



# Depth and the Uncertainty of Statistical Knowledge on Musical Creativity Fluctuate Over a Composer's Lifetime

Tatsuya Daikoku\*

Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

## OPEN ACCESS

### Edited by:

Matjaž Perc,  
University of Maribor, Slovenia

### Reviewed by:

Haroldo Valentin Ribeiro,  
Universidade Estadual de Maringá,  
Brazil

Rodrigo Laje,  
Universidad Nacional de Quilmes  
(UNQ), Argentina

### \*Correspondence:

Tatsuya Daikoku  
daikoku@cbs.mpg.de

**Received:** 06 December 2018

**Accepted:** 11 April 2019

**Published:** 30 April 2019

### Citation:

Daikoku T (2019) Depth and the Uncertainty of Statistical Knowledge on Musical Creativity Fluctuate Over a Composer's Lifetime.  
*Front. Comput. Neurosci.* 13:27.  
doi: 10.3389/fncom.2019.00027

Brain models music as a hierarchy of dynamical systems that encode probability distributions and complexity (i.e., entropy and uncertainty). Through musical experience over lifetime, a human is intrinsically motivated in optimizing the internalized probabilistic model for efficient information processing and the uncertainty resolution, which has been regarded as rewards. Human's behavior, however, appears to be not necessarily directing to efficiency but sometimes act inefficiently in order to explore a maximum rewards of uncertainty resolution. Previous studies suggest that the drive for novelty seeking behavior (high uncertain phenomenon) reflects human's curiosity, and that the curiosity rewards encourage humans to create and learn new regularities. That is to say, although brain generally minimizes uncertainty of music structure, we sometimes derive pleasure from music with uncertain structure due to curiosity for novelty seeking behavior by which we anticipate the resolution of uncertainty. Few studies, however, investigated how curiosity for uncertain and novelty seeking behavior modulates musical creativity. The present study investigated how the probabilistic model and the uncertainty in music fluctuate over a composer's lifetime (all of the 32 piano sonatas by Ludwig van Beethoven). In the late periods of the composer's lifetime, the transitional probabilities (TPs) of sequential patterns that ubiquitously appear in all of his music (familiar phrase) were decreased, whereas the uncertainties of the whole structure were increased. Furthermore, these findings were prominent in higher-, rather than lower-, order models of TP distribution. This may suggest that the higher-order probabilistic model is susceptible to experience and psychological phenomena over the composer's lifetime. The present study first suggested the fluctuation of uncertainty of musical structure over a composer's lifetime. It is suggested that human's curiosity for uncertain and novelty seeking behavior may modulate optimization and creativity in human's brain.

**Keywords:** statistical learning, entropy, mutual information, information theory, Markovian, Bayesian, order, n-gram

## INTRODUCTION

### Statistical Learning and Uncertainty in the Brain

The brain models external phenomena as a hierarchy of dynamical systems that encode probability distributions and complexity (i.e., entropy and uncertainty) over states in the world. Based on the internalized hidden model, it can predict a future state and optimize behavior and action to resolve the uncertainty (Friston, 2010). Within predictive-coding framework, this behavior mandates the suppression of prediction errors (prediction of content) and uncertainty (prediction of the context or precision of predictability and uncertainty) through updating internal model that generates predictions and the belief (Kanai et al., 2015). It has been considered that aesthetic appreciation of music can be modulated by these brain function: Through musical experience over lifetime, a human is intrinsically motivated in optimizing the internalized probabilistic model for efficient information processing and the uncertainty resolution, which has been regarded as rewards. For example, previous studies demonstrated that the precise prediction (Przyssinda et al., 2017) and uncertainty perception (Hansen and Pearce, 2014) in music is stronger in proficient musicians than non-musicians.

This generative model could cover statistical learning (SL) theory of brain (Saffran et al., 1996; Cleeremans et al., 1998; Perruchet and Pacton, 2006). The SL is an implicit process by which the brain automatically calculates the statistical distribution of sequential phenomena based on Bayesian inference (Daikoku et al., 2012, 2014, 2016, 2017a, 2018; Yumoto and Daikoku, 2016; Daikoku and Yumoto, 2017), grasps the uncertainty (Hasson, 2017), predicts a future state based on the internal statistical model, and optimize action for achieving a given goal (Monroy et al., 2017a,b). By SL, generation of culture (Feher et al., 2016), individuality of creativity (Daikoku, 2018b) can be originated. Although brain tries to realize valuable behaviors at the lowest possible informational cost and uncertainty, it also seeks slightly suboptimal solution if the solution can be afforded at a significantly low uncertainty (Tishby and Polani, 2011). In other words, human's behavior appears to be not necessarily directing to efficiency but sometimes act unefficiently to explore a maximum rewards of uncertainty resolution. Previous studies suggest that the drive for novelty seeking behavior (high uncertain phenomenon) reflects human's curiosity and that the curiosity rewards encourage humans to create and learn new regularities (Kagan, 1972; Wittmann et al., 2008; Krebs et al., 2009; Schwartenbeck et al., 2013). Furthermore, a certain degree of uncertainty generates excitement and pleasure (Shen et al., 2015) because we explore a maximum curiosity rewards. Although brain generally minimizes prediction errors and uncertainty (Friston, 2010), we sometimes derive pleasure from prediction errors under conditions such as enjoying music listening due to curiosity and motivation for novelty seeking behavior by which we anticipate the resolution of uncertainty. Some literatures propose the hypothesis that the recurrent resolution of uncertainty activates

reward networks that underwrite pleasure induced by listening to music (Koelsch, 2014; Salimpoor et al., 2015). It has been suggested that creativity can be explained as by-products of such intrinsic curiosity rewards (Schmidhuber, 2006). That is, human seems to look for some forms of optimality between uncertain and certain situations through action by which we are expected a maximum curiosity rewards, and hence our action gives rise to increasing as well as decreasing uncertainty. Recent studies imply that the curiosity rewards encourage humans to create and learn new regularities (Schmidhuber, 2006), and the fluctuations in uncertainty of predictions could contribute to aesthetic appreciation of art and music (Koelsch, 2014). Thus, it is hypothesized that human's intrinsic curiosity and motivation may modulate optimization and efficiency of prediction and action involved in SL. Recent computational studies on music suggest that, from early to late periods in the composer's lifetime, the transitional probabilities (TPs) of familiar phrase that ubiquitously appears in all of his music were gradually decreased (Daikoku, 2018d). This suggests that the statistical knowledge (Daikoku, 2018a) may be susceptible to long-term experience that modulates brain's probabilistic model (Hansen and Pearce, 2014). A neurophysiological study also suggested that sequences with higher entropy were learned based on higher-order TP, whereas those with lower entropy were learned based on lower-order TP (Daikoku et al., 2017b; Daikoku and Yumoto, 2019). Another study suggested that certain regions or networks perform specific computations of entropy (i.e., uncertainty), which are different from TP (i.e., prediction) of each content (Hasson, 2017). Thus, interaction between prediction and uncertainty in perceptive systems is an important topic to understand whole process of brain SL in both computational and neurophysiological areas (Daikoku, 2018c; Yumoto and Daikoku, 2018). Nevertheless, to our knowledge, few study examined relationships between SL, uncertainty and musical creativity and how curiosity for uncertain and novelty seeking behavior modulates musical creativity. The present study investigated how the probabilistic model and the uncertainty in music fluctuate over a composer's lifetime (all of the 32 piano sonatas by Ludwig van Beethoven).

### Computational Model

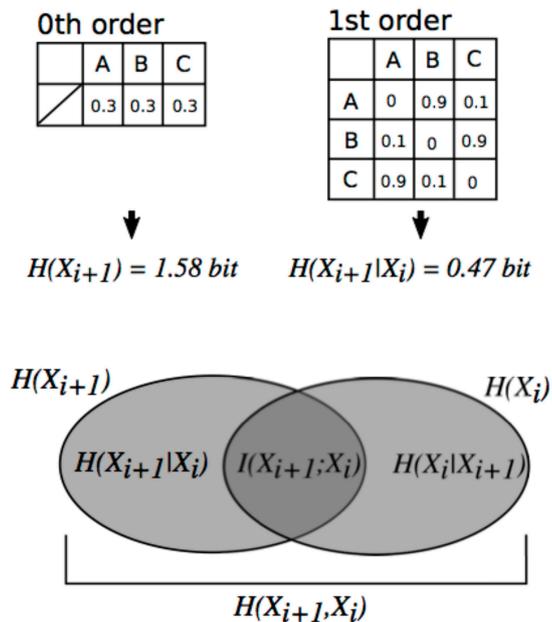
The computational model and simulation have been used to understand SL systems (e.g., Pearce and Wiggins, 2012; Rohrmeier and Rebuschat, 2012; Daikoku, 2018a; Wiggins, 2018; Daikoku and Yumoto, 2019). Although experimental approaches are necessary for understanding the real-world brain's function, the modeling approaches partially outperform experimental results under conditions that are impossible to replicate in an experimental approach (e.g., long-term statistical variation over the decades within a person and across cultures) and serves an important dual role in providing a quantitative account of observed empirical effects and in generating novel predictions to guide empirical research (e.g., Elman, 1990; Thiessen et al., 2013; Carreiras et al., 2014). Computational modeling can also express the relevant neural networks and neural hardware of sensory cortices (Turk-browne et al., 2009; Roux and Uhlhaas, 2014). For example, simple recurrent

network (SRN), which is classified as a neural network and was firstly devised by Elmer Elman (1990), learns sequential co-occurrence statistics by error-driven learning in which the gap between the prediction of a next input and the actual input drives changes to the weights on its internal connections. The SRN (Rogers and McClelland, 2004) and a modified SRN (Altmann, 1999; Dienes et al., 1999) implement a similarity space in which words referring to similar objects or actions were located more closely to one another than to words referring to dissimilar objects or actions. The neural network and deep learning such as Long-Short Term Memory (LSTM) (Hochreiter and Jürgen Schmidhuber, 1997), on the other hand, is not intended to be a model of the relationship between human episodic and semantic memory although they proceed in this direction. Corpus-based approaches such as hyperspace analog to language (HAL) (Lund and Burgess, 1996), bound encoding of the aggregate language environment (BEAGLE) (Jones and Mewhort, 2007), Latent Semantic Analysis (LSA) (Landauer and Dumais, 1997) are based on abstraction of episodic memory of input information and encoding in a multidimensional semantic space as semantic memory. Their models could also generate semantic similarity spaces in the similar way. For instance, when a verb of “drink” occurs, the models predict subsequent words that can be drunken. PARSER (Perruchet and Vinter, 1998), Competitive Chunker (Servan-Schreiber and Anderson, 1990), Information Dynamics of Music (IDyOM) (Pearce and Wiggins, 2012), Information Dynamics of Thinking (IDyOT) (Wiggins, 2018), and other *Markovian* models including the n-gram and nth-order Markov models (Daikoku, 2018b), can implement chunking hypotheses that learning is based on extracting, storing, and combining small chunks. Particularly, information-theoretical models including *Markovian* processes have been applied to neurophysiological studies of SL in human brain as well as computational simulation (Pearce et al., 2010a; Pearce and Wiggins, 2012; Daikoku et al., 2014, 2016, 2017a, 2018; Yumoto and Daikoku, 2016, 2018; Daikoku and Yumoto, 2017; Daikoku, 2018c). These neurophysiological experiments showed consistent evidence: neural activities for stimuli with high information content (i.e., low probability) are larger than those with low information content (i.e., high probability). Furthermore, these SL effects were larger when humans are exposed stimulus sequence with less information entropy (uncertainty), compared with when they are exposed stimulus sequence with high information entropy (Daikoku et al., 2017c). The mutual information of information theory, which has been assumed as the reduction of uncertainty afforded by observations (see section Mutual Information of nth-order SL model), is also correlated with neuronal activity in limbic cortex (Harrison et al., 2006). This neural phenomenon is in agreement with a *Bayesian* hypothesis in theoretical neurobiology that the brain encodes probabilities (beliefs) about the causes of sensory data, and that these beliefs are updated in response to new sensory evidence based on Bayesian inference (Kersten et al., 2004; Knill and Pouget, 2004; Doya et al., 2007; Friston, 2010; O’Reilly et al., 2012; Parr and Friston, 2018; Parr et al., 2018). Formulations of self-organization (Karl Friston, 2013; Kirchhoff et al., 2018) and brain connectivity

(Parr and Friston, 2018) are also expressed using an information-theoretical concept called Markov blankets (Pearl, 1988). The blanket of a state is the only knowledge necessary to predict the behavior of that state and the adjacent state. If we know everything within a blanket, knowledge about things outside the blanket becomes uninformative about things inside the blanket. For example, Parr and Friston (2018) hypothesized that a neuronal population reflecting a given variable only need receive connections from those populations representing its blanket and explained this notion from perception, planning, attention, and movement. The Markov blanket may also represent in part chunk formation although it’s not sufficient. Markov decision process (MDP) (Schwartenbeck et al., 2013; Karl Friston et al., 2014, 2015; Pezzulo et al., 2015), which has often been used for reinforcement learning in AI and robotics, extends the simple perceptive process by adding active process (controlling predictability by choice, called “*policy*”) and “*rewards*” (giving motivation). The IDyOM is also an extension of Markov model to precisely modeling SL of musical sequence combining several concomitant information such as pitch, duration, onset, scale degree, and so on. The SL based on IDyOM could also be reflected in neurophysiological responses within the predictive-coding framework (Pearce et al., 2010b). The IDyOT also takes advantage of information theory to represent domain-general SL mechanisms that cover both language and music (Wiggins, 2018). Particularly, this model implements semantic and episodic memory systems, and captures hierarchical SL process from lower- to higher-level using boundary entropy: spectrum of auditory sequence is chunked into phonemes, then morphemes, then words (Wiggins, in press). In summary, information-theoretical models including Markovian processes can capture a variety of neurophysiological phenomena on SL such as prediction, uncertainty, a part of chunk formation, and policy of action, across domains, and modality.

A previous study reported that SL effects based on TPs occurs action as well as perception (Monroy et al., 2017c). This suggests that SL also contributes production of sequences. In other words, from psychological perspective, TP distribution sampled from music based on Markov models may also refer to the characteristics of a composer’s statistical knowledge: a high-probability transition in music may be one that a composer is more likely to predict and choose based on the latest n states, compared to a low-probability transition. Thus, the Markov model is used in the interdisciplinary realms of neuroscience, behavioral science, engineering, and informatics. A computational study using nth-order Markov or n-gram models suggested that time-course variations of statistics in music reflect time-course variations of a composer’s statistical knowledge (Daikoku, 2018d). Neurophysiological studies also suggested that time-course variations of statistics of auditory sequence modulate SL effects (Daikoku et al., 2017c) and that the SL effects of sequences with higher entropy were lower than those with lower entropy, even when TP itself is same between these two sequences (Daikoku et al., 2017c). These studies suggest that time-course variations of TPs and entropy may partially be able to predict the SL model in human’s brain.

sequence: (C)ABCABCABABCABCACABCABCABCBCABC



**FIGURE 1** | Relationship between order of transitional probabilities, conditional entropy, and mutual information illustrated using a Venn diagram. The degree of dependence on  $X_i$  for  $X_{i+1}$  is measured by mutual information [mutual information  $I(X; Y) = \text{entropy}(H(X_{i+1})) - \text{conditional entropy}(H(X_{i+1}|X_i))$ ]. The mutual information of sequences in this figure is more than 0. Thus, each event  $X_{i+1}$  in the sequence is dependent on a preceding event  $X_i$ .

## Mathematical Interpretation of Brain's Statistical Learning Based on Information Theory

### Nth-Order Transitional Probability

According to SL theory, the brain automatically computes both lower- and higher-order TPs in sequences (Furl et al., 2011; Yumoto and Daikoku, 2016, 2018, grasps uncertainty/entropy in the whole sequences Hasson, 2017, and predicts a future state based on the internalized statistical model Friston, 2010). The TP is a conditional probability of an event B given that the latest event A has occurred, written as  $P(B|A)$ . The nth-order TP distributions sampled from sequential information such as music and language can be expressed by nth-order Markov models. The nth-order Markov model is based on the conditional probability of an event  $e_{n+1}$ , given the preceding  $n$  events based on Bayes' theorem ( $P(e_{n+1}|e_n)$ ). From psychological perspective, the conditional probability ( $P(e_{n+1}|e_n)$ ) can be interpreted as positing that the brain predicts a subsequent event  $e_{n+1}$  based on the preceding events  $e_n$  in a sequence. In other words, learners expect the event with the highest TP based on the latest  $n$  states, whereas they are likely to be surprised by an event with lower TP. Furthermore, TPs are often translated as information contents (IC) ( $-\log_2 I/P(e_{n+1}|e_n)$ ) of information theory (Pearce and Wiggins, 2006). The lower IC (i.e., higher TPs) means higher predictabilities and smaller surprising, whereas the higher IC (i.e., lower TPs) means lower predictabilities and larger surprising. In the end, a tone with lower IC may be

one that a composer is more likely to predict and choose as a next tone, compared to tones with higher IC. IC can be used in computational studies of music to discuss psychological phenomena involved in prediction and SL.

### Entropy and Uncertainty

Entropy as well as TP of each event is used to understand predictability of a sequence (Pearce, 2005). Entropy (e.g., see Figure 1) is calculated from probability distribution, interpreted as uncertainty (Friston, 2010), and used to evaluate neurophysiological effects of uncertainty in SL (Harrison et al., 2006) and curiosity (Loewenstein, 1994). A previous study reported that neural systems of uncertainty perception were partially independent of those of prediction of each content (Hasson, 2017). Some articles, however, suggest that uncertainty modulates predictability of each content in SL (Daikoku et al., 2017c). Furthermore, uncertainty of auditory and visual statistics is coded by modality-general, as well as modality-specific, neural systems (Strange et al., 2005; Nastase et al., 2014). This suggests that the neural basis that codes uncertainty as well as prediction, is a domain-general system. Thus, there seems to be neural and psychological interactions of perceptions between prediction and uncertainty.

### Mutual Information of nth-order SL Model

Mutual information (MI) and pointwise Mutual information (PMI) are a measure of the mutual dependence between the two variables. The PMI refers to each event in sequence (local

dependence), whereas MI refers to the average of all events in the sequence (global dependence). In the framework of SL based on TPs ( $P(e_{n+1}|e_n)$ ), MI explains how an event  $e_{n+1}$  is dependent on the preceding event  $e_n$ . Thus, MI is a key to understanding order of SL. For instance, conventional oddball sequence, which consists of a frequent stimulus with high probability of appearance and a deviant stimulus with low probability of appearance, has weak dependence between two adjacent events ( $e_n, e_{n+1}$ ) and shows low MI, because event  $e_{n+1}$  appears independently of preceding events  $e_n$ . In

contrast, SL sequence based on TPs, but not probabilities of appearance, has strong dependence between two adjacent events and shows larger MI. For example, typical SL paradigm that consists of concatenation of pseudo-words with three stimuli has large MI until 2nd-order Markov or tri-gram models [i.e.,  $P(C|AB)$ ], whereas it has low MI from 3rd-order Markov or four-gram models [i.e.,  $P(D|ABC)$ ]. Thus, MI is sometimes used to evaluate hierarchical SL in both neurophysiological and computational studies (Harrison et al., 2006; Pearce et al., 2010b).

#### Beethoven's Piano Sonata No.2 in A major, Op.2-2: 1st Movement

#### Beethoven's Piano Sonata No.3 in C major, Op.2-3: 3rd Movement

#### Beethoven's Piano Sonata No.7 in D major, Op.10-3: 4th Movement

#### Beethoven's Piano Sonata No.10 in G major, Op.14-2: 1st Movement

**FIGURE 2** | Representative sequences of  $[0, -2, -3, -5, -7]$ ,  $[0, -2, -4, -5, -7]$ , and  $[0, -1, -3, -5, -6]$  in Beethoven's piano sonatas in the early period.

In this section, the three types of information-theoretical evaluation of SL models (i.e., IC, entropy, and MI) were explained from psychological aspects. In sum, (1) IC reflects surprising/predictability. A tone with lower IC (i.e., higher TPs) may be one that a composer is more likely to predict and choose as a next tone, compared to tones with higher IC. (2) Entropy reflects uncertainty of whole sequences. (3) MI reflects hierarchy of statistics and is interpreted as dependence of preceding sequential events in SL. Using them, the present study investigated how prediction, uncertainty, and the depth of implicit knowledge in music vary over a composer's lifetime (all of the 32 piano sonatas by Ludwig van Beethoven).

## Ludwig Van Beethoven's Piano Sonata

The German composer and pianist Ludwig van Beethoven (1770–1827) remains one of the most famous and influential of all composers. It is believed that his music strongly expresses the psychological variations and visions of his life (Sullivan, 1927; Boucourechliev, 1963). Beethoven's compositional career is often divided into the early (around 1802), middle (around 1802–1814), and late periods (from about 1814) (Dahlhaus, 1991; Adorno-Wiesengrund, 1993). It is generally thought that his works in the early period were strongly influenced by his predecessors in classicism, such as Wolfgang Amadeus Mozart (1756–1792) and Franz Joseph Haydn (1732–1809), whereas his works in the late period show his personal character and experience (Sullivan, 1927) and accompanying intellectual depth and personal expression (Dahlhaus, 1991; Adorno-Wiesengrund, 1993). Thus, his psychological variations on thinking and experience may form the statistical characteristics of his music that may reflect a composer's statistical knowledge (Johnson et al., 1985). It is believed that he always explored new directions of musical composition and gradually expanded his scope of music over his lifetime (Dahlhaus, 1991; Adorno-Wiesengrund, 1993). Using Beethoven's piano sonatas over his lifetime, the present study examined time-course variations of three types of statistics in music: TPs (ICs) of sequential patterns that appear in all 32 sonatas, entropy of whole TP distribution, and the MI. It was hypothesized that, because of his exploration of new directions in musical composition over his lifetime, TP of phrases that frequently appear in the early period (i.e., sequences with high TP) might decrease in the late period (i.e., decreasing TP), whereas entropy might increase in the late period. It would be very interesting if the psychological variations in which Beethoven explored new directions and gradually expanded his scope of music over his lifetime were reflected in the SL models of his music.

## METHODS

The Piano Sonata with all of its movements by Ludwig van Beethoven (No.1 in F minor, Op.2-1 to No.32 in C minor, Op.111, composed 1795–1822) was used in the present study. Using a scorewriter (Finale version 25, MI Seven Japan, Inc.), electronic scoring data of the sequences of highest pitch were extracted from the XML files. The highest pitches were chosen

based on the following definitions: the highest pitches that can be played at a given point in time, the pitches with slurs can be counted as one, and the grace notes were excluded. Although melody is sometimes not highest pitches e.g., bass melodies), the present study only analyzed the highest pitch because different melodies could concurrently appear in some titles of music, and melody is often played in highest pitches. Using all the sequences of highest pitches in a movement of a Sonata, sequential patterns based on uni- to four-grams were extracted. For each type of the sequential patterns, all pitches were numbered so that the first pitch was 0 in each transition, and an increase or decrease in a semitone was 1 and  $-1$  based on the first pitch, respectively. The representative examples were shown in **Figure 2**. This revealed interval patterns but not pitch pattern. This procedure was employed to eliminate the effects of the change of key on transitional patterns. The interpretation of the change of key depends on musicians, and it is difficult to define in an objective manner. Thus, the results in the present study may represent a variation of statistics associated with relative pitch rather than absolute pitch. Then, the TPs of the sequential patterns were calculated based on 0th- to 3rd-order Markov chains. Furthermore, TPs of all the movements in each piece of sonata (No.1 to No.32) were weighted averaged: an average in which probability of each phrase is multiplied by a weight before summing to a single average value. That is weightings are the equivalent of having that many like items with the same value involved in the average. The  $n$ th-order Markov chain is the conditional probability of an event  $e_{n+1}$ , given the preceding event  $e_n$ :

$$P(e_{n+1}|e_n) = \frac{P(e_{n+1} \cap e_n)}{P(e_n)} \quad (1)$$

The ICs ( $I$ ) and conditional entropy ( $H$ ) in the  $n$ th-order TP distribution (hereafter, Markov entropy) were calculated using TPs in the framework of information theory:

$$I(e_{n+1}|e_n) = -\log_2 P(e_{n+1}|e_n) \text{ (bit)} \quad (2)$$

$$H(B|A) = -\sum_i \sum_j P(ai)P(bj|ai) \log_2 P(bj|ai) \text{ (bit)} \quad (3)$$

where  $P(bj|ai)$  is a conditional probability of sequence "ai bj." Then, MI ( $I(X;Y)$ ) were calculated in 1st-, 2nd-, and 3rd-order Markov models. MI is an information theoretic measure of dependency between two variables. From entropy values, the MI can also be expressed as

$$\begin{aligned} I(X; Y) &= \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \\ &= \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)} \right) - \sum_{x,y} p(x,y) \log p(y) \\ &= \sum_{x,y} p(x) p(y|x) \log p(y|x) - \sum_{x,y} \log p(y) p(x,y) \\ &= \sum_x p(x) \left( \sum_y p(y|x) \log p(y|x) \right) \end{aligned}$$

$$\begin{aligned}
& - \sum_y \log p(y) \left( \sum_x p(x, y) \right) \\
& = - \sum_x p(x) H(Y|X = x) - \sum_y p(y) \log p(y) \\
& = -H(Y|X) + H(Y) \\
& = H(Y) - H(Y|X) \text{ (bit)} \tag{4}
\end{aligned}$$

where  $p(x, y)$  is the joint probability function of  $X$  and  $Y$ ,  $p(x)$ , and  $p(y)$  are the marginal probability distribution functions of  $X$  and  $Y$  respectively,  $H(X)$  and  $H(Y)$  are the marginal entropies,  $H(X|Y)$  and  $H(Y|X)$  are the conditional entropies, and  $H(X, Y)$  is the joint entropy of  $X$  and  $Y$  (Figure 1) (Daikoku, 2018a). Based on psychological and information-theoretical concepts, the Equation (4) can be regarded that the amount of entropy (uncertainty) remaining about  $Y$  after  $X$  is known. That is, the MI is corresponding to reduction in entropy (uncertainty). In each order of Markov models, the sequential patterns that ubiquitously appear in all 32 sonatas (hereafter, familiar phrase) were extracted. Then, TPs of the familiar phrases were averaged (0th: 20 phrases, 1st: 37 phrases, 2nd: 12 phrases, and 3rd: 3 phrases) (Table 1). The 32 sonatas were divided based on the well-known 3 periods: early (No.1 to 12, No.19, and No.20), middle (No.13 to 18 and No. 21 to 27), and late (No. 28 to 32). Then, I conducted analysis of variances (ANOVAs) with a within-subject factor order (0th vs. 1st vs. 2nd vs. 3rd) and a between-subjects factor composition period of the sonatas (early vs. middle vs. late) for the TPs of familiar phrases and entropy of whole music, and an ANOVA with a within-subject factor order (1st vs. 2nd vs. 3rd) and a between-subjects factor composition period (early vs. middle vs. late) for the mutual information. When we detected significant effects, Bonferroni-corrected *post-hoc* tests were conducted for further analysis. Then, in each order of Markov models, the TPs of familiar phrase and the uncertainty of whole music were compared by Pearson's correlation analysis. Statistical significance levels were set at  $p = 0.05$  for all analyses.

## RESULTS

### TPs of Familiar Phrases

In the TPs of familiar phrases, An ANOVA with a within-subject factor order (0th vs. 1st vs. 2nd vs. 3rd) and a between-subjects factor composition period (early vs. middle vs. late) was conducted. As a result, the main effect of period was significant [ $F(2, 29) = 6.02$ ,  $p = 0.007$ ,  $\text{partial}\eta^2 = 0.29$ , early > late,  $p = 0.005$ ; middle > late,  $p = 0.032$ ] (Figure 3D). The period-order interaction was also significant [ $F(6) = 6.82$ ,  $p < 0.001$ ,  $\text{partial}\eta^2 = 0.32$ ] (Figure 3A). The 3rd-order TPs in late period were significantly lower than those in early ( $p < 0.001$ ) and middle periods ( $p = 0.003$ ). That is to say, the 3rd-order TPs of familiar phrases in the late period only decrease during lifetime. The main effect of order was significant [ $F(3, 87) = 1108.35$ ,  $p < 0.001$ ,  $\text{partial}\eta^2 = 0.98$ ]. The 0th-order TPs were significantly lower than the 1st-, 2nd-, and 3rd-order TPs (all:  $p < 0.001$ ). The 1st-order TPs were significantly lower than the 2nd-, 3rd-order TPs (all:  $p < 0.001$ ). The 2nd-order TPs were significantly lower than the 3rd-order TPs ( $p = 0.007$ ).

**TABLE 1** | Sequential patterns that appear in all 32 sonatas (i.e., phrases) in each order of Markov models.

Order	Sequential patterns
0th	[-2], [-1], [1], [0], [2], [-3], [3], [5], [-4], [4], [-5], [12], [-7], [7], [9], [-12], [8], [-6], [6], [-9]
1st	[1, 1], [-1, 2], [-1, -1], [1, -2], [0, 1], [3, 3], [0, 2], [1, -4], [0, 3], [0, 5], [1, -1], [-2, 2], [-2, 5]
2nd	[-2, -4, -5], [0, 0, 0], [-1, -3, -5], [2, 4, 5], [-2, -3, -5], [1, 3, 5], [2, 3, 5], [2, 0, -1], [-2, -3, -2], [-1, 0, 2], [1, 3, 1], [-1, 0, -1]
3rd	[-2, -3, -5, -7], [-2, -4, -5, -7], [-1, -3, -5, -6]

### Entropy and Uncertainty

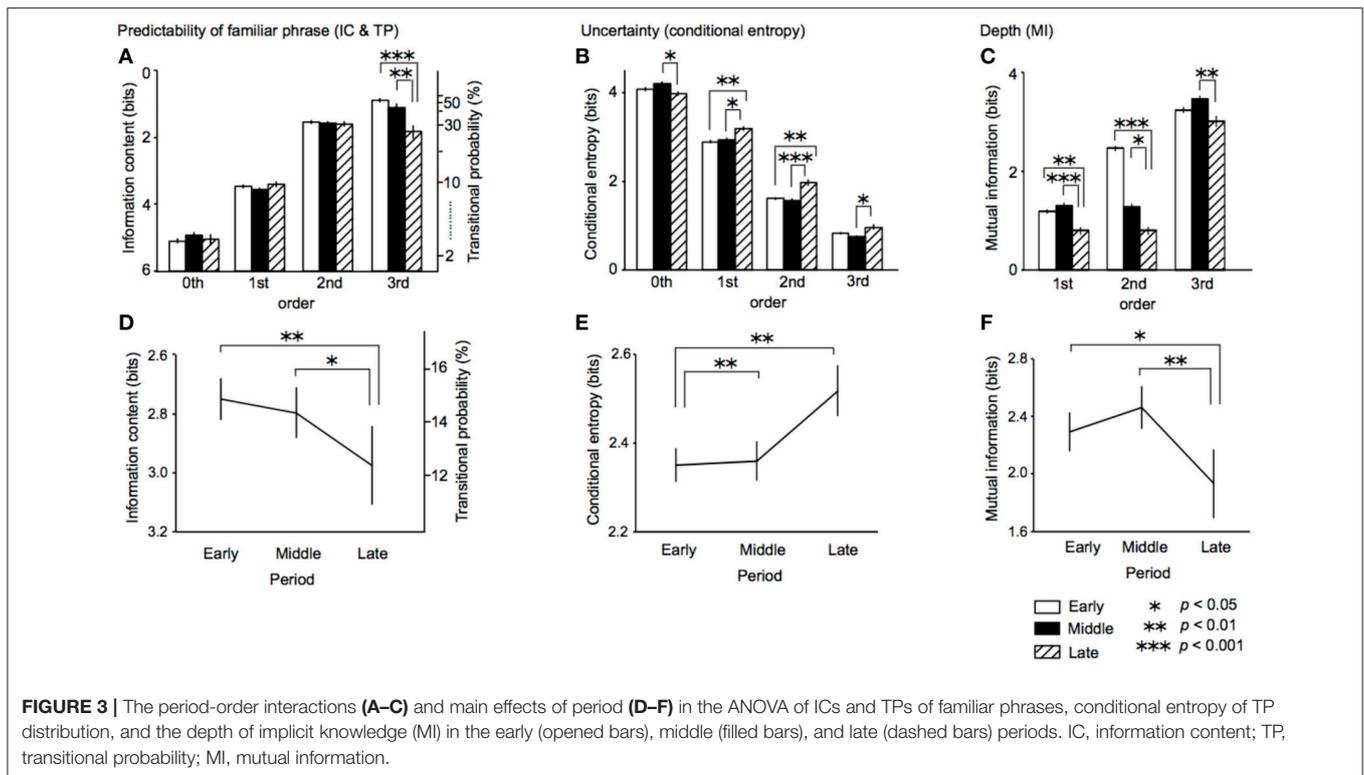
In entropy of whole TP distribution, ANOVA with a within-subject factor order and a between-subjects factor composition period of sonatas was performed. The main effect of period was significant [ $F(2, 29) = 7.58$ ,  $p = 0.002$ ,  $\text{partial}\eta^2 = 0.34$ , early < middle,  $p = 0.005$ ; early < late,  $p = 0.002$ ] (Figure 3E). The period-order interaction was also significant [ $F(6) = 6.68$ ,  $p < 0.001$ ,  $\text{partial}\eta^2 = 0.32$ ] (Figure 3B). The entropies of 0th-order TPs in late period were significantly lower than those in the middle periods ( $p = 0.034$ ). The entropies of 1st-order TPs in late period were significantly higher than those in the early ( $p = 0.004$ ) and middle periods ( $p = 0.014$ ). The entropies of 2nd-order TPs in late period were significantly higher than those in the early ( $p = 0.001$ ) and middle periods ( $p < 0.001$ ). The entropies of 3rd-order TPs in late period were significantly higher than those in the middle periods ( $p = 0.017$ ). The main effect of order was significant [ $F(1.73, 50.30) = 2329.84$ ,  $p < 0.001$ ,  $\text{partial}\eta^2 = 0.99$ ]. The entropies of 0th-order TPs were significantly lower than the 1st-, 2nd-, and 3rd-order TPs (all:  $p < 0.001$ ). The entropies of 1st-order TPs were significantly lower than the 2nd-, 3rd-order TPs (all:  $p < 0.001$ ). The entropies of 2nd-order TPs were significantly lower than the 3rd-order TPs ( $p = 0.007$ ).

### Hierarchy of Statistics: Mutual Information

In the mutual information, an ANOVA with a within-subject factor order and a between-subjects factor composition period was conducted. The main effect of period was significant [ $F(2, 29) = 9.08$ ,  $p = 0.001$ ,  $\text{partial}\eta^2 = 0.39$ , early > late,  $p = 0.020$ ; middle > late,  $p = 0.001$ ] (Figure 3F). The period-order interaction was also significant [ $F(4) = 2.80$ ,  $p = 0.034$ ,  $\text{partial}\eta^2 = 0.16$ ] (Figure 3C). The mutual information of 1st- and 2nd-order TPs in late period were significantly lower than those in the early (1st:  $p = 0.004$ ; 2nd:  $p = 0.012$ ) and middle periods (1st:  $p < 0.001$ ; 2nd:  $p < 0.001$ ). The mutual information of 3rd-order TPs in late period was significantly higher than those in the middle periods ( $p = 0.008$ ). The main effect of order was significant [ $F(1.32, 38.16) = 2350.56$ ,  $p < 0.001$ ,  $\text{partial}\eta^2 = 0.99$ ]. The mutual information of 1st-order TPs were significantly lower than the 2nd-, 3rd-order TPs (all:  $p < 0.001$ ). The 2nd-order TPs were significantly lower than the 3rd-order TPs ( $p < 0.001$ ).

## DISCUSSION

Brain encodes probability distributions and the entropy/uncertainty of musical information (Koelsch et al.,



2018) and mandates the suppression of prediction errors and uncertainty by updating the internal probabilistic model of music that generates predictions and the belief (Kanai et al., 2015). In other words, through musical experience over lifetime, a human generally tries to optimize the internalized probabilistic model for efficient information processing and the uncertainty resolution, which has been regarded as rewards. On the other hand, to explore the maximum rewards of uncertainty resolution, human's behavior appears to be not necessarily directing to efficiency, but sometimes be drove by inefficient, uncertain, and novelty information, which is thought as curiosity (Kagan, 1972; Wittmann et al., 2008; Krebs et al., 2009; Schwartenbeck et al., 2013). Thus, although brain typically minimizes uncertainty of music structure, we sometimes derive pleasure from music with uncertain structure due to curiosity for novelty-seeking behavior by which we anticipate further rewards by uncertainty resolution. The present study, using all the Beethoven's piano sonatas over his lifetime, examined how the probabilistic model and the uncertainty in music fluctuate over a composer's lifetime. The transitional probability and information content (TP), information content (IC), conditional entropy, and mutual information (MI) can be calculated based on  $n$ th-order Markov models. Based on psychological and neurophysiological studies on SL (Harrison et al., 2006; Pearce et al., 2010b; de Zubicaray et al., 2013; Daikoku et al., 2015; Monroy et al., 2017c), these three information can be translated to psychological indices: a tone with lower IC (i.e., higher TPs) may be one that a composer is more likely to choose as a next tone, compared to tones with higher IC, whereas the entropy

and MI are interpreted as uncertainty and the order of the SL, respectively. It was hypothesized that probability, uncertainty, and the order of SL models is fluctuated over Beethoven's lifetime. If so, it may suggest that his curiosity for uncertain and novelty seeking behavior modulate optimization and creativity in human's brain.

The TPs of familiar phrase (i.e., sequences that appear in all 32 sonatas) were decreased in the late periods (Figure 3D), whereas the entropies in music were increased in the late periods (Figure 3E). In other words, there was no significant difference between early and middle periods, while there was significant difference between middle and late periods. Particularly, the 3rd-order TPs in the late period decrease during lifetime (Figure 3A). According to musicological studies, his works in the early period were strongly influenced by his predecessors in classicism whereas his works in the late period show his personal character and experience (Sullivan, 1927). It is believed that he always explored new directions of musical composition and gradually expanded his scope of music over his lifetime (Dahlhaus, 1991; Adorno-Wiesengrund, 1993). The findings of the present study may suggest the hypothesis that the psychological variations over lifetime are reflected in the statistical structure in music. The decreasing of the TPs of familiar phrase and increasing of the entropies may imply that, in the late period, he tried novel composition strategies in which he avoided familiar sequences in the early period, and tried various transitional patterns by which  $n$ th-order TPs are broadly distributed. The previous study detected time-course variation of predictability of familiar phrases over his lifetime (Daikoku, 2018d). The present study,

furthermore, suggested that there seems to be interactions between prediction and uncertainty.

The decreasing of TPs of familiar phrase over Beethoven's lifetime was obvious in the higher-, but not lower-, order models (**Figure 3A**). This may suggest that a higher-, rather than lower-, order statistical structure reflects specific statistical knowledge that is susceptible to experience and novelty seeking behavior. The entropy (i.e., uncertainty) of TP distribution may also support the hypothesis (**Figure 3B**). The entropies of higher-order (1st to 3rd), but not lower-order (0th) models in late period increased compared with those in early period. Furthermore, MI in late periods was lower compared with those early and middle periods (**Figure 3F**). This suggests that, in the late period, each event of tone hardly depends on preceding successive events of tones. Typical Western-classical music has strict syntactic rules based on music theory. Therefore, a forthcoming tone can partially be predicted from preceding successive tones based on the rules. According to previous studies, syntax of musical sequences is partially expressed by conditional probabilities (Rohrmeier and Cross, 2008), although it is not sufficient to account for all of the music syntax. The findings on MI in this study may suggest that, in the late period, the composer avoided a tone that can easily be predicted based on typical transition rules involved in music syntax.

In sum, the present study detected time-course variation of predictability of familiar phrases, uncertainty of whole music, and the depth of SL in music that were composed over Beethoven's lifetime. According to corpus studies, the historical characteristics of music can be extracted based on the era (e.g., Albrecht and Huron, 2014; Gjerdingen, 2014; White, 2014). This indicates that strategies of composition and musical knowledge depend on the era. The present study also suggests that the characteristics of music can be extracted based on the periods within a composer's lifetime. In addition, the higher-order hierarchical structure showed larger time-course variations of both predictability of familiar phrases and uncertainty of whole music. From the psychological perspective, it would be interesting if the higher- (i.e., deep), rather than lower-order statistical knowledge was susceptible to experience and novelty seeking behavior. The present study also suggested that there are interactions between prediction and uncertainty. It is of note, however, that the present study did not directly investigate the composer's statistical knowledge of music, as only the statistics of musical scores were analyzed. Furthermore, the present

study only analyzed one composer, therefore could not discuss universal phenomenon on SL. This suggests that there may be other possible explanations for the findings of the present study. For instance, it might have been Beethoven's plan to compose the sonatas from familiar and lower entropy to unfamiliar and larger entropy based on the statistical structure of music. Future study should investigate SL of music from many composers using interdisciplinary approaches in parallel.

## CONCLUSION

The present study investigated how predictability of familiar phrases that was used in all of music, uncertainty of whole structure, and the order of the probabilistic models in music fluctuates over a composer's lifetime, and discussed the results from psychological perspective within SL framework. The results suggest that the higher-, rather than lower-order statistical knowledge may be susceptible to experience and novelty seeking behavior. The present study also suggested that there might be interactions between prediction and uncertainty. The present study first suggested that uncertainty may be increased in a composer's lifetime, and that the higher-order probabilistic model may be susceptible to experience and novelty seeking behavior over the composer's lifetime. It is suggested that human's curiosity for uncertain and novelty seeking behavior may modulate creativity in human's brain, and that the fluctuations of uncertainty could reflect aesthetic appreciation of music. To more understand brain's predictive function, future study is needed to examine relationships between prediction of familiar phrases and uncertainty perception, using both modeling and experimental approaches in parallel.

## AUTHOR CONTRIBUTIONS

The methodology of the present study and were considered by the author. The author analyzed all of the data and prepared the figures, and wrote the manuscript text.

## FUNDING

This work was supported by Grant-in-Aid for Nakayama Foundation for Human Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCES

- Adorno-Wiesengrund, T. (1993). *Beethoven: The Philosophy of Music Fragments and Texts*. Cambridge: Polity Press.
- Albrecht, J. D., and Huron, D. (2014). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400–1750. *Music Percept. Interdiscipl. J.* 31, 223–243. doi: 10.1525/mp.2014.31.3.223
- Altmann, G. T. (1999). Rule learning by seven-month-old infants and neural networks. *Science* 284, 875a. doi: 10.1126/science.284.5416.875a
- Boucourechliev, A. (1963). *Beethoven (in French)*. Seuil.
- Carreiras, M., Armstrong, B. C., Perea, M., and Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends Cogn. Sci.* 18, 90–98. doi: 10.1016/j.tics.2013.11.005
- Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: news from the front. *Trends Cogn. Sci.* 2, 406–416. doi: 10.1016/S1364-6613(98)01232-7
- Dahlhaus, C. (1991). *Ludwig van Beethoven: Approaches to His Music*. Translated by Mary Whittall. New York, NY: Oxford University Press.
- Daikoku, T. (2018a). Entropy, uncertainty, and the depth of implicit knowledge on musical creativity : computational study of improvisation in melody and rhythm. *Front. Comput. Neurosci.* 12:97. doi: 10.3389/fncom.2018.00097

- Daikoku, T. (2018b). Musical creativity and depth of implicit knowledge: spectral and temporal individualities in improvisation. *Front. Comput. Neurosci.* 12:89. doi: 10.3389/fncom.2018.00089
- Daikoku, T. (2018c). Neurophysiological markers of statistical learning in music and language: hierarchy, entropy, and uncertainty. *Brain Sci.* 8:114. doi: 10.3390/brainsci8060114
- Daikoku, T. (2018d). Time-course variation of statistics embedded in music: corpus study on implicit learning and knowledge. *PLoS ONE* 13:e0196493. doi: 10.1371/journal.pone.0196493
- Daikoku, T., Ogura, H., and Watanabe, M. (2012). The variation of hemodynamics relative to listening to consonance or dissonance during chord progression. *Neurol. Res.* 34, 557–563. doi: 10.1179/1743132812Y.0000000047
- Daikoku, T., Okano, T., and Yumoto, M. (2017c). “Relative difficulty of auditory statistical learning based on tone transition diversity modulates chunk length in the learning strategy,” in *Proceedings of the Biomagnetic* (Sendai), 75.
- Daikoku, T., Takahashi, Y., Futagami, H., Tarumoto, N., and Yasuda, H. (2017a). Physical fitness modulates incidental but not intentional statistical learning of simultaneous auditory sequences during concurrent physical exercise. *Neurol. Res.* 39, 107–116. doi: 10.1080/01616412.2016.1273571
- Daikoku, T., Takahashi, Y., Tarumoto, N., and Yasuda, H. (2018). Auditory statistical learning during concurrent physical exercise and the tolerance for pitch, tempo, and rhythm changes. *Motor Control* 22, 233–244. doi: 10.1123/mc.2017-0006
- Daikoku, T., Yatomi, Y., and Yumoto, M. (2014). Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia* 63, 194–204. doi: 10.1016/j.neuropsychologia.2014.08.028
- Daikoku, T., Yatomi, Y., and Yumoto, M. (2015). Statistical learning of music- and language-like sequences and tolerance for spectral shifts. *Neurobiol. Learn. Mem.* 118, 8–19. doi: 10.1016/j.nlm.2014.11.001
- Daikoku, T., Yatomi, Y., and Yumoto, M. (2016). Pitch-class distribution modulates the statistical learning of atonal chord sequences. *Brain Cogn.* 108, 1–10. doi: 10.1016/j.bandc.2016.06.008
- Daikoku, T., Yatomi, Y., and Yumoto, M. (2017b). Statistical learning of an auditory sequence and reorganization of acquired knowledge: a time course of word segmentation and ordering. *Neuropsychologia* 95, 1–10. doi: 10.1016/j.neuropsychologia.2016.12.006
- Daikoku, T., and Yumoto, M. (2017). Single, but not dual, attention facilitates statistical learning of two concurrent auditory sequences. *Sci. Rep.* 7:10108. doi: 10.1038/s41598-017-10476-x
- Daikoku, T., and Yumoto, M. (2019). Concurrent statistical learning of ignored and attended sound sequences: an MEG study. *Front. Hum. Neurosci.* 13:102. doi: 10.3389/fnhum.2019.00102
- de Zubicaray, G., Arciuli, J., and McMahon, K. (2013). Putting an “End” to the motor cortex representations of action words. *J. Cogn. Neurosci.* 25, 1957–1974. doi: 10.1162/jocn\_a\_00437
- Dienes, Z., Altmann, G., and Gao, S. (1999). Mapping model across domains a neural feedback : network of implicit of transfer of implicit knowledge. *Cogn. Sci.* 23, 53–82. doi: 10.1207/s15516709cog2301\_3
- Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Cambridge, MA: MIT Press.
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402\_1
- Feher, O., Ljubičić, I., Suzuki, K., Okanoya, K., and Tchernichovski, O. (2016). Statistical learning in songbirds : from self- tutoring to song culture1. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160053. doi: 10.1098/rstb.2016.0053
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–224. doi: 10.1080/17588928.2015.1020053
- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2014). *The Anatomy of Choice: Dopamine and Decision-Making Subject Collections The Anatomy of Choice: Dopamine and Decision-Making*. Available online at: [http://rstb.royalsocietypublishing.org/content/369/1655/20130481.full.html#ref-list-1%5Cnhttp://dx.doi.org/10.1098/rstb.2013.0481](http://rstb.royalsocietypublishing.org/content/369/1655/20130481.full.html#related-urls%5Cnhttp://rstb.royalsocietypublishing.org/content/369/1655/20130481.full.html#ref-list-1%5Cnhttp://dx.doi.org/10.1098/rstb.2013.0481)
- Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., and Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage* 54, 2267–2277. doi: 10.1016/j.neuroimage.2010.10.038
- Gjerdingen, R. O. (2014). “Historically Informed” corpus studies. *Music Percept. Interdiscipl. J.* 31, 192–204. doi: 10.1525/MP.2014.31.3.192
- Hansen, N. C., and Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Front. Psychol.* 5:1052. doi: 10.3389/fpsyg.2014.01052
- Harrison, L. M., Duggins, A., and Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Netw.* 19, 535–546. doi: 10.1016/j.neunet.2005.11.002
- Hasson, U. (2017). The neurobiology of uncertainty : implications for statistical learning. *Philos. Trans. R. Soc. B* 372:20160048. doi: 10.1098/rstb.2016.0048
- Hochreiter, S., and Uergen Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Johnson, D., Tyson, A., and Wnter, R. (1985). *The Beethoven Sketchbooks*. Oxford: Clarendon.
- Jones, M. N., and Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* 114, 1–37. doi: 10.1037/0033-295X.114.1.1
- Kagan, J. (1972). Motives and development. *J. Personal. Soc. Psychol.* 22, 51–66. doi: 10.1037/h0032356
- Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140169. doi: 10.1098/rstb.2014.0169
- Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.* 55, 271–304. doi: 10.1146/annurev.psych.55.090902.142005
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface* 15:20170792. doi: 10.1098/rsif.2017.0792
- Knill, D. C., and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666
- Koelsch, S., Vuust, P., and Friston, K. (2018). Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* 23, 63–77. doi: 10.1016/j.tics.2018.10.006
- Krebs, R. M., Schott, B. H., Schütze, H., and Düzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia* 47, 2272–2281. doi: 10.1016/j.neuropsychologia.2009.01.015
- Landauer, T. K., and Dumais, S. T. (1997). A solution to Platos problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211
- Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* 116, 75–98. doi: 10.1037/0033-2909.116.1.75
- Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208. doi: 10.3758/BF03204766
- Monroy, C., Meyer, M., Gerson, S., and Hunnius, S. (2017b). Statistical learning in social action contexts. *PLoS ONE* 12:e0177261. doi: 10.1371/journal.pone.0177261
- Monroy, C. D., Gerson, S. A., Domínguez-Martínez, E., Kaduk, K., Hunnius, S., and Reid, V. (2017a). Sensitivity to structure in action sequences: an infant event-related potential study. *Neuropsychologia* 126, 92–101. doi: 10.1016/j.neuropsychologia.2017.05.007
- Monroy, C. D., Meyer, M., Schröer, L., Gerson, S. A., and Hunnius, S. (2017c). The infant motor system predicts actions based on visual statistical learning. *Neuroimage* 185, 947–954. doi: 10.1016/j.neuroimage.2017.12.016
- Nastase, S., Iacovella, V., and Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* 35, 1111–1128. doi: 10.1002/hbm.22238
- O’Reilly, J. X., Jbabdi, S., and Behrens, T. E. (2012). How can a Bayesian approach inform neuroscience? *Eur. J. Neurosci.* 35, 1169–1179. doi: 10.1111/j.1460-9568.2012.08010.x
- Parr, T., and Friston, K. J. (2018). The anatomy of inference: generative models and brain structure. *Front. Comput. Neurosci.* 12:90. doi: 10.3389/fncom.2018.00090

- Parr, T., Rees, G., and Friston, K. J. (2018). Computational neuropsychology and bayesian inference. *Front. Hum. Neurosci.* 12:61. doi: 10.3389/fnhum.2018.00061
- Pearce, M., and Wiggins, G. A. (2006). Expectation in melody: the influence of context and learning. *Music Percept.* 23, 377–405. doi: 10.1525/mp.2006.23.5.377
- Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Unpublished Doctoral Thesis, City University London.
- Pearce, M. T., Müllensiefen, D., and Wiggins, G. A. (2010b). The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. *Perception* 39, 1367–1391. doi: 10.1068/p6507
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., and Bhattacharya, J. (2010a). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* 50, 302–313. doi: 10.1016/j.neuroimage.2009.12.019
- Pearce, M. T., and Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Top. Cogn. Sci.* 4, 625–652. doi: 10.1111/j.1756-8765.2012.01214.x
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference SanMateo*. Morgan Kaufmann Series in Representation and Reasoning. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* 10, 233–238. doi: 10.1016/j.tics.2006.03.006
- Perruchet, P., and Vinter, A. (1998). PARSER: A model for word segmentation. *J. Mem. Lang.* 39, 246–263.
- Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progr. Neurobiol.* 134, 17–35. doi: 10.1016/j.pneurobio.2015.09.001
- Przynda, E., Zeng, T., Maves, K., Arkin, C., and Loui, P. (2017). Jazz musicians reveal role of expectancy in human creativity. *Brain Cogn.* 119, 45–53. doi: 10.1016/j.bandc.2017.09.008
- Rogers, T. T., and McClelland, J. L. J. (2004). Semantic cognition: a parallel distributed processing approach. *Attend. Perform.* 42:5:439. doi: 10.7551/mitpress/6161.001.0001
- Rohrmeier, M., and Cross, I. (2008). “Statistical Properties of Tonal Harmony in Bach’s Chorales,” in *Proc 10th Intl Conf on Music Perception and Cognition* (Sapporo). 6, 123–1319.
- Rohrmeier, M., and Rebuschat, P. (2012). Implicit learning and acquisition of music. *Top. Cogn. Sci.* 4, 525–553. doi: 10.1111/j.1756-8765.2012.01223.x
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations:  $\alpha$ -gamma versus  $\theta$ -gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., and McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends Cogn. Sci.* 19, 86–91. doi: 10.1016/j.tics.2014.12.001
- Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connect. Sci.* 18, 173–187. doi: 10.1080/09540090600768658
- Schwartenbeck, P., FitzGerald, T., Dolan, R. J., and Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Front. Psychol.* 4:710. doi: 10.3389/fpsyg.2013.00710
- Servan-Schreiber, E., and Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 592–608. doi: 10.1037/0278-7393.16.4.592
- Shen, L., Fishbach, A., and Hsee, C. K. (2015). The motivating-uncertainty effect: uncertainty increases resource investment in the process of reward pursuit. *J. Consumer Res.* 41, 1301–1315. doi: 10.1086/679418
- Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., and Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw.* 18, 225–230. doi: 10.1016/j.neunet.2004.12.004
- Sullivan, J. (1927). *Beethoven: His Spiritual Development*. New York, NY: A.A. Knopf.
- Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.* 139, 792–814. doi: 10.1037/a0030801
- Tishby, N., and Polani, D. (2011). “Information theory of decisions and actions,” in *Perception-action Cycle*, eds V. Cutsuridis, A. Hussain, and J. G. Taylor (New York, NY: Springer), 601–636.
- Turk-browne, N. B., Scholl, B. J., Chun, M. M., and Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* 21, 1934–1945. doi: 10.1162/jocn.2009.21131
- White, C. W. (2014). Changing styles, changing corpora, changing tonal models. *Music Percept. Interdiscipl. J.* 31, 244–253. doi: 10.1525/mp.2014.31.3.244
- Wiggins, G. A. (2018). Creativity, information, and consciousness: the information dynamics of thinking. *Phys. Life Rev.* 1, 1–39. doi: 10.1016/j.pprev.2018.05.001
- Wiggins, G. A., and Sanjekdar, A. (in press). Consolidation as re-representation: revising the meaning of memory. *Front. Psychol.* doi: 10.3389/fpsyg.2019.00802
- Wittmann, B. C., Daw, N. D., Seymour, B., and Dolan, R. J. (2008). striatal activity underlies novelty-based choice in humans. *Neuron* 58, 967–973. doi: 10.1016/j.neuron.2008.04.027
- Yumoto, M., and Daikoku, T. (2016). “IV Auditory system. 5 basic function,” in *Clinical Applications of Magnetoencephalography*, eds S. Tobimatsu and R. Kakigi (Springer).
- Yumoto, M., and Daikoku, T. (2018). Neurophysiological Studies on Auditory Statistical Learning [in Japanese] 聴覚刺激の統計学習の神経生理学的研究 *Jpn. J. Cogn. Neurosci. 認知神経科学.* 20, 38–43. doi: 10.11253/ninchishinkeikagaku.20.38

**Conflict of Interest Statement:** 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.

Copyright © 2019 Daikoku. 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.