. \quad (13)$$

Therefore, the uncertainty always decreases in the recognition process from the prior to the posterior, $q(\theta) : p(\theta) \rightarrow p(\theta|x)$.

The surprisal, the second term of Eq. 12, is a negative log likelihood marginalized with a prior, $-\ln p(x) = -\ln \langle p(x|\theta) \rangle_{p(\theta)}$. When a prior is updated to a posterior, $p(\theta) : p(\theta) \rightarrow p(\theta|x)$, the posterior uncertainty is equivalent to the surprisal of the model prediction with posterior; it corresponds to a negative log likelihood of a predictive distribution.

$$U^* := -\ln \langle p(x|\theta) \rangle_{p(\theta|x)} = -\ln p^*(x), \quad (14)$$

where $p^*(\cdot) = \langle p(\cdot|\theta) \rangle_{p(\theta|x)}$ is a predictive distribution. By applying Jensen's inequality, I find that the updated uncertainty, U^* , corresponds to the lower term of the uncertainty, U . Thus, the uncertainty decreases in the prior updating process.

$$\begin{aligned} U_{post} &= \langle -\ln p(x|\theta) \rangle_{p(\theta|x)} \\ &\geq -\ln \langle p(x|\theta) \rangle_{p(\theta|x)} = U^*. \end{aligned} \quad (15)$$

Jensen's inequality holds for any priors. Therefore, for any priors, the uncertainty is always decreased by Bayesian prior updating.

$$\forall p(\theta), \langle -\ln p(x|\theta) \rangle_{p(\theta)} \geq \ln \langle -p(x|\theta) \rangle_{p(\theta)}. \quad (16)$$

Hence, the uncertainty is always decreased by both the recognition (perception) and the prior updating (learning) process. By alternately repeating the two processes and given the data from the same stimuli source, the uncertainty minimized to the surprisal of true distribution of data.

Uncertainty Convergence to Perceived Complexity

Consider the case where one obtains n data D^n with respect to m values, $x_i \in R$ ($i = 1, \dots, m$). The empirical probability for each value is $f(x_i)$, ($i = 1, 2, \dots, m$), where f is the empirical distribution of obtained data x . From Eq. 14, the lower-bound uncertainty of a prior updating is the negative log likelihood of the predictive distribution with respect to the data, D^n . I find that the lower-bound uncertainty corresponds to the (n times) cross entropy (Chen et al., 2012).

$$\begin{aligned} U^*(D^n) &= -\ln p^*(D^n) \\ &= -\ln \prod_{i=1}^m p^*(x_i)^{n \cdot f(x_i)} \\ &= -n \sum_{i=1}^m f(x_i) \ln p^*(x_i) \\ &= n \langle -\ln p^*(x) \rangle_{f(x)}, \end{aligned} \quad (17)$$

where $p^*(x) = \langle p(x|\theta) \rangle_{p(\theta|D^n)}$ is a predictive distribution given data D^n of a random variable, x , and $p^*(x_i)$ is the predictive probability of value x_i . The cross entropy is a summation of a KL divergence and entropy.

$$\langle -\ln p^*(x) \rangle_{f(x)} = D_{KL}(f(x)||p^*(x)) + H(f(x)). \quad (18)$$

The KL divergence represents the deviation of the model prediction from the data (empirical) distributions, which is always greater than zero. Thus, the lower bound of the cross entropy is the entropy of the sample distribution.

$$\langle -\ln p^*(x) \rangle_{f(x)} \geq H(f(x)). \quad (19)$$

As discussed, the lower-bound uncertainty implies that the uncertainty after a prior is replaced with the Bayesian posterior; thus, the uncertainty monotonically decreases with the Bayesian updating of repeated data input given from the same stimuli source (i.e., a cause target of perception). Therefore, the cross-entropy converges to the entropy of the distribution of the data. In other words, when the model prediction (predictive distribution) fits the data distribution by updating the prior, the KL divergence is minimized to zero, and the cross-entropy approaches the entropy of the data distribution. By the statistical law of great numbers, the empirical distribution converges to the true distribution of the stimuli source. Therefore, after sufficiently observing a stimuli source (a target of perception or a cause of sensory data), the uncertainty converges to the entropy.

The entropy reflects the disorder or complexity of a data distribution. Therefore, the uncertainty minimized by Bayesian updating, given sufficient input data, is interpreted as the *perceived complexity* of the data. Previous studies have shown that the curvature entropy of a shape correlates to the complexity perceived by human evaluators (Ujii et al., 2012; Okano et al., 2020). Those experimental results support the interpretation that the uncertainty minimized by observing sufficient data (i.e., evaluating a shape) represents perceived complexity.

Summary

Table 1 shows a summary of free-energy decomposition and its meanings in the context of emotion arousal potential. Free energy, representing sensory surprisal, is equivalent to the summation of information gain or Bayesian surprise (e.g., novelty and incongruity). The inverse accuracy and the uncertainty of causes are the negative log-likelihoods weighted over recognition. After observing sufficient sensory data from a stimuli source, the uncertainty represents perceived complexity. Novelty, incongruity, complexity, and the uncertainty are the collative properties (i.e., sources of arousal potential) that Berlyne exemplified. This concordance suggests that the free energy explains the arousal potential induced by collative properties.

TABLE 1 | Summary of semantics of free energy in emotion arousal potential.

Mathematical terms	Information contents	Arousal potential
$-\ln p(x) \geq nH(x)$	Free energy, surprisal, negative log likelihood	Arousal potential of collative variables
$D_{KL}(q(\theta) p(\theta)) \geq 0$	Bayesian surprise, information gain, prior/posterior uncertainty	Novelty, incongruity, unfamiliarity
$\langle -\ln p(x \theta) \rangle_{q(\theta)} \geq nH(x)$	Inverse accuracy entropy	Uncertainty, Perceived complexity

Arousal potential is formulated as free energy minimized in perception: surprisal. Free energy comprises the summation of information gained from novelty, incongruity, and unfamiliarity and inverse accuracy representing perceived uncertainty and perceived complexity.

ANALYSIS OF FREE ENERGY USING A GAUSSIAN GENERATIVE MODEL

Gaussian Free-energy Function

In the free-energy principle literature, Gaussian posterior, the so-called Laplace approximation, is assumed (Friston et al., 2006, 2007; Buckley et al., 2017). The limitation of this approximation is that it can be inaccurate when the mode is not near the majority of the probability mass. However, the Gaussian form is useful because the parameters provide an exact form of prediction errors (distance between means) and precisions (inverse variance). Here, I assume Gaussian distributions for the prior and posterior to analyze how predictability and sensory variance (i.e. inverse precisions) affect free energy and its components: the information gain and the uncertainty. Gaussian distribution is a probability distribution formed by two parameters: mean and variance. The difference in the mean of prior and likelihood represents prediction errors. Variances of the prior and the likelihood represent the prior uncertainty and the sensory variance, respectively. Hence, a Gaussian distribution can be used to explicitly analyze how predictability (prediction errors and prior uncertainty) and sensory variance affect free energy by partial differentiations of Gaussian parameters. To be simple, I assume that a causal state of stimuli does not change over time.

I assume that one obtains n sensory data samples, $x^n = (x_1, x_2, \dots, x_n)$, from a stimulus source that conforms to a Gaussian distribution, $N(\mu, \sigma^2)$, and has a generative model, $p(x, \mu) = p(x|\mu)p(\mu)$. Here, the mean (μ) represents a perception target (i.e., the cause of sensory data). When the conditional probability, $p(x|\mu)$, conforms to a normal distribution, the likelihood function of a cause (μ) given the sensory data is:

$$\begin{aligned}
 p(x^n|\mu) &= \prod_{i=1}^n p(x_i|\mu) \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi s_l}} \exp\left[-\frac{(x_i - \mu)^2}{2s_l}\right] \\
 &= \left(\frac{1}{\sqrt{2\pi s_l}}\right)^n \exp\left[-\frac{1}{2s_l} \{n(\mu - \bar{x})^2 + nS\}\right], \quad (20)
 \end{aligned}$$

where s_l is the variance of $p(x|\mu)$ (i.e., sensory variance), \bar{x} is the sample mean, and S is the sample variance. I assume that the sensory variance estimates the stimulus's source variance, $s_l \simeq \sigma^2$.

The sensory data's free energy, using the generative model, is represented as follows:

$$F(x^n) = D_{KL}(q(\mu)||p(\mu)) + \langle -\ln p(x^n|\mu) \rangle_{q(\mu)}, \quad (21)$$

where $p(\mu)$ is a prior, and $q(\mu)$ is the recognition distribution of μ . When the prior follows the Gaussian distribution, $N(\eta, s_p)$, the Bayesian theorem estimates the posterior, $p(\mu|x^n)$, that follows a Gaussian distribution, $N(\eta_{post}, s_{post})$, where

$$\text{Average : } \eta_{post} = \frac{ns_p\bar{x} + s_l\eta}{ns_p + s_l}, \text{ Variance : } s_{post} = \frac{s_p s_l}{ns_p + s_l}. \quad (22)$$

When the recognition distribution is approximated to the Bayesian posterior, the free energy is expressed as a summation of the information gain (i.e., Bayesian surprise) and the uncertainty.

$$\begin{aligned}
 D_{KL}(N(\eta_{post}, s_{post})||N(\eta, s_p)) + \langle -\ln p(x|\mu) \rangle_{N(\eta_{post}, s_{post})} \\
 = G + U. \quad (23)
 \end{aligned}$$

The information gain representing human surprise to novelty is derived as a quadratic function of prediction error, $\delta = \bar{x} - \eta$, which represents the difference between the prior mean and the likelihood peak, with coefficients as functions of prior variance (i.e., prior uncertainty), s_p , and sensory variance, s_l (Yanagisawa et al., 2019).

$$G_n = D_{KL}(N(\eta_{post}, s_{post})||N(\eta, s_p)) = A_G \delta^2 + B_G$$

$$A_G := \frac{n^2 s_p}{2(ns_p + s_l)^2}, B_G := -\frac{1}{2} \left(\ln \frac{s_l}{ns_p + s_l} + \frac{ns_p}{ns_p + s_l} \right). \quad (24)$$

The uncertainty is also derived as a quadratic function of prediction error (see the **Appendix** for the derivation).

$$U_n := \langle -\ln p(x|\mu) \rangle_{N(\eta_{post}, s_{post})} = A_U \delta^2 + B_U$$

$$A_U := \frac{ns_l}{2(ns_p + s_l)^2}, B_U := \frac{n}{2} \left(\frac{s_p}{ns_p + s_l} + \ln 2\pi s_l + \frac{S}{s_l} \right). \quad (25)$$

Therefore, free energy, as the summation of the information gain (novelty) and the uncertainty of μ , is also a quadratic

function of the prediction error with coefficients that are functions of the variances, s_p and s_l .

$$F_n := G_n + U_n = A_F \delta^2 + B_F$$

$$A_F := \frac{1}{2} \frac{n}{ns_p + s_l},$$

$$B_F := \frac{1}{2} \left\{ \ln(ns_p + s_l) + (n-1) \ln s_l + n \ln 2\pi + n \frac{S}{s_l} \right\}. \quad (26)$$

Convergence to Entropy and Perceived Complexity

Here, I demonstrate how the uncertainty converges to entropy to represent perceived complexity. A prior updated by Bayesian theorem with n samples follows $N(\eta_n, s_n)$ with an average of $\eta_n = (ns_p \bar{x} + s_l \eta) / (ns_p + s_l)$ and a variance of $s_n = s_p s_l / (ns_p + s_l)$. Given additional m samples from the stimulus source, the posterior follows $N(\eta_{n+m}, s_{n+m})$ with an average of $\eta_{n+m} = \frac{(n+m)s_p \bar{x} + s_l \eta}{(n+m)s_p + s_l}$ and a variance of $s_{n+k} = \frac{s_p s_l}{(n+m)s_p + s_l}$. With the prior and posterior, the information gain (G_{n+m}) and the uncertainty (U_{n+m}) can be derived:

$$G_{n+m} := D_{KL}(N(\eta_{n+m}, s_{n+m}) || N(\eta_n, s_n))$$

$$= \frac{1}{2} \left[m^2 \frac{s_p s_l}{ns_p + s_l} \left(\frac{\delta^2}{(n+m)s_p + s_l} \right)^2 - \ln \frac{ns_p + s_l}{(n+m)s_p + s_l} + \frac{ns_p + s_l}{(n+m)s_p + s_l} - 1 \right], \quad (27)$$

$$U_{n+m} := \langle -\ln p(x^{n+m} | \mu) \rangle_{N(\eta_{n+m}, s_{n+m})}$$

$$= \frac{m}{2} \left[\frac{s_l}{((n+m)s_p + s_l)^2} \delta^2 + \frac{s_l}{(n+m)s_p + s_l} + \ln 2\pi s_l + \frac{S}{s_l} \right]. \quad (28)$$

When the sample number (n) is large enough (i.e., $n \rightarrow \infty$), the information gain converges to zero, and the uncertainty converges to m times the entropy of the stimulus source population.

$$\lim_{n \rightarrow \infty} G_{n+m} = 0, \quad (29)$$

$$\lim_{n \rightarrow \infty} U_{n+m} = \frac{m}{2} (\ln 2\pi s_l + 1) = mH(x), \quad (30)$$

where the entropy of the sensory data is:

$$\begin{aligned} H(x) &= \int_{-\infty}^{\infty} -p(x) \ln p(x) dx \\ &= \int_{-\infty}^{\infty} -N(\mu, \sigma^2) \ln N(\mu, \sigma^2) dx \\ &= \frac{1}{2} \left(\ln 2\pi \sigma^2 + \frac{S}{\sigma^2} \right) \cong \frac{1}{2} (\ln 2\pi s_l + 1) \\ &\because \sigma^2 \cong s_l, \lim_{n \rightarrow \infty} \frac{S}{\sigma^2} = 1. \end{aligned} \quad (31)$$

Entropy is proportional to the logarithm of sensory variance, $H(x) \propto \ln s_l$, when a Gaussian distribution is assumed. As discussed, the entropy of the sensory-data distributions represents the perceived complexity of the stimulus.

The free energy, as the summation of the information gain and the uncertainty, minimizes to entropy. In other words, $\lim_{n \rightarrow \infty} F_{n+m} = \lim_{n \rightarrow \infty} (G_{n+m} + U_{n+m}) = mH(x)$. By obtaining the sensory data, the information gain decreases to zero because the prior comes close to the true distribution of the stimuli source *via* Bayesian updates. The uncertainty decreases to the lower limit (i.e., entropy of the true distribution) because the predictive distribution comes close to the sample distributions, and the KL divergence becomes zero. With Gaussian posteriors and likelihoods, the uncertainty monotonically decreases as n increases. Therefore, when both information gain and the KL divergence between the predictive and data distribution converges to zero, free energy is minimized to the entropy of the data distribution.

Effect of Predictability

Because the uncertainty is a quadratic function of prediction error, as shown in Eq. 25, deviations from prior expectations increase the uncertainty. Its gradient A_U is always positive and means the sensitivity of the effect of prediction error on the uncertainty. The gradient of A_U with respect to the prior variation s_p is always negative:

$$\frac{\partial A_U}{\partial s_p} = \frac{-n^2 s_l}{2(ns_p + s_l)^3} < 0. \quad (32)$$

s_p is proportional to the logarithm of prior entropy (e.g., prior uncertainty; see Eq. 6). Thus, a high prior uncertainty indicates that the prediction error's effect on the uncertainty is low. In other words, a certain expectation (low prior variance) refers to a significant effect of prediction error on the uncertainty. By increasing the amount of data sampled, the gradient converges to zero, $\lim_{n \rightarrow \infty} A_U = 0$.

The gradient of B_U (the intercept) with respect to the prior variation s_p is always positive or zero:

$$\frac{\partial B_U}{\partial s_p} = \frac{ns_l}{2(ns_p + s_l)^2} \geq 0. \quad (33)$$

Therefore, small or zero prediction errors and higher prior variations (i.e., prior uncertainty) lead to higher uncertainty. Conversely, large prediction errors and higher prior variation means smaller uncertainty, owing to greater gradients. By increasing the number of data samples, the intercept converges to the entropy times the sample number (i.e., infinite).

$$\lim_{n \rightarrow \infty} B_U = \frac{1}{2} (\ln 2\pi s_l + 1) \lim_{n \rightarrow \infty} n = H(x) \lim_{n \rightarrow \infty} n. \quad (34)$$

Hence, the intercept is dominated by the uncertainty when the sample is large.

Regarding the free energy function, Eq. 26, the gradient of A_F with respect to the prior variation (s_p) is always negative:

$$\frac{\partial A_F}{\partial s_p} = -\frac{n^2}{2(ns_p + s_l)^2} < 0. \quad (35)$$

The gradient of B_F with respect to the prior variation (s_p) is always positive:

$$\frac{\partial B_F}{\partial s_p} = \frac{n}{2(ns_p + s_l)} > 0. \quad (36)$$

Therefore, small or zero prediction errors and higher prior variations mean smaller free energy. Conversely, large prediction errors and higher prior variation means higher free energy. This inversion is observed in the effect of predictability on the information gain when $s_p > s_l$ (Yanagisawa et al., 2019). **Figure 2** shows an example of the inversion of information gain, uncertainty, and free energy as functions of prediction error varied by prior uncertainty.

Effect of Sensory Variance

The entropy of sensory data is proportional to the logarithm of sensory variance (s_l) as the estimate of sensory data population variation. With large data samples, s_l comes close to the sensory data source variation. The entropies of sensory data sources represent perceived complexity. In this section, I analyze how the sensory variance affects the uncertainty, the information gain, and the free energy.

Regarding the uncertainty function, Eq. 25, the gradient of B_U with respect to sensory variance is always positive:

$$\frac{\partial B_U}{\partial s_l} = \frac{n s_p^2 + s_l^2 + s_l s_p}{2 (ns_p + s_l)^2} \geq 0. \quad (37)$$

Therefore, zero prediction errors and higher data variances mean higher uncertainty. The gradient of A_U with respect to sensory variance depends on the magnitude relation of ns_p and s_l .

$$\frac{\partial A_U}{\partial s_l} = \frac{n}{2} \frac{ns_p - s_l}{(ns_p + s_l)^3}. \quad (38)$$

In most cases, the prior variance is greater than the likelihood variance, $ns_p > s_l$, where n is a large natural number; therefore, partial derivative is greater than zero (i.e., the gradient of A_U is always positive). In this case, sensory variance increases A_U . It also increases the uncertainty when the prediction error is greater than zero. Therefore, in most cases, $ns_p > s_l$, and sensory variance and the uncertainty increase (**Figure 3A**).

Regarding the information gain function, Eq. 24, the gradient of both A_G and B_G with respect to the sensory variation is always negative:

$$\frac{\partial A_G}{\partial s_l} = -\frac{n^2 s_p (ns_p + s_l)}{(ns_p + s_l)^4} \leq 0, \quad (39)$$

$$\frac{\partial B_G}{\partial s_l} = \frac{ns_p}{2} \left\{ \frac{1 - (ns_p + s_l)}{(ns_p + s_l)^2} \right\}. \quad (40)$$

Therefore, when $(ns_p + s_l) > 1$, sensory variance decreases the information gain (**Figure 3B**). This is likely when n is a large natural number.

Regarding the free energy function, Eq. 26, the gradient of A_F with respect to the sensory variation (s_l) is always negative:

$$\frac{\partial A_F}{\partial s_l} = \frac{-n}{2(ns_p + s_l)} < 0. \quad (41)$$

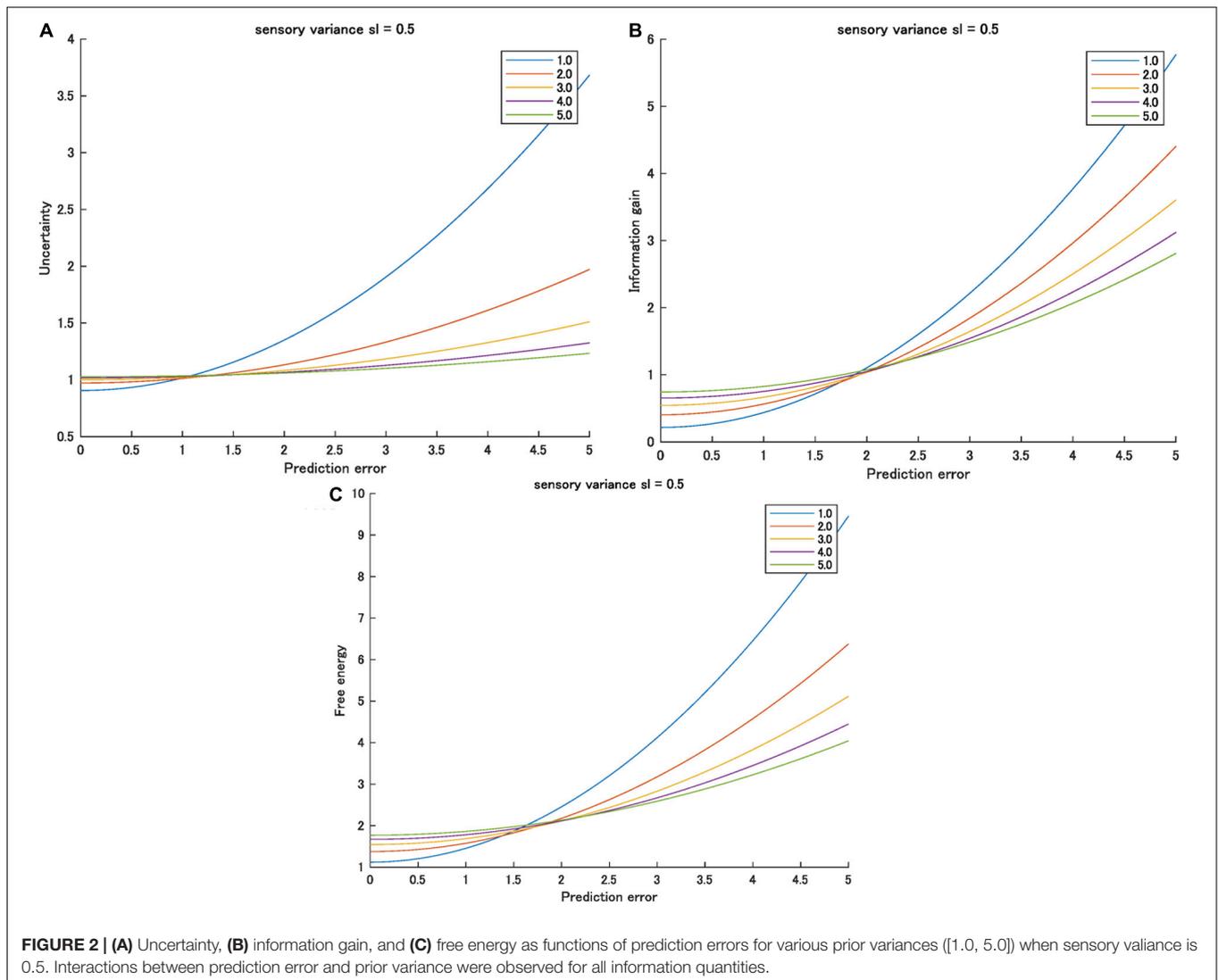
Therefore, sensory variance decreases the sensitivity of prediction errors on free energy. The gradient of B_F with respect to sensory variation (s_l) is always positive:

$$\frac{\partial B_F}{\partial s_l} = \frac{1}{2} \left(\frac{1}{ns_p + s_l} + \frac{n-1}{s_l} \right) > 0. \quad (42)$$

Therefore, when the prediction error is zero, sensory variance increases the free energy. However, the gradient is higher when the sensory variance is smaller. As a result, when the prediction error is large, this relationship is reversed. The higher the sensory variance, the smaller the free energy (**Figure 3C**). **Table 2** summarizes the effect of sensory variance on the effects of prediction errors on the three information quantities.

DISCUSSION

This study presents the relationship between the information content of perceived sensory stimuli (i.e., the free energy) and arousal potential (i.e., the potential of the primary emotional dimension). Information content to be processed after the perception of sensory stimuli (i.e., sensory surprisal) corresponds to the formulation of free energy commonly found in various disciplines, such as physics, statistics, and neuroscience (Friston et al., 2006). Free energy can be represented as a summation of the two terms of information content: information gain (Bayesian surprise) and surprisal (negative-log-likelihood averaged over the Bayesian posterior) (Eq. 10). Our previous study showed that the former represents novelty as a source of arousal potential (Yanagisawa et al., 2019; Ueda et al., 2021). I confirmed that the term also means expectation incongruity and unfamiliarity. The latter represents uncertainty of the estimated cause of sensory data. I demonstrated that the uncertainty decreases with the recognition and prior updating process based on Bayesian updating by adding sensory data from a stimuli source (i.e., a target of perception or cause of sensory data) in Section “Uncertainty Reduction in Bayesian Updating.” The uncertainty after the recognition and prior update (i.e., learning) is decomposed into KL divergence, representing deviation of sensory data from model prediction and entropy (Eq. 19). By reducing the model deviation to zero through iterations of Bayesian updating with enough data, the uncertainty minimizes to the entropy of sensory-data distribution as shown in Section “Uncertainty Convergence to Perceived Complexity.” From the analysis using the Gaussian generative model, I confirmed that the uncertainty converges to the entropy when sufficient data is provided in Section “Convergence to Entropy and Perceived Complexity.” Previous studies showed that the entropy



represents the perceived complexity of a stimuli (Ujii et al., 2012; Okano et al., 2020). Therefore, the uncertainty minimization by sufficiently observing a target of perception is considered to represent the perceived complexity.

All these factors (i.e., novelty, incongruity, unfamiliarity, uncertainty, and complexity) are collative variables exemplified by Berlyne as sources of arousal potential. Our previous model (Yanagisawa et al., 2019) considered only novelty. Hence, this model is more general than our previous model, including the second term regarding perceived complexity and uncertainty. Indeed, previous empirical studies have shown that the hedonic function of both perceived complexity and perceived novelty have inverse-U shapes. Mathematical formulations using free energy suggest that the sum of information content induced by novelty and complexity functions represents the arousal potential.

From the analysis of a Gaussian generative model, I found that all three information quantities (i.e., free energy, uncertainty, and information gain) are formulated as quadratic functions of prediction error, and the coefficients (i.e., gradient and

intercept) are functions of prior variance representing prior uncertainty and sensory variance, which is the variance of the sensory-data distribution proportional to the entropy. Prediction errors increase all three information quantities. Prior variance decreases the effect that prediction errors have on these information quantities. In other words, the inverse prior variance functions as the sensitivity of prediction errors on the arousal potentials. However, the intercept increases as prior variance increases. Therefore, small or zero prediction errors mean lower prior variance, which in turn means lower information quantities.

This causes an interaction effect between the prior uncertainty and prediction error on the information contents. In our previous study (Yanagisawa et al., 2019), we experimentally verified the interaction effect on information gain, representing surprise to novelty using event-related potential, P300, and subjective reports of surprise from human participants. Free energy, representing arousal potential, (e.g., novelty, uncertainty, and complexity) involves the same interaction effect. Prior

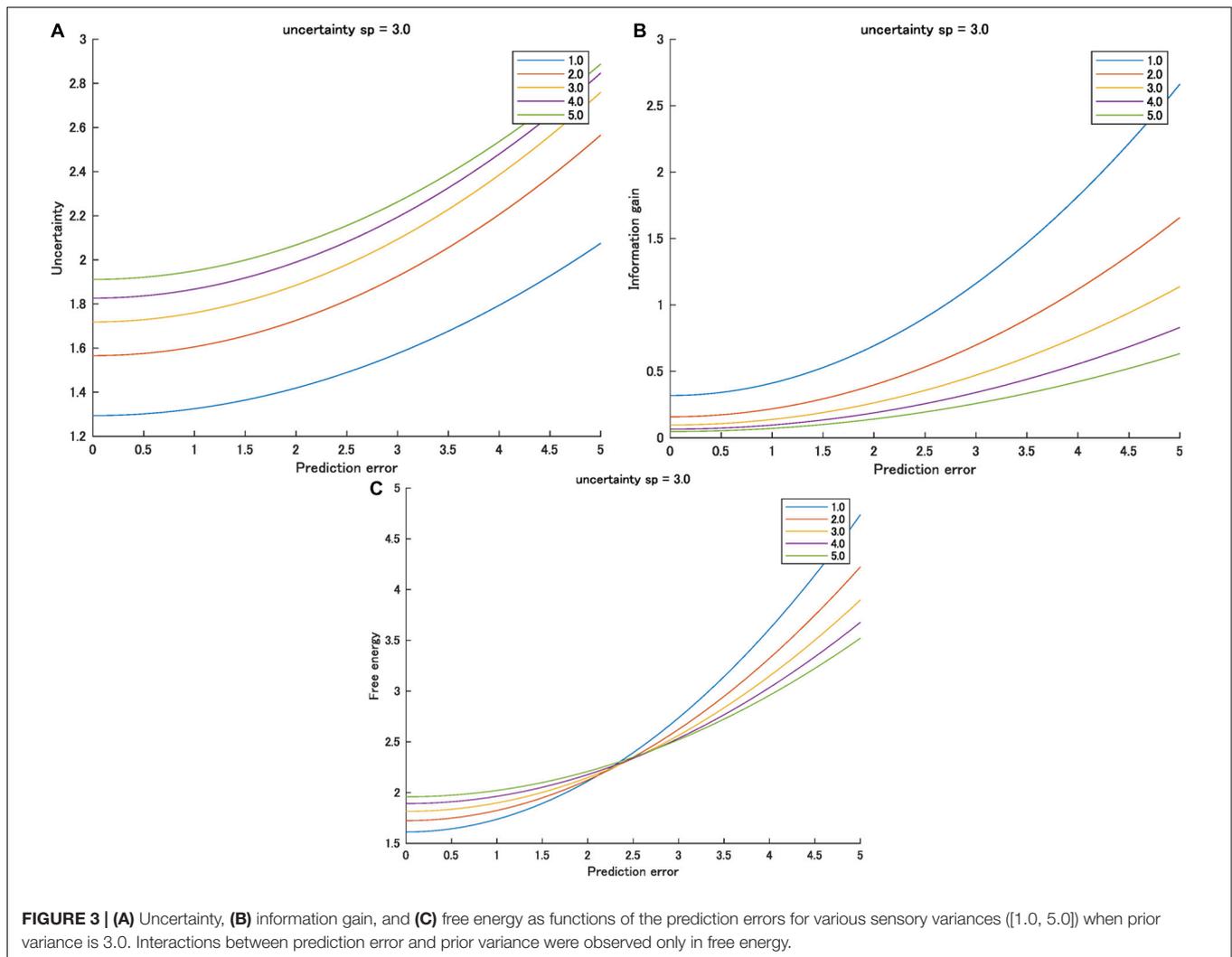


TABLE 2 | Effect of sensory variance on *uncertainty*, *information gain*, and free energy and its condition.

Uncertainty		Information gain		Free energy	
Gradient	Intercept	Gradient	Intercept	Gradient	Intercept
increase	Increase ($ns_p + s_i$) ≥ 1	decrease	Decrease ($s_p > s_i$)	decrease	increase

The gradient represents the sensitivity of prediction errors. The intercept is the effect when the prediction error is zero. Interaction of prediction error and sensory variance on free energy is predicted regardless any conditions.

knowledge and experience decrease prior variance (i.e., prior uncertainty). Therefore, the more solidified the prior beliefs that one has from prior knowledge and experience, the less arousal potential that one elicits from events having small prediction errors. By contrast, the more certain prior belief one has, the greater arousal potential one elicits from an event with large prediction error. This characteristic explains the fact that prior knowledge and experience increase the acceptance of complexity (Tuorila et al., 1994; Silvia, 2005). Small prediction errors with certain prior uncertainties mean that one is familiar with the object (e.g., one is frequently exposed to the object), or one has sufficient knowledge about the object (i.e., one is

an expert on the object). In such conditions, one’s information gain is small, and one has the capability to accept or enjoy more complex stimuli based on the Wundt curve shown in **Figure 1**.

I analyzed the effects of sensory variance on all three information quantities and summarized the results in **Table 2**. Sensory variance increases the uncertainty (intercept) and the sensitivity of prediction errors on the uncertainty (gradient). By contrast, sensory variance decreases information gain and the sensitivity of prediction errors on the information gain. Regarding the free energy (i.e., a summation of the uncertainty and the information gain), the sensory variance decreases the

sensitivity of prediction errors (gradient), but it increases the intercept. When the prediction error is zero or small, sensory variance increases free energy (arousal potential). Conversely, when the prediction error is high, sensory variance decreases the free energy (arousal potential). This interaction effect is the same as the effect of the prior variance on the free energy (see **Figures 2C, 3C**). In conclusion, both the prior and sensory data are considered information sources or cues for estimating the cause of sensory stimuli in Bayesian perception (Knill and Pouget, 2004; Körding and Wolpert, 2004). Inverse variances represent the precision of each source. In this sense, it is reasonable that the effect of variance (inverse precision) of both sources on arousal potential is the same.

The model predictions shown in **Figures 2, 3** provide a general scale of arousal (formalized as the free energy or surprisal) comprising novelty (the information gain) and perceived complexity (the uncertainty). This scale explains the difference in acceptable novelty and perceived complexity of stimuli based on predictability (i.e., prior uncertainty and prediction errors) caused by knowledge and experience (e.g., experts vs. novices) For a detailed analysis of the difference in acceptable novelty between experts and novices, see Miyamoto and Yanagisawa (2021). An experimental verification of this scale should be performed in future studies. This scale can help adjust to the preferable condition of novelty and complexity when designing artifacts (e.g., products and services) accepted by a target population (e.g., customers and users).

The human brain is an organ that processes significant amounts of information. This information involves a high cognitive load; therefore, a large amount of energy is required for processing. According to Friston's free-energy minimization (Friston et al., 2006), the brain perceives the causes of sensory data such that the free energy is variationally minimized. Free-energy minimization equilibrium is a law that can be applied to both physical and biological systems. Surprisal corresponds to the minimized free energy and is concerned with the information content remaining after the perception of sensory stimuli (or recognition) is performed where the recognition distributions can be approximated to the Bayesian posterior. My expectation is that the minimized or remaining free energy can be used to activate emotional arousal and subsequent emotions, such as valence, and for free energy to be used as the general principle of emotion potential. This prediction supports the theory of a relationship between valence and free energy variation (Joffily and Coricelli, 2013).

Emotions motivate certain actions, such as approach and avoidance. Active inferencing suggests that the free energy can

be reduced by acting to gain sensory evidence (Friston et al., 2015). This implies that emotions are functions that initiate actions for reducing the free energy, and the remaining free energy activates the said functions. In a model setting where an agent's action changes hidden states, and the agent's preference of sensory outcomes can be predefined as a prior, the confidence (or precision) of an agent's action model or the notion of affective charge is adopted to estimate emotional valence as a state of metacognitive level of a deep hierarchical generative model (Hesp et al., 2021).

By contrast, the present study considered valence as a function of arousal potential characterized by collative properties (i.e., novelty and complexity) of sensory stimuli given a static cause (a hidden state), which does not change over time. See the analysis of Sections "Uncertainty Convergence to Perceived Complexity" and "Analysis of Free Energy Using a Gaussian Generative Model." Preference to appropriate levels of arousal may be caused by curiosity and epistemic value in an active agent because these components comprise expected free energy (Friston et al., 2017), which is assumed to be arousal potential in the present study. A future direction to extend the arousal model should consider an active agent model that elucidates a mechanism of valence caused by arousal in a dynamic environment.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://arxiv.org/abs/2003.10073>.

AUTHOR CONTRIBUTIONS

HY designed and conducted the study and drafted and revised the manuscript.

FUNDING

This study was supported by JSPS KAKEN grant numbers 18H03318 and 21H03528.

ACKNOWLEDGMENTS

I thank Professor Tamotsu Murakami, Dr. Kazutaka Ueda, Dr. Takeo Kato, and members of Design Engineering Lab. at University of Tokyo for supporting this project.

REFERENCES

- Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Percept. Psychophys.* 8, 279–286.
- Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: a mathematical review. *J. Math. Psychol.* 81, 55–79. doi: 10.1016/j.jmp.2017.09.004
- Chen, X., Kar, S., and Ralescu, D. A. (2012). Cross-entropy measure of uncertain variables. *Inf. Sci.* 201, 53–60. doi: 10.3390/e22010108
- Clark, J. E., Watson, S., and Friston, K. J. (2018). What is mood? A computational perspective. *Psychol. Med.* 48, 2277–2284.
- Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215
- Fox, C. W., and Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artif. Intell. Rev.* 38, 85–95. doi: 10.3934/mbe.2020207
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11:127. doi: 10.1038/nrn2787

- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234.
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214.
- Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Comput.* 29, 2633–2683. doi: 10.1162/neco_a_00999
- Giacalone, D., Duerlund, M., Bøegh-Petersen, J., Bredie, W. L. P., and Frøst, M. B. (2014). Stimulus collative properties and consumers' flavor preferences. *Appetite* 77, 20–30. doi: 10.1016/j.appet.2014.02.007
- Hekkert, P., Snelders, D., and Van Wieringen, P. C. W. (2003). 'Most advanced, yet acceptable': typicality and novelty as joint predictors of aesthetic preference in industrial design. *Br. J. Psychol.* 94, 111–124. doi: 10.1348/000712603762842147
- Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., and Ramstead, M. J. D. (2021). Deeply felt affect: the emergence of valence in deep active inference. *Neural Comput.* 33, 398–446. doi: 10.1162/neco_a_01341
- Hung, W.-K., and Chen, L.-L. (2012). Effects of novelty and its dimensions on aesthetic preference in product design. *Int. J. Des.* 6, 81–90.
- Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vis. Res.* 49, 1295–1306.
- Joffily, M., and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Comput. Biol.* 9:e1003094. doi: 10.1371/journal.pcbi.1003094
- Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719.
- Körding, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.
- Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *Am. Psychol.* 50:372. doi: 10.1037/0003-066x.50.5.372
- Marin, M. M., Lampatz, A., Wandl, M., and Leder, H. (2016). Berlyne revisited: evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Front. Hum. Neurosci.* 10:536. doi: 10.3389/fnhum.2016.00536
- Miyamoto, M., and Yanagisawa, H. (2021). Modeling acceptable novelty based on Bayesian information—experimental verification of acceptance of deviation from musical rules. *Int. J. Affect. Eng.* 20, 265–274. doi: 10.5057/ijae.IJAE-D-21-00001
- Okano, A., Matsumoto, T., and Kato, T. (2020). Gaussian curvature entropy for curved surface shape generation. *Entropy* 22:353. doi: 10.3390/e22030353
- Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of dynamic causal models. *PLoS Comput. Biol.* 6:e1000709. doi: 10.1371/journal.pcbi.1000709
- Russell, J. A. (1980). A circumplex model of affect. *J. Personal. Soc. Psychol.* 39:1161.
- Sekoguchi, T., Sakai, Y., and Yanagisawa, H. (2019). "Mathematical model of emotional habituation to novelty: modeling with Bayesian update and information theory," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (Bari: IEEE). doi: 10.1109/SMC.2019.8914626
- Seth, A. K., and Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20160007.
- Shannon, C. E., Weaver, W., Blahut, R. E., and Hajek, B. (1949). *The Mathematical Theory of Communication*. Champaign, IL: University of Illinois Press Urbana.
- Silvia, P. J. (2005). Emotional responses to art: from collation and arousal to cognition and emotion. *Rev. Gen. Psychol.* 9:342. doi: 10.1037/1089-2680.9.4.342
- Tuorila, H., Meiselman, H. L., Bell, R., Cardello, A. V., and Johnson, W. (1994). Role of sensory and cognitive information in the enhancement of certainty and linking for novel and familiar foods. *Appetite* 23, 231–246. doi: 10.1006/appe.1994.1056
- Ueda, K., Sekoguchi, T., and Yanagisawa, H. (2021). How predictability affects habituation to novelty. *PLoS One* 16:e0237278. doi: 10.1371/journal.pone.0237278
- Ujii, Y., Kato, T., Sato, K., and Matsuoka, Y. (2012). Curvature entropy for curved profile generation. *Entropy* 14, 533–558.
- Valone, T. J. (2006). Are animals capable of Bayesian updating? An empirical review. *Oikos* 112, 252–259.
- Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., and Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Comput. Biol.* 11:e1004066. doi: 10.1371/journal.pcbi.1004066
- Wilson-Mendenhall, C. D., Barrett, L. F., and Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychol. Sci.* 24, 947–956.
- Yanagisawa, H. (2016). A computational model of perceptual expectation effect based on neural coding principles. *J. Sens. Stud.* 31, 430–439. doi: 10.1016/j.neunet.2020.06.008
- Yanagisawa, H., Kawamata, O., and Ueda, K. (2019). Modeling emotions associated with novelty at variable uncertainty levels: a Bayesian approach. *Front. Comput. Neurosci.* 13:2. doi: 10.3389/fncom.2019.00002

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yanagisawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Derivation of uncertainty using the Gaussian generative model:

$$\begin{aligned}
 U &= \langle -\ln p(x|\mu) \rangle_{N(\eta_{post}, s_{post})} \\
 &= \int_{-\infty}^{\infty} -\ln \left[\left(\frac{1}{\sqrt{2\pi s_l}} \right)^n \exp \left[-\frac{1}{2s_l} \{n(\mu - \bar{x})^2 + nS\} \right] \right] N(\eta_{post}, s_{post}) d\mu \\
 &= \frac{n}{2} \left[\frac{1}{s_l} \int_{-\infty}^{\infty} n(\mu - \bar{x})^2 N(\eta_{post}, s_{post}) d\mu + \ln 2\pi s_l + \frac{S}{s_l} \right]
 \end{aligned}$$

With $(\mu - \bar{x})^2 = (\mu - \eta_{post})^2 + 2(\eta_{post} - \bar{x})\mu - \eta_{post}^2 + \bar{x}^2$, the integral is derived as

$$\begin{aligned}
 &\int_{-\infty}^{\infty} (\mu - \bar{x})^2 N(\eta_{post}, s_{post}) d\mu \\
 &= \int_{-\infty}^{\infty} (\mu - \eta_{post})^2 N(\eta_{post}, s_{post}) d\mu + 2(\eta_{post} - \bar{x}) \int_{-\infty}^{\infty} \mu N(\eta_{post}, s_{post}) d\mu + (\bar{x}^2 - \eta_{post}^2) \int_{-\infty}^{\infty} N(\eta_{post}, s_{post}) d\mu \\
 &= s_{post} + 2(\eta_{post} - \bar{x})\eta_{post} + \bar{x}^2 - \eta_{post}^2 \\
 &= \frac{s_p s_l}{n s_p + s_l} + \frac{s_l^2 (\eta - \bar{x})^2}{(n s_p + s_l)^2} \\
 &= \left\{ \frac{s_p}{n s_p + s_l} + \frac{s_l}{(n s_p + s_l)^2} \delta^2 \right\} s_l,
 \end{aligned}$$

where prediction error $\delta = \eta - \bar{x}$. Therefore,

$$U = \frac{n}{2} \left[\frac{s_l}{(n s_p + s_l)^2} \delta^2 + \frac{s_p}{n s_p + s_l} + \ln 2\pi s_l + \frac{S}{s_l} \right].$$