Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation

Marcus T. Pearce[1,2]

[1]Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. [2]Centre for Music in the Brain, Aarhus University, Aarhus, Denmark

Address for correspondence: Dr. Marcus T. Pearce, Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK. marcus.pearce@qmul.ac.uk

Music perception depends on internal psychological models derived through exposure to a musical culture. It is hypothesized that this musical enculturation depends on two cognitive processes: (1) statistical learning, in which listeners acquire internal cognitive models of statistical regularities present in the music to which they are exposed; and (2) probabilistic prediction based on these learned models that enables listeners to organize and process their mental representations of music. To corroborate these hypotheses, I review research that uses a computational model of probabilistic prediction based on statistical learning (the information dynamics of music (IDyOM) model) to simulate data from empirical studies of human listeners. The results show that a broad range of psychological processes involved in music perception—expectation, emotion, memory, similarity, segmentation, and meter—can be understood in terms of a single, underlying process of probabilistic prediction using learned statistical models. Furthermore, IDyOM simulations of listeners from different musical cultures demonstrate that statistical learning can plausibly predict causal effects of differential cultural exposure to musical styles, providing a quantitative model of cultural distance. Understanding the neural basis of musical enculturation will benefit from close coordination between empirical neuroimaging and computational modeling of underlying mechanisms, as outlined here.

Keywords: music perception; enculturation; statistical learning; probabilistic prediction; IDyOM

Introduction

Musical styles comprise cultural constraints on the compositional choices made by composers, which can be distinguished both from constraints reflecting universal laws (of nature and human perception or production of sound) and specific within-culture, nonstyle-defining compositional strategies employed by particular (groups of) composers in particular circumstances.[1] As recognized by Leonard Meyer in his early writing,[2] these constraints can be viewed as complex, probabilistic grammars defining the syntax of a musical style,[3,4] which are acquired as internal cognitive models of the style by composers, performers, and listeners. This enables successful communication of musical meaning between composers and performers and between performers and listeners.[2,5–8]

Unlike many other general theories of music cognition,[9–12] this approach elegantly encompasses the idea that listeners exposed to different musical styles will differ in their psychological processing of music. It provides naturally for musical enculturation, the process by which listeners internalize the regularities and constraints defining and distinguishing musical styles and cultures. My purpose here is to elaborate Meyer's proposals by putting forward a computational model that is capable of learning the probabilistic structure of musical styles and examining whether the model successfully simulates the perception of mature, enculturated listeners across a broad range of cognitive processes and whether the model also simulates enculturation in musical styles.

I propose two hypotheses about the psychological and neural mechanisms involved in musical enculturation. According to these hypotheses, listeners use implicit statistical learning through passive exposure to acquire internal cognitive models of

1

the regularities defining the syntax of a musical style; furthermore, they use probabilistic prediction based on the learned internal model to generate probabilistic predictions that underlie their perception and emotional experience of music. In other words, while existing theoretical approaches propose several distinct cognitive mechanisms underlying perception and emotional experience of music,[6,9,12] here probabilistic prediction is put forward as a foundational mechanism underpinning other psychological processes in music perception. To substantiate these rather bold proposals, I introduce a computational model of probabilistic prediction based on statistical learning and present empirical results showing that the same model simulates a wide range of key cognitive processes in music perception (expectation, uncertainty, emotional experience, recognition memory, similarity perception, phrase-boundary perception, and metrical inference). Finally, I demonstrate how the same model can be used to simulate enculturation and generate predictions about individual differences in perception resulting from enculturation in different musical styles.

Statistical learning and predictive processing

Two hypotheses guide the present approach to understanding music cognition. The statistical learning hypothesis (SLH) states that musical enculturation is a process of implicit statistical learning in which listeners progressively acquire internal models of the statistical and structural regularities present in the musical styles to which they are exposed, over short (e.g., an individual piece of music) and long time scales (e.g., an entire lifetime of listening). The probabilistic prediction hypothesis (PPH) states that, while listening to new music, an enculturated listener applies models learned via the SLH to generate probabilistic predictions that enable them to organize and process their mental representations of the music and generate culturally appropriate responses.

Probabilistic prediction is the process by which the brain estimates the likelihood with which an event is likely to occur. With respect to musical listening, this corresponds to the probability of different possible continuations of the music (e.g., the next note or chord and its temporal position). But where

do the probabilities come from? *Statistical learning* is the process by which individuals learn the statistical structure of the sensory environment and is thought to proceed automatically and implicitly.[13,14] This makes the theory general purpose in that it can potentially apply to any musical style, but also beyond music to other domains, such as language or visual perception. It also means that the theory can explicitly account for the effects of experience on music perception, including differences between listeners of different ages and different musical cultures and with different levels of musical training and stylistic exposure.

Research has established statistical learning and predictive processing as important mechanisms in many areas of cognitive science and cognitive neuroscience,[15–17] including language processing,[13,18–21] visual perception,[22–25] and motor sequencing.[26] In particular, *predictive coding*[15,17,27–29] is a general theory of the neural and cognitive processes involved in perception, learning, and action. According to the theory, an internal model of the sensory environment compares top-down predictions about the future with the actual events that transpire, and error signals generated from the comparison drive learning to improve future predictions by updating the model to reduce error. These prediction errors occur at a series of hierarchical levels, each reflecting an integration of information over successively larger temporal or spatial scales. Top-down predictions are precision weighted such that more specific predictions (i.e., those more sharply focused on a single outcome) generate greater predictions errors. In the auditory modality, there is some evidence supporting hierarchical predictive coding for perception of nonmusical pitch sequences[30,31] and speech,[32] though not all aspects of the theory have been empirically substantiated.[33] Vuust and colleagues have proposed a predictive coding theory of rhythmic incongruity.[34]

As noted above, the idea that musical appreciation depends on probabilistic expectations has a venerable history, going back at least to Meyer's 1957 article.[2] However, until relatively recently, empirical psychological research had been limited by the lack of a plausible computational model that simulates the psychological processes of statistical learning and probabilistic prediction. Recent research using the information dynamics of music (IDyOM) model[35] has successfully implemented and extended

Meyer's proposals and subjected them to empirical testing.

IDyOM

IDyOM[35] is a computational model of auditory cognition that uses statistical learning and probabilistic prediction to acquire and process internal representations of the probabilistic structure of a musical style. Given exposure to a corpus of music, IDyOM learns the syntactic structure present in the corpus in terms of sequential regularities determining the likelihood of a particular event appearing in a particular context (e.g., the pitch or timing of a note at a particular point in a melody). IDyOM is designed to capture several intuitions about human predictive processing of music.

First, expectations are dependent on knowledge acquired during long-term exposure to a musical style,[36–38] but listeners are also sensitive to repeated patterns within a piece of music.[39–41] Therefore, IDyOM acquires probabilistic knowledge about a musical style through statistical learning from a large corpus reflecting a listener's long-term exposure to a musical style (simulated by IDyOM's long-term model (LTM), which is exposed to a large corpus of music in a given style). IDyOM also acquires knowledge about the structure of the music it is currently processing through short-term incremental, dynamic statistical learning of repeated structure experienced during the current listening episode (simulated by IDyOM's short-term model, which is emptied of any learned content before processing each new piece of music). Second, expectations are dependent on the preceding context, such that different expectations are generated when the context changes.[42] In modeling terms, the length of the context used to make a prediction is called the *order* of the model. For example, a model that predicts continuations based on the preceding two events is a second-order model (sometimes referred to as a trigram model). IDyOM is a variable-order Markov model[43–46] that adaptively varies the order used for each context encountered during prediction. IDyOM also combines higher order predictions, which are structurally very specific to the context but may be statistically unreliable (because longer contexts appear less frequently, with fewer distinct continuations, in the prior experience of the model), with lower order predictions (based on shorter contexts) that are more structurally generic but also

more statistically robust (since they have appeared more frequently with a wider range of continuations). IDyOM computes a weighted mixture of the predictions made by models of all orders lower than the adaptively selected order for the context.

Third, research has demonstrated that listeners process music using multiple psychological representations of pitch[37,47,48] (e.g., pitch height, pitch chroma, pitch interval, and pitch contour scale degree) and time[49] (e.g., absolute duration-based and relative beat-based representations). Accordingly, IDyOM is able to create models for multiple attributes of the musical surface and combine the predictions made by these models. For example, it can be configured to predict pitch with a combination of two models for pitch interval and scale degree (see pi and sd in the third panel of Fig. 1). Alternatively, it can be configured to predict note onsets with a combination of two models for interonset interval and sequential interonset interval ratios (see ioi and ioi-ratio in the second panel of Fig. 1).[35,50] Each of the models generates predictive distributions for a single property of the next note (e.g., pitch or onset time), which are combined separately for the long-term and short-term models before being combined into the final pitch distribution. Finally, listeners generate expectations for both the pitch[37] and the timing of notes.[36] Therefore, IDyOM applies the same process of probabilistic prediction described above in parallel to predict the pitch and onset time of the next note and computes the final probability of the note as the joint likelihood of its pitch and onset time. Given evidence that pitch structure and temporal structure are processed by listeners independently in some situations but interactively in others,[51–53] IDyOM can process pitch and temporal attribute independently (using separate models whose probabilistic output is subsequently combined) or interactively using a single model of an attribute that links the two domains (e.g., by representing notes as a pair of scale degree and interonset interval ratio, see sd ⊗ ioi-ratio in the lower panel of Fig. 1).

IDyOM acquires knowledge about the structure of music through statistical learning of variable-length sequential dependencies between events in the music to which it is exposed and, while processing music event by event, generates expectations for the next event (e.g., the note that continues a melody) in the form of a probability distribution
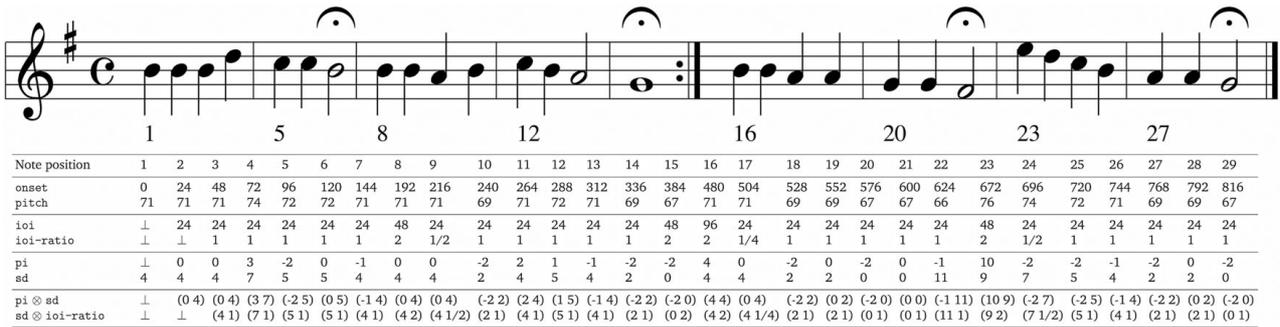
Note position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
onset	0	24	48	72	96	120	144	192	216	240	264	288	312	336	384	480	504	528	552	576	600	624	672	696	720	744	768	792	816
pitch	71	71	74	72	72	71	71	71	71	69	71	72	71	69	71	71	71	69	69	67	67	66	74	74	72	71	69	69	69
ioi	⊥	24	24	24	24	24	24	48	24	24	24	24	24	24	48	96	24	24	24	24	24	48	24	24	24	24	24	24	
ioi-ratio	⊥	⊥	1	1	1	1	1	2	1/2	1	1	1	1	1	2	2	1/4	1	1	1	1	1	2	1/2	1	1	1	1	1
pi	⊥	0	0	3	-2	0	-1	0	0	-2	2	1	-1	-2	-2	4	0	-2	0	-2	0	-1	10	-2	-2	0	-2		
sd	4	4	4	7	5	5	4	4	4	2	4	5	4	2	0	4	4	2	2	0	0	11	9	7	5	4	2	2	0
pi ⊗ sd	⊥	(0 4)	(0 4)	(3 7)	(-2 5)	(0 5)	(-1 4)	(0 4)	(0 4)	(-2 2)	(2 4)	(1 5)	(-1 4)	(-2 2)	(-2 0)	(4 4)	(0 4)	(-2 2)	(0 2)	(-2 0)	(0 0)	(-1 11)	(10 9)	(-2 7)	(-2 5)	(-1 4)	(-2 2)	(0 2)	(-2 0)
sd ⊗ ioi-ratio	⊥	⊥	(4 1)	(7 1)	(5 1)	(5 1)	(4 1)	(4 2)	(4 1/2)	(2 1)	(4 1)	(5 1)	(4 1)	(2 1)	(0 2)	(4 2)	(4 1/4)	(2 1)	(2 1)	(0 1)	(0 1)	(11 1)	(9 2)	(7 1/2)	(5 1)	(4 1)	(2 1)	(2 1)	(0 1)

Figure 1. A chorale harmonized by J. S. Bach (BWV 379) showing examples of the input representations used by IDyOM. The first vertical panel shows the basic event space in which musical events are represented in terms of their chromatic pitch (pitch as an MIDI note number, where 60 = middle C) and onset time (onset, where 24 corresponds to a crotchet duration in this example). The second panel shows attributes derived from onset, including the interonset interval (ioi) and the ratio between successive interonset intervals (ioi-ratio). Note that ioi is undefined (denoted by ⊥) for the first note in a melody, while ioi-ratio is undefined for the first two notes. The third panel shows attributes derived from pitch, including the pitch interval in semitones formed between a note and its immediate predecessor (pi) and chromatic scale degree (sd) or distance in semitones from the tonic pitch (G or 67 in this example). The final panel shows two examples of linked attributes: first, linking pitch interval with scale degree (pi ⊗ sd) affording learning of combined melodic and tonal structure (the IDyOM models used Figs. 2–4 to use this linked attribute); second, linking pitch and temporal attributes (sd ⊗ ioi-ratio), affording learning of combined tonal and rhythmic structure.

(P) that assigns a probability to each possible next event, conditioned upon the preceding musical context and the prior musical experience of the model. The information-theoretic quantity *entropy* ($H = -\sum_{p \in P} p \log p$) reflects the uncertainty of the prediction before the next event is heard—if every continuation is equiprobable, entropy will be maximum and the prediction highly uncertain, while if one continuation has very high probability, entropy will be low and the prediction very certain.[54,55] When the next event actually arrives, it may have a high probability, making it expected, or a low probability, making it unexpected. Rather than dealing with raw probabilities, *information content* ($h = -\log_2 p$) provides a measure that is more numerically stable and has a meaningful information-theoretic interpretation in terms of compressibility.[44,54] Information content (IC) reflects how unexpected the model finds an event in a particular context. Compression involves removing redundant information from a signal, which has been proposed as a central part of perceptual pattern recognition, and it has been argued that compression provides a measure of the strength of evidence for psychological interpretations of perceptual data (see also below).[56–58]

Figure 2 applies IDyOM to excerpts from Schubert's *Octet for Strings and Winds*, which is discussed in detail by Leonard Meyer in his book *Explaining Music* (p. 219, example 121).[59] Since Meyer's analysis pertains to pitch structure, IDyOM is configured only to predict pitch in this example. Referring to the penultimate note in the second bar (Fig. 2A), Meyer writes, "The continuation is triadic–to G– but in the wrong register. The realization therefore is only provisional." IDyOM reflects this analysis, estimating a lower probability for the G_4 that actually follows than for the G_5 that is anticipated (0.015 versus 0.186). When the theme returns in bars 21– 22 (Fig. 2B), Meyer writes that "The triadic implications of the motive are satisfactorily realized . . . But instead of the probable G, A follows—as part of the dominant of D minor (V/II)." IDyOM reflects this analysis, estimating a lower probability for the A_5 that actually follows than for the G_5 that is, again, anticipated (0.013 versus 0.186). The relatively high probability (0.344) assigned by IDyOM to the D_5 can be attributed to another melodic process discussed by Meyer called *gap-fill* in which a larger interval that spans more than one adjacent scale degree (the gap, C_5–E_5 in this case) creates an implication for the subsequent melodic movement to fill in the intervening scale degrees skipped over (here D_5). The relatively high probability (0.189) assigned by IDyOM to the E_5 reflects a general implication for small intervals (here a unison, the smallest interval possible).[10] Meyer adds that "The poignancy of the A is the result not only of its deviant character and its harmonic context, but of the fact that the larger interval—a sixth rather than a fifth–acts both as a triadic continuation and as a gap implying
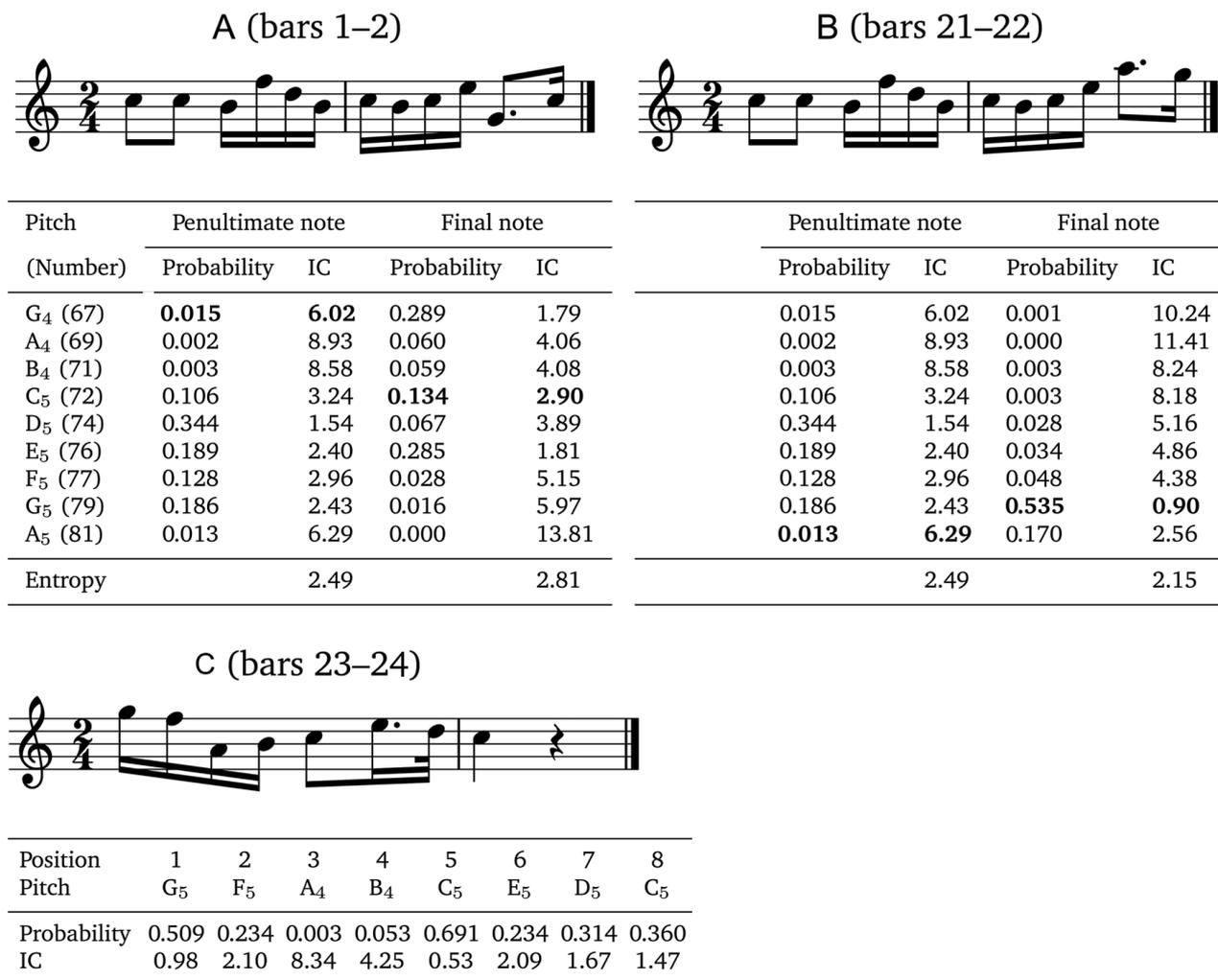
A (bars 1–2)

Pitch	Penultimate note		Final note	
(Number)	Probability	IC	Probability	IC
G_4 (67)	**0.015**	**6.02**	0.289	1.79
A_4 (69)	0.002	8.93	0.060	4.06
B_4 (71)	0.003	8.58	0.059	4.08
C_5 (72)	0.106	3.24	**0.134**	**2.90**
D_5 (74)	0.344	1.54	0.067	3.89
E_5 (76)	0.189	2.40	0.285	1.81
F_5 (77)	0.128	2.96	0.028	5.15
G_5 (79)	0.186	2.43	0.016	5.97
A_5 (81)	0.013	6.29	0.000	13.81
Entropy		2.49		2.81

B (bars 21–22)

Penultimate note		Final note	
Probability	IC	Probability	IC
0.015	6.02	0.001	10.24
0.002	8.93	0.000	11.41
0.003	8.58	0.003	8.24
0.106	3.24	0.003	8.18
0.344	1.54	0.028	5.16
0.189	2.40	0.034	4.86
0.128	2.96	0.048	4.38
0.186	2.43	**0.535**	**0.90**
0.013	**6.29**	0.170	2.56
	2.49		2.15

C (bars 23–24)

Position	1	2	3	4	5	6	7	8
Pitch	G_5	F_5	A_4	B_4	C_5	E_5	D_5	C_5
Probability	0.509	0.234	0.003	0.053	0.691	0.234	0.314	0.360
IC	0.98	2.10	8.34	4.25	0.53	2.09	1.67	1.47

Figure 2. Three excerpts from the fourth movement of Schubert's *Octet in F Major* (D.803) taken from bars 1–2 (A), 21–22 (B), and 23–24 (C). (A and B) Probabilities and corresponding information content (IC) and entropy generated by IDyOM for the penultimate and final notes in each excerpt. At each point in processing, IDyOM estimates a probability distribution for the 37 chromatic pitches from B_2 (47) to B_5 (83), most of which have very low probabilities. For purposes of illustration, only the diatonic pitches between G_4 and A_5 are shown, including those that actually appear in the octet (highlighted in bold font). The entropy of the prediction is computed over the full 37-pitch alphabet. (C) The probability and IC for each note appearing in the final two bars of the theme. In all cases, IDyOM was configured to predict pitch with an attribute linking melodic pitch interval and chromatic scale degree (pi ⊗ sd, see Fig. 1) using both the short-term and long-term models, the latter trained on 903 folk songs and chorales (data sets 1, 2, and 9 from table 4.1 in Ref. 35 comprising 50,867 notes).

descending motion toward closure." Again, IDyOM reflects Meyer's analysis: the penultimate A_5 in bar 22 allows IDyOM to predict the continuation with greater certainty than it could following the G_4 in bar 2 (reflected in the lower entropy of 2.15 compared with 2.81), making the subsequent descent to the G_5 (finally making its appearance, resolving the tension introduced by the preceding deviations from anticipated continuation) much more probable than it would have been following the penultimate G_4 in bar 2 (0.535 versus 0.016) and indeed more probable than the C_5 that actually followed in bar 2 (0.535 versus 0.134). As shown in Figure 2C, IDyOM also strongly anticipates the restatement of the G_5 on the downbeat of bar 23, while the cadence toward tonal closure in the final two bars is characterized overall by high probability in IDyOM analysis (average probability = 0.3).

The features described above make IDyOM capable of simulating human cognitive processing of music to an extent that was simply not possible when Meyer was writing in the 1950s. Nonetheless, there are limits to the kinds of music (and musical structure) that IDyOM can process. To date,

research has focused on modeling melodic music, generating predictions for the pitch and timing of individual notes based on the preceding melodic context (Figs. 1 and 2). However, recent research has extended IDyOM to modeling expectations for harmonic movement[60] and has simulated melodic and harmonic expectations separately for tonal cadences in classical string quartets.[61] Current research is also extending IDyOM to polyphonic music represented as parallel sequences, each containing a voice or perceptual stream, for which separate predictions are generated.[62] In time, this approach may be capable of modeling complex aspects of polyphonic structure, such as stream segregation, and interactions between harmony and melody (e.g., the ways in which harmonic syntax constrains melodic expectations). IDyOM does require its musical input to be represented symbolically, which means that it cannot process aspects of music that rely on timbral, dynamic, or textual changes. Meyer refers to these parameters as secondary, since they do not usually take primary responsibility for bearing the syntax of a musical style (at least in the Western styles he is concerned with), and suggests that they operate differently from primary parameters (e.g., melody, harmony, and rhythm), though they may reinforce or diminish the effects of these syntactic parameters (which could be simulated as an independent process that is subsequently combined with IDyOM's predictive output). Where they take a prominent role in a musical style (e.g., electroacoustic music, electronic music, and soundscapes), I would predict that expectations are psychologically generated in a rather different way (based on extrapolation of physical properties, such as continuous changes in timbre, dynamics, or texture) that is not captured by IDyOM's structural processing of music.

Finally, it is instructive to draw parallels and contrasts between IDyOM and other modeling approaches, including rule-based models, adaptive oscillator models, and general probabilistic theories of brain function. Rule-based models have been proposed for simulating pitch expectations[10,42,63–65] and temporal expectations.[9,12,66–68] Such models are characterized by a collection of fixed rules for determining the onset and pitch of a musical event in a given context. Examples for pitch expectations are the *implication-realization* theory[10,63] consisting of numerical rules defining the implications made by one pitch interval for the successive interval and the

tonal pitch space theory[69] consisting of numerical rules characterizing harmonic and melodic tension in terms of tonal stability and attraction. An example of a rule-based approach to modeling temporal expectations is Melisma,[70] which uses preference rules to select the preferred meter for a rhythm from a set of possible meters defined by well-formedness rules. Rule-based models depend heavily on the expertise of their designers and are often useful for analytical purposes, since the degree to which a musical example follows the rules can be interrogated perspicuously. However, since the rules are fixed and impervious to experience, such models cannot be used to simulate the acquisition of cognitive models of musical styles through enculturation (though they may describe the end result of this process for a given culture).

A rather different approach to simulating expectation is to use nonlinear dynamical systems, consisting of oscillators operating at different periods with specific phase and period relations.[71–74] In this approach, metrical expectations emerge from the resonance of coupled oscillators that entrain to temporal periodicities in the stimulus. A related oscillatory approach has been used to predict cross-cultural invariances in perceived tonal stability.[75] Since these models naturally imply an explanation of pitch and temporal processing in terms of stimulus structure, they do not provide a compelling account of enculturation (though it has been claimed that it is potentially compatible with Hebbian learning).[71] It is possible that oscillator-based models and the mechanisms of statistical learning and probabilistic processing implemented in IDyOM are complementary in simulating different aspects of expectation (e.g., enculturated versus nonenculturated processing) or by operating at different Marrian levels of description.[76]

More broadly, there are relationships between IDyOM and the general mechanisms of brain function hypothesized by predictive coding theory. First, although the representations in IDyOM input are particular to auditory stimuli, there is nothing else domain-specific in IDyOM's design and, in fact, variable-order Markov models are widely used in statistical language modeling[77,78] and universal lossless data compression.[44–46] Second, IC is a measure of prediction error,[15] as posited by predictive coding theory, between the event that actually follows and the top-down prediction made by IDyOM based

on prior learning: high IC implies greater prediction error and vice versa. Third, the combination of distributions produced by the subcomponent models within IDyOM is weighted by entropy such that models generating more certain predictions have higher weights.[35,50] This is similar to the precision weighting of prediction errors in predictive coding theory.[15]

Probabilistic prediction in music cognition

To substantiate the proposal that probabilistic prediction constitutes a foundational process in music perception, the following sections review empirical results in which IDyOM models, after training on a corpus of Western tonal music, account well for the performance of Western participants (with long-term exposure to Western tonal music) on a range of tasks, reflecting key psychological processes involved in music perception.

Expectation and uncertainty

IDyOM has been shown to predict accurately Western listeners' melodic pitch expectations in behavioral, physiological, and electroencephalography (EEG) studies using a range of experimental designs, including the probe-tone paradigm,[35,79] visually guided probe-tone paradigm,[80,81] a gambling paradigm,[35] continuous expectedness ratings,[82,83] and an implicit reaction-time task to judgments of timbral change.[81] In these studies, IC accounts for up to 83% of the variance in listeners' pitch expectations. Furthermore, listeners show greater uncertainty when generating pitch expectations in low-entropy contexts than they do in high-entropy contexts, as predicted by IDyOM.[79] In many circumstances, IDyOM provides a more accurate model of listeners' pitch expectations than static rule-based models,[10,63] which cannot account for enculturation.[35,79,80] Figure 3 illustrates the relationship between IC and listeners' expectations throughout a Bach chorale melody, using data from an empirical study of pitch expectations reported by Manzara et al.[84]

Furthermore, there is evidence that IC predicts neural measures of expectation violation. EEG studies with artificially constructed stimuli have identified an increased early negativity emerging around the latency of the auditory N1 (80–120 ms) for incongruent melodic endings in artificially composed stimuli.[85–90] Omigie et al.

generalized these findings to more complex, real-world musical stimuli, taking continuous EEG recordings while participants listened to a collection of isochronous English hymn melodies.[91] The peak amplitude of the N1 component decreased significantly from high-IC events through medium-IC events to low-IC events, and this effect was slightly right lateralized. Furthermore, across all notes in all 58 stimuli, the amplitude of the early negative potential correlated significantly with IC. Alongside the behavioral studies reviewed above,[35,79–83] these results show that IDyOM's IC also accounts well for neural markers of pitch expectation. It remains to be seen whether this holds true for neural measures of temporal expectation.[92]

Emotional experience

Expectation is thought to be one of the principal psychological mechanisms by which music induces emotions.[6,38,93–95] In spite of this, there has been very little empirical research that robustly links quantitative measures of expectation with induced emotion, partly due to the previous lack of a reliable computational model capable of simulating listeners' musical expectations. Research has shown greater physiological arousal and subjective tension for Bach chorales manipulated to contain harmonic endings that violated principles of Western music theory[96] and also for extracts from romantic and classical piano sonatas.[97] However, as the stimulus categories were derived from music-theoretic analysis, this does not provide insight into the underlying cognitive processes, especially with respect to the SLH and the PPH.

Egermann et al. took continuous ratings of subjective emotion (arousal and valence) and physiological measures (skin conductance and heart rate) while participants listened to live performances of music for solo flute. IDyOM was used to obtain pitch IC profiles reflecting the unexpectedness of the pitch of each note in the stimuli.[82] The results showed that high-IC passages were associated with higher subjective and physiological arousal and lower valence than low-IC passages. This has been replicated in a controlled, laboratory-based behavioral study of continuous responses to folk song melodies selected to vary systematically in terms of pitch and rhythmic predictability (assessed using IDyOM IC).[83] The results showed that arousal was higher and valence lower for unpredictable compared with predictable
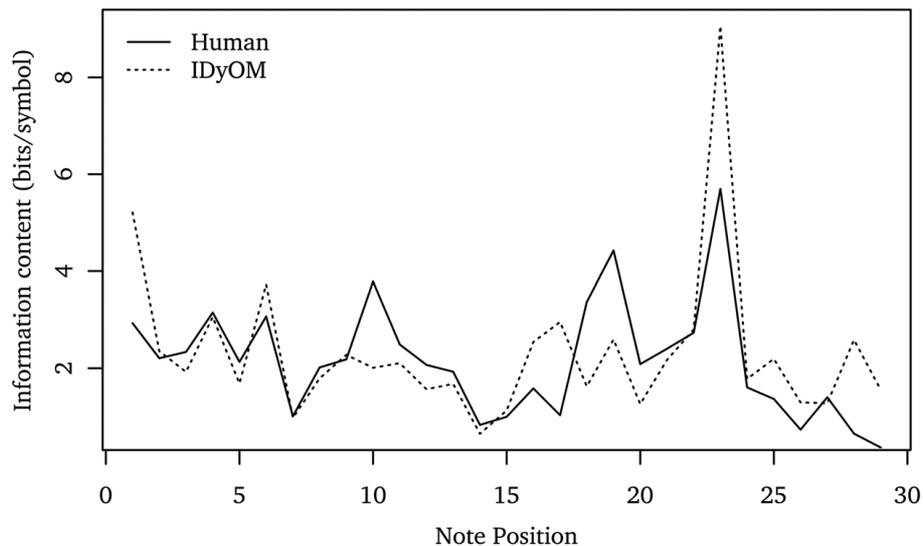
Figure 3. Information content generated by IDyOM for the Bach chorale shown in Figure 1, together with mean perceived expectedness from an empirical study reported by Manzara and colleagues.[84] In this study, 15 participants were given a capital sum of virtual currency $S_0 = 0$ and bet a proportion p of their capital on the pitch of each successive note in a melody (presented via a computer interface), continuing to place bets until the correct note was predicted, at which point they moved to the next note. At each note position n, incorrect predictions resulted in the loss of p, while the correct prediction was rewarded by incrementing the capital sum in proportion to the amount bet: $S_n = 20 p S_{n-1}$ (there were 20 pitches to choose from). The measure of information content plotted is derived by taking $\log_2 20 - \log_2 S$, where S is the capital won for a given note averaged across participants. As in Figure 2, IDyOM was configured to predict pitch with an attribute linking melodic pitch interval and chromatic scale degree (pi ⊗ sd, see Fig. 1) using both the short-term and long-term models, the latter trained on 903 folk songs and chorales (data sets 1, 2, and 9 in table 4.1 of Ref. 35 comprising 50,867 notes). IDyOM was configured to predict pitch only, since the participants in the Manzara *et al.* study were given the task of predicting pitch only.

melodies and that this effect was stronger for rhythmic predictability than pitch predictability. Furthermore, causal manipulations of the stimuli had the predicted effects on valence responses: transforming a melody to be more predictable resulted in increased valence ratings. Theoretical proposals of an inverted U-shaped relationship between predictability and pleasure[98] have received empirical support in some[99] but not all[100] studies of music perception. The results reviewed above show lower valence for more unpredictable musical passages, which may be because the particular combination of stimuli and participants reflects only the right-hand side of a putative underlying an inverted U-shaped relationship.

These results confirm the hypothesized role of probabilistic prediction in communicating musical affect, linking the predictability of musical events, assessed quantitatively in terms of IC, with the valence and arousal of listeners' continuous emotional responses. Gingras *et al.* report a study that examines the relationship between compositional structure, expressive performance timing, and per-

ceived tension in this communicative process.[8] IDyOM was used to characterize, in terms of IC and entropy, the compositional structure of the *Prélude non mesuré No. 7* by Louis Couperin, which was then performed by 12 professional harpsichordists whose performances were rated continuously for tension experienced by 50 listeners. IC and entropy were predictive of continuous changes in performance timing (performers slowed down in anticipation of high-IC events, and timing was more variable across performers around points of high IC and entropy), which, in turn, were predictive of perceived tension. Since the prelude is unmeasured, there is generous scope for expressive timing in performance, and, since the piece was performed on a harpsichord, performance expression is channeled primarily through timing, since there is little scope for expressive variations in dynamics and timbre. These design choices provide experimental control, but the results need to be generalized to a broader range of musical and instrumental styles.

It is important to note that expectation is not the only psychological mechanism by which

music can induce emotions,[6,93] and future research should examine the ways in which expectation-based induction of emotion interacts with other psychological mechanisms, such as imagery, contagion, and episodic memory, to generate complex aesthetic experiences of music.

Recognition memory

As noted above, IDyOM uses computational techniques originally developed for use in universal lossless data compression, where IC has a well-defined information-theoretic interpretation.[44,54] A sequence with low IC is predictable and thus does not need to be encoded in full, since the predictable portion can be reconstructed with an appropriate predictive model; the sequence is compressible and can be stored efficiently. Conversely, an unpredictable sequence with high IC is less compressible and requires more memory for storage. Therefore, there are theoretical grounds for using IDyOM as a model of musical memory. Empirical research has shown that more complex musical examples are more difficult to hold in memory for later recognition,[101–104] and this appears to be related to features that are stylistically unusual.[105] Furthermore, there is a strong link between information-theoretic measures of predictability and perceived complexity of musical structure.[106] Therefore, there are also empirical grounds for using IDyOM to simulate the relationship between stimulus predictability (as a measure of complexity) and memory for music.

Loui and Wessel used artificial auditory grammars to demonstrate that listeners show better recognition memory for previously experienced sequences generated by a grammar and that this generalizes to new exemplars from the grammar.[107] Furthermore, in an EEG study, generalization performance correlated with the amplitude of an early anterior negativity (FCz, 150–210 milliseconds).[89] However, this research did not explicitly relate degrees of predictability with memory performance. Agres *et al.* report a study that investigates recognition memory for artificial tone sequences varying systematically in information-theoretic complexity across three sessions in each of which listeners were presented with 12 sequences, followed by a recognition test consisting of the same 12 sequences and 12 foils.[108] To simulate listeners' responses, an IDyOM model with no prior training was exposed

to the stimulus set, learning the structure of the artificial style dynamically throughout the course of the session. In the first session, memory performance—measured by d' scores—did not correlate with the average IC of the stimuli. However, over time, listeners learned the structure of the artificial musical style to the extent that, by the third session, IC accounted for 85% of the variance in memory performance, such that memory was better for predictable stimuli (those with low IC).

This suggests a strong relationship between the stylistic unpredictability of the stimulus, again represented by IDyOM IC, and accuracy of encoding or retrieval in memory. However, these results need to be replicated with actual music varying systematically in stylistic predictability.

Perceptual similarity

Similarity perception is considered a fundamental process in cognitive science because it provides the psychological basis for classifying perceptual and cognitive phenomena into categories.[109] Recent theories view the process of comparing two perceptual stimuli as a process of transformation such that similarity emerges as the complexity of the simplest transformation between them.[110–112] This process can be simulated using information-theoretic models as the *compression distance* between the two stimuli.[56,113,114] Informally, IDyOM can be used to derive a compression distance $D(x, y)$ between two musical stimuli x and y by training a model on x, using that model to predict y, and taking the average IC across all notes in y (see Ref. 115 for a formal presentation of the model). If x and y are very similar, the IC will be low; if they are very dissimilar, the IC will be high.

Pearce and Müllensiefen tested this model by comparing compression distance with pairwise similarity ratings provided by listeners in three studies for stimuli consisting of one original pop melody and a manipulated version (containing rhythm, interval, contour, phrase order, and modulation errors).[115] The results showed very high correlations between compression distance and perceptual similarity (with coefficients ranging from 0.87 to 0.94), especially for IDyOM models configured to combine probabilistic predictions of pitch and timing.

To further assess generalization performance, IDyOM's measure of compression distance was tested on a very different set of data:[115] the MIREX

2005 similarity task designed to evaluate melodic similarity algorithms in music information retrieval research.[116,117] In this task, algorithms must rank the similarity of 558 candidate melodies to each of 11 queries (all taken from the RISM A/II catalog of incipits from music manuscripts dated from 1600 onward), and performance is assessed by comparison with a canonical order compiled from the responses of 35 musical experts. Without any prior optimization for this task, IDyOM performed comparably to the best-performing algorithms originally submitted (which took advantage of prior optimization on a comparable set of training data that is no longer available).

Phrase-boundary perception

The idea that perceptual grouping (or segment) boundaries occur at points of uncertainty or prediction error has been investigated in several areas of cognitive science, including modeling of phrase and word boundary perception in language.[118–120] Research has also demonstrated that children and adults learn the statistical structure of novel artificial auditory sequences, identifying sequential grouping boundaries on the basis of low transition probabilities.[13,121]

IDyOM has been used to test the hypothesis that perceived grouping boundaries in music (defining phrases) occur before contextually unpredictable events (those with high IC).[122] The principle is illustrated clearly in Figure 3, in which phrase boundaries (marked by fermata in the score shown in Fig. 1) are preceded by a fall in IC to the final note of a phrase, followed by a marked rise in IC for the first note of the subsequent phrase. IDyOM was configured to predict both pitch and timing of notes and used to identify points where IC increased markedly compared with the recent trend.[122] Comparing the boundaries predicted for 15 pop and folk songs with those indicated by 25 participants in an empirical study, IDyOM predicted perceived phrase boundaries with reasonable success. In most cases, performance was not as high as rule-based models,[12,123] though these have been optimized specifically for phrase-boundary detection based on expert knowledge and do not provide any account of enculturation or cross-cultural differences in boundary perception.[124] By contrast, IDyOM was not optimized in any way for boundary detection, and this research did not make full use of IDyOM's ability to simultaneously predict multiple attributes of musical events, leaving much scope for further development of IDyOM's phrase-boundary detection model. Simulating boundary perception at one level opens the door to simulating perception of hierarchical structure in music by inferring embedded groups at different hierarchical levels of abstraction[11] and using these as units in a multilayer predictive model.

Metrical inference

The IDyOM models used to predict phrase-boundary perception[122] and similarity perception[115] generate combined predictions of pitch and temporal position. In these models, the timing of notes is predicted using a model of statistical regularities in rhythm, but note timing is also influenced heavily by meter, a hierarchically embedded structure of periodically recurring accents that is inferred and aligned with a piece of music[9] and is also an important influence on temporal expectations. Palmer and Krumhansl[36] examined probe-tone ratings for events whose timing was varied systematically in relation to the meter implied by the preceding rhythmic context. Ratings reflected the hierarchical structure of the meter and the statistical distribution of onsets in music, leading to the suggestion that listeners' metrical expectations reflect learned temporal distributions.

Consistent with this proposal, cross-cultural differences in meter perception have been observed using a task in which listeners detect changes to rhythmic patterns that either preserve or violate metrical structure.[125,126] American adults show better detection in *isochronous* meters (e.g., 6/8) than *nonisochronous* meters (e.g., 7/8), while adults from Turkey and the Balkans (where such meters are common) show no such difference[125] but only for nonisochronous meters that appear in the culture.[127] American 6-month-olds show no such difference in processing of isochronous and non-isochronous meters; 12-month-olds do show a difference, but it is eliminated by 2 weeks of listening to Balkan music, while this was not the case for U.S. adults.[126] There is also evidence for cross-cultural differences in rhythm production as a function of enculturation.[128,129]

Can such enculturation effects be accurately simulated using computational models? As noted above, rule-based models of meter

perception[9,12,66–68] are not sensitive to experience and therefore cannot plausibly account for enculturation, while approaches that simulate meter perception as emerging from the resonance of coupled oscillators that entrain to temporal periodicities[71,73,130,131] naturally imply an explanation of meter in terms of stimulus structure rather than the experience of the listener.

Recent research has extended IDyOM with an empirical Bayesian scheme for inferring meter.[132] The metrical interpretation of a rhythm is treated as a hidden variable, consisting of both the metrical category itself (i.e., the time signature) and a phase aligning it to the rhythm. Metrical inference involves computing the posterior probability of a metrical interpretation at a given point in a rhythm through Bayesian combination of a prior distribution over meters (estimated empirically from a corpus) with the likelihood of an onset given the meter (estimated empirically by IDyOM). By virtue of IDyOM's statistical modeling framework, both the likelihood and the prior are also conditional on the preceding rhythmic context; therefore, metrical inference can vary dynamically event by event during online processing of music, taking into account the previous rhythmic context. Furthermore, the model naturally combines IDyOM's temporal predictions arising through repetition of rhythmic motifs with temporal predictions arising from the inferred meter. Unlike other probabilistic approaches, which are hand-crafted specifically for meter finding,[133,134] this approach derives metrical inference from a general-purpose model of sequential statistical learning and probabilistic prediction (implemented in IDyOM).

Computational simulations suggest that the model of metrical inference performs well. In a collection of 4966 German folk songs from the Essen Folk Song Collection,[135–137] it correctly predicted the notated time signature in 71% of the corpus, with performance increasing for higher order models (tested up to an order bound of four). Furthermore, and of greater theoretical interest, metrical inference substantially reduces IC (or prediction error) at all order bounds compared with a comparable IDyOM model of temporal prediction that does not perform metrical inference. This provides concrete, quantitative evidence that metrical inference is a profitable strategy for improving accuracy of temporal prediction in processing music. It is important to generalize these findings to musical styles exhibiting a greater range of meters (including nonisochronous meters), as well as styles exhibiting high levels of metrical uncertainty (e.g., through syncopation or polyrhythm), making metrical induction more challenging.

Statistical learning in musical enculturation

Most research on music cognition has been conducted on Western musical styles guided, implicitly or otherwise, by the particularities of Western music theory. However, the syntactic structure of musical styles varies among musical cultures. According to the SLH, this structure is learned through exposure producing observable differences among listeners from different musical cultures. Demorest and Morrison capture the effects of the SLH in their *cultural distance* hypothesis: "the degree to which the musics of any two cultures differ in the statistical patterns of pitch and rhythm will predict how well a person from one of the cultures can process the music of the other."[138] While cross-cultural research has found evidence of differences in music perception between listeners as a function of their culture,[40,41,64,65,124–129,139–146] the psychological mechanisms underlying the acquisition of these differences are currently poorly understood.

The research reviewed to this point demonstrates that exactly the same underlying model of probabilistic prediction provides a plausible account of a wide range of different psychological processes in music perception, including expectation, emotion, recognition memory, similarity perception, phrase-boundary perception, and metrical inference. In this research, the responses of Western listeners have been simulated using IDyOM models trained on Western tonal music (that approximates, within a tolerable degree of error, the stylistic properties of the music to which a typical Western listener is exposed). The IDyOM results reviewed above, therefore, are consistent with statistical learning as a mechanism for musical enculturation but the relationship is correlational rather than causal (with the exception of Ref. 108, which examined statistical learning directly but using an artificial musical system). In the following, I will outline a new modeling approach for a causal empirical investigation of the SLH of enculturation in musical styles.
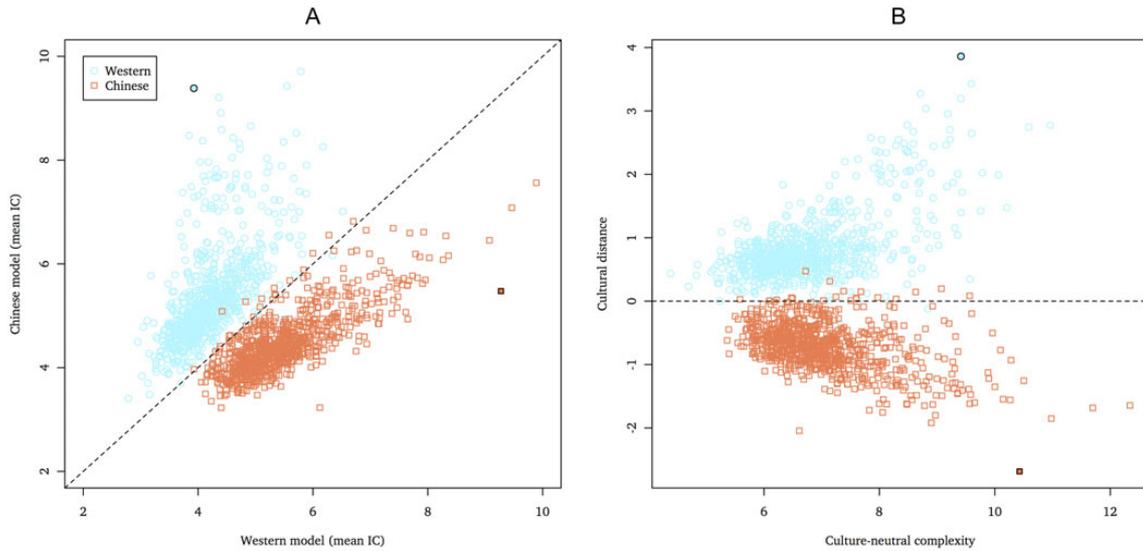
Figure 4. Simulating cultural distance between Western and Chinese listeners. (A) The information content of the Western model plotted against that of the Chinese model with the line of equality shown. (B) A 45° rotation of A such that the ordinate represents cultural distance and the abscissa culture-neutral complexity. For each style, the composition with the most extreme cultural distance is highlighted, and corresponding musical scores are shown for these two melodies. The Western corpus consists of 769 German folk songs from the Essen Folk Song Collection[135–137] (data sets *fink* and *erk*). The Chinese corpus consists of 858 Chinese folk songs from the Essen Folk Song Collection (data sets *han* and *natmin*). In a prior step, duplicate compositions were removed from the full data sets using a conservative procedure that considers two composition duplicates if they share the same opening for melodic pitch intervals, regardless of rhythm. IDyOM is configured to predict pitch with an attribute linking pitch interval with scale degree (pi ⊗ sd) and onset with the ioi-ratio attribute (Fig. 1) using the long-term model only trained on the Western and Chinese corpora, respectively, for the Western and Chinese models.

In order to test whether IDyOM is capable of simulating enculturation effects through statistical learning, IDyOM models were trained on corpora reflecting different musical cultures, simulating listeners from those cultures. A Western listener was simulated by training a model on a corpus of Western folk songs (the Western model) and a Chinese listener was simulated by training a model on a corpus of Chinese folk songs (the Chinese model). Each model was used to make both within-culture and between-culture predictions. For the within-culture predictions (i.e., the Western model processing Western folk songs or the Chinese model processing Chinese folk songs), IDyOM was used to estimate the IC

Table 1. IDyOM simulations of cultural distance between the Chinese and Western corpora (Fig. 4)

	Western example (deut1445)			Chinese example (han0418)			Overall	
	Western model IC	Chinese model IC	Cultural distance	Western model IC	Chinese model IC	Cultural distance	Accuracy	Cultural distance
Pitch	2.44	6.53	2.89	4.77	2.36	1.70	97.91	0.62
Onset	1.49	2.86	0.97	4.51	3.11	0.99	84.27	0.15
Pitch and onset	3.93	9.39	3.86	9.27	5.48	2.69	98.52	0.77

NOTE: Results are shown for IDyOM models configured to predict pitch only (using an attribute linking pitch interval with scale degree, pi ⊗ sd, see Fig. 1), onset only (using the attribute ioi-ratio), and both pitch and onset. Overall accuracy and cultural distance are shown as well as results for a Western and a Chinese piece with high cultural distance (Fig. 4) including the information content (IC) for the Western and Chinese models (trained on the Western and Chinese corpora, respectively) and cultural distance.

of every event in every composition in the corpus (using 10-fold cross-validation[147] to create training and test sets from the same corpus). For between-culture predictions, IDyOM was first trained on the within-culture corpus (e.g., the Western corpus for the Western model) and then used to estimate the IC of every note in every composition in the other corpus representing the comparison culture (e.g., the Chinese corpus for the Western model). IDyOM was configured to use only its LTM trained on the appropriate corpus; the short-term model was not used. In all cases, IC was averaged across notes, yielding a mean IC value representing the unpredictability of each composition for each model. The results are shown in Figure 4 and Table 1.

For the comparison between cultures (Western versus Chinese), the data are plotted in Figure 4 for each composition in the two corresponding corpora: IC for one model is plotted on the abscissa, while IC for the second model is plotted on the ordinate. The line of equality ($x = y$) indicates equivalence between the two models: compositions lying on this line are equally predictable for each model and do not distinguish the two cultures; in other words, they should be equally familiar and predictable to listeners enculturated in either musical style. Positions near the origin represent compositions that are predictable within both cultures, while positions far from the origin represent compositions that are unpredictable within both cultures. Positions farther away from the line of equality represent compositions that are predictable for the simulated model of one culture but unpredictable for the simulated model of the other culture. Distance from the line of equality, therefore, provides a quantitative measure of cultural distance[138] based on information-theoretic modeling of enculturation in

musical styles. Figure 4A illustrates how cultural distance is computed for a comparison between IDyOM models trained on Western and Chinese corpora and, by rotating the data points through 45°, Figure 4B shows the same data with cultural distance on the ordinate and culture-neutral complexity on the abscissa. In this example, IDyOM correctly classifies 98.52% of the folk songs by culture (Chinese versus Western). Moreover, classification accuracy and cultural distance are greater for IDyOM models configured to predict both pitch and time than models configured to predict pitch or time in isolation (Table 1), suggesting both that a combination of temporal and pitch regularities distinguishes the styles and that IDyOM is capable of learning such distinctive regularities in pitch and timing.

This approach provides a formal, computational model of enculturation, which guides the proposition of hypotheses about cultural familiarity and processing fluency. For example, referring to the examples shown in Figure 4, stimuli with strongly positive cultural distance should prove culturally familiar and easy to process for Western listeners but culturally unfamiliar and difficult to process for Chinese listeners and vice versa for stimuli with strongly negative cultural distance.

Van der Weij *et al.* developed empirical simulations of the effects of enculturation on metrical inference, using the computational model of metrical inference described above.[132] A Western model trained on 1136 German folk songs is compared with a Chinese model trained on 1136 Chinese folk songs (all stimuli taken from the Essen Folk Song Collection[135–137]). When tested on 200 unseen folk songs from each culture, the Western model shows greater IC (prediction error) for Chinese music

(1.72 bits per symbol) than for German music (1.34), while the Chinese model shows greater prediction error for the German music (1.70) than the Chinese music (1.49). Furthermore, the Western model also shows better meter-finding performance for German music (73% correct) than Chinese music (72%), while the Chinese model performs better on Chinese music (75%) than German music (47%).

These simulations demonstrate that IDyOM provides a plausible computational model of enculturation effects through statistical learning, though further empirical studies are required to fully corroborate SLH.

Conclusions

I have proposed two hypotheses about the psychological processes underlying enculturation in musical styles: (1) that probabilistic prediction is a foundational process in music perception underpinning other psychological processes (PPH) and (2) that statistical learning is the mechanism by which listeners acquire probabilistic models of musical styles (SLH). A review of the empirical evidence demonstrates that many different aspects of music perception—expectation, emotional response, recognition memory, phrase boundary perception, perceptual similarity, and, potentially, meter perception—can be simulated in terms of a single underlying process of probabilistic prediction, implemented in IDyOM. While these results are consistent with the SLH, since an IDyOM model trained on Western music accurately simulates Western listeners across a range of tasks, they do not provide causal evidence for the SLH. However, the results of a recognition memory study[108] show that memory performance is causally related to dynamic statistical learning of an artificial musical system. Finally, I presented data from computational simulations suggesting that statistical learning can plausibly predict causal effects of differential cultural exposure to musical styles on perception, providing a formal, quantitative model of cultural distance.[138]

Therefore, there are increasingly valid empirical and theoretical grounds to propose probabilistic prediction based on statistical learning as a foundational psychological process in a general theory of music perception. However, several areas remain open for future research. The results reviewed in this paper have been obtained for discrete, symbolic representations of melodic musical styles. To generalize the approach to a wider variety of musical styles, the representational capacity of IDyOM must be expanded to polyphonic music[61] but also to musical cultures that have no written tradition, where the distinction between composition and performance is blurred or nonexistent, or where music is inextricably combined with other modes of communication.[148,149] Doing so would open up the approach to a much broader range of musical cultures and traditions while also introducing significant computational challenges in modeling statistical learning and probabilistic prediction. It is also important to understand in more detail how musical training (active and explicit) and musical exposure (passive and implicit) exert a combined influence in musical enculturation.[150,151] Questions also arise over the effects of exposure to more than one musical style during enculturation. Further research is required to examine whether IDyOM's statistical learning mechanism distinguishes the styles sufficiently to account for such cases of bimusicalism or whether separate IDyOM models simulate bimusical listeners more accurately.[152,153] Finally, there is a fast-growing body of neuroscientific research on predictive processing in music,[80,88,89,91,92,96,154,155] and further progress in understanding the neural processes underlying the SLH and the PPH in musical enculturation will benefit significantly from closely coordinated combination of empirical neuroimaging with computational modeling of the underlying mechanisms as outlined in this paper.

Acknowledgments

Competing interests

The author declares no competing interests.

References

1. Meyer, L.B. 1989. *Style and Music: History, Theory and Ideology.* Chicago: University of Chicago Press.
2. Meyer, L.B. 1957. Meaning in music and information theory. *J. Aesthet. Art Crit.* **15:** 412–424.

3. Rohrmeier, M. & M.T. Pearce. 2017. Musical syntax I: theoretical perspectives. In *Springer Handbook of Systematic Musicology*. R. Bader, Ed.: 473–486. Berlin: Springer.

4. Pearce, M.T. & M. Rohrmeier. 2017. Musical syntax II: empirical perspectives. In *Springer Handbook of Systematic Musicology*. R. Bader, Ed.: 487–505. Berlin: Springer.

5. Kendall, R.A. & E.C. Carterette. 1990. The communication of musical expression. *Music Percept.* **8:** 129–163.

6. Juslin, P.N. & D. Västfjäll. 2008. Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* **31:** 559–621.

7. Juslin, P.N. 2000. Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform.* **26:** 1797–1813.

8. Gingras, B., M.T. Pearce, M. Goodchild, *et al.* 2015. Linking melodic expectation to expressive performance timing and perceived musical tension. *J. Exp. Psychol. Hum. Percept. Perform.* **42:** 594–609.

9. Lerdahl, F. & R. Jackendoff. 1983. *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

10. Narmour, E. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realisation Model.* Chicago: University of Chicago Press.

11. Narmour, E. 1992. *The Analysis and Cognition of Melodic Complexity: The Implication-Realisation Model.* Chicago: University of Chicago Press.

12. Temperley, D. 2001. *The Cognition of Basic Musical Structures.* Cambridge, MA: MIT Press.

13. Lany, J. & J.R. Saffran. 2013. Statistical learning mechanisms in infancy. In *Neural Circuit Development and Function in the Healthy and Diseased Brain.* J. Rubenstein & P. Rakic, Eds.: 231–248. Amsterdam: Elsevier.

14. Perruchet, P. & S. Pacton. 2006. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* **10:** 233–238.

15. Friston, K. 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11:** 127–138.

16. Bar, M., Ed. 2011. *Predictions in the Brain: Using Our Past to Generate a Future.* Oxford: Oxford University Press.

17. Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36:** 181–204.

18. DeLong, K.A., T.P. Urbach & M. Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* **8:** 1117–1121.

19. Hale, J. 2006. Uncertainty about the rest of the sentence. *Cogn. Sci.* **30:** 643–672.

20. Levy, R. 2008. Expectation-based syntactic comprehension. *Cognition* **16:** 1126–1177.

21. Cristià, A., G.L. McGuire, A. Seidl, *et al.* 2011. Effects of the distribution of acoustic cues on infants' perception of sibilants. *J. Phon.* **39:** 388–402.

22. Bar, M. 2007. The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* **11:** 280–289.

23. Bubic, A., D.Y. von Cramon & R.I. Schubotz. 2010. Prediction, cognition and the brain. *Front. Hum. Neurosci.* **4:** 25.

24. Summerfield, C. & T. Egner. 2016. Feature-based attention and expectation. *Trends Cogn. Sci.* **20:** 401–404.

25. Summerfield, C. & T. Egner. 2009. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13:** 403–409.

26. Wolpert, D.M. & J.R. Flanagan. 2001. Motor prediction. *Curr. Biol.* **11:** R729–R732.

27. Friston, K. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B* **360:** 815–836.

28. Huang, Y. & R.P.N. Rao. 2011. Predictive coding. *Wiley Interdiscip. Rev. Cogn. Sci.* **2:** 580–593.

29. Rao, R.P.N. & D.H. Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2:** 79–87.

30. Furl, N., S. Kumar, K. Alter, *et al.* 2011. Neural prediction of higher-order auditory sequence statistics. *Neuroimage* **54:** 2267–2277.

31. Kumar, S., W. Sedley, K.V. Nourski, *et al.* 2011. Predictive coding and pitch processing in the auditory cortex. *J. Cogn. Neurosci.* **23:** 3084–3094.

32. Blank, H. & M.H. Davis. 2016. Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biol.* **14:** e1002577.

33. Heilbron, M. & M. Chait. 2017. Great expectations: is there evidence for predictive coding in auditory cortex. *Neuroscience* https://doi.org/10.1016/j.neuroscience.2017.07.061.

34. Vuust, P., M. Witek, M. Dietz, *et al.* 2018. Now you hear it: a novel predictive coding model for understanding rhythmic incongruity. *Ann. N.Y. Acad. Sci.* https://doi.org/10.1111/nyas.13622.

35. Pearce, M.T. 2005. The construction and evaluation of statistical models of melodic structure in music perception and composition. Doctoral dissertation. Department of Computer Science, City University of London.

36. Palmer, C. & C.L. Krumhansl. 1990. Mental representations for musical meter. *J. Exp. Psychol. Hum. Percept. Perform.* **16:** 728–741.

37. Krumhansl, C.L. 1990. *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.

38. Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

39. Oram, N. & L.L. Cuddy. 1995. Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychol. Res.* **57:** 103–118.

40. Castellano, M.A., J.J. Bharucha & C.L. Krumhansl. 1984. Tonal hierarchies in the music of North India. *J. Exp. Psychol. Gen.* **113:** 394–412.

41. Kessler, E.J., C. Hansen & R.N. Shepard. 1984. Tonal schemata in the perception of music in Bali and the West. *Music Percept.* **2:** 131–165.

42. Krumhansl, C.L. 2000. Tonality induction: a statistical approach applied cross-culturally. *Music Percept.* **17:** 461–479.

43. Begleiter, R., R. El-Yaniv & G. Yona. 2004. On prediction using variable order Markov models. *J. Artif. Intell. Res.* **22:** 385–421.

44. Bell, T.C., J.G. Cleary & I.H. Witten. 1990. *Text Compression* Englewood Cliffs, NJ: Prentice Hall.

45. Cleary, J.G. & W.J. Teahan. 1997. Unbounded length contexts for PPM. *Comput. J.* **40:** 67–75.

46. Bunton, S. 1997. Semantically motivated improvements for PPM variants. *Comput. J.* **40:** 76–93.

47. Shepard, R.N. 1982. Structural representations of musical pitch. In *Psychology of Music*. D. Deutsch, Ed.: 343–390. New York: Academic Press.

48. Levitin, D.J. & A.K. Tirovalas. 2009. Current advances in the cognitive neuroscience of music. *Ann. N.Y. Acad. Sci.* **1156:** 211–231.

49. Teki, S., M. Grube, S. Kumar, *et al.* 2011. Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* **31:** 3805–3812.

50. Conklin, D. & I.H. Witten. 1995. Multiple viewpoint systems for music prediction. *J. New Music Res.* **24:** 51–73.

51. Palmer, C. & C.L. Krumhansl. 1987. Independent temporal and pitch structures in determination of musical phrases. *J. Exp. Psychol. Hum. Percept. Perform.* **13:** 116–126.

52. Jones, M.R. & M.G. Boltz. 1989. Dynamic attending and responses to time. *Psychol. Rev.* **96:** 459–491.

53. Boltz, M.G. 1999. The processing of melodic and temporal information: independent or unified dimensions? *J. New Music Res.* **28:** 67–79.

54. MacKay, D.J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

55. Shannon, C.E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* **27:** 379–656.

56. Chater, N. & P. Vitányi. 2003. Simplicity: a unifying principle in cognitive science? *Trends Cogn. Sci.* **7:** 19–22.

57. Attneave, F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* **61:** 183–193.

58. Barlow, H.B. 1961. Possible principles underlying the transformation of sensory messages. In *Sensory Communication*. W. Rosenblith, Ed.: 217–234. Cambridge MA: MIT Press.

59. Meyer, L.B. 1973. *Explaining Music: Essays and Explorations*. Berkeley, CA: University of California Press.

60. Harrison, P.M.C. & M.T. Pearce. 2017. A statistical-learning model of harmony perception. In *Proceedings of DMRN+12: Digital Music Research Network One-Day Workshop*. P. Kudamakis & M. Sandler, Eds.: 15. London: Queen Mary University of London.

61. Sears, D.R.W., M.T. Pearce, W.E. Caplin, *et al.* 2018. Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *J. New Music Res.* **47:** 29–52.

62. Sauvé, S. 2018. Prediction in polyphony: modelling musical auditory scene analysis. Doctoral dissertation. School of Electronic Engineering and Computer Science, Queen Mary University of London.

63. Schellenberg, E.G. 1997. Simplifying the implication-realisation model of melodic expectancy. *Music Percept.* **14:** 295–318.

64. Krumhansl, C.L., P. Toivanen, T. Eerola, *et al.* 2000. Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* **76:** 13–58.

65. Krumhansl, C.L., J. Louhivuori, P. Toiviainen, *et al.* 1999. Melodic expectation in Finnish spiritual hymns: convergence of statistical, behavioural and computational approaches. *Music Percept.* **17:** 151–195.

66. Lee, C. 1991. The perception of metrical structure: experimental evidence and a model. In *Representing Musical Structure*. P. Howell, R. West & I. Cross, Eds.: 59–127. London: Academic Press.

67. Povel, D.-J.J. & P.J. Essens. 1985. Perception of temporal patterns. *Perception* **2:** 411–440.

68. Parncutt, R. 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* **11:** 409–464.

69. Lerdahl, F. 1988. Tonal pitch space. *Music Percept.* **5:** 315–350.

70. Temperley, D. & D. Sleator. 1999. Modeling meter and harmony: a preference-rule approach. *Comput. Music J.* **23:** 10–27.

71. Large, E.W., J.A. Herrera & M.J. Velasco. 2015. Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* **9:** 159.

72. Large, E.W. & M.R. Jones. 1999. The dynamics of attending: how we track time-varying events. *Psychol. Rev.* **106:** 119–159.

73. Large, E.W. & J.F. Kolen. 1994. Resonance and the perception of musical meter. *Conn. Sci.* **6:** 177–208.

74. Large, E.W. & C. Palmer. 2002. Perceiving temporal regularity in music. *Cogn. Sci.* **26:** 1–37.

75. Large, E.W., J.C. Kim, J.J. Barucha, *et al.* 2016. A neurodynamic account of music tonality. *Music Percept.* **33:** 319–331.

76. Marr, D. 1982. *Vision*. San Francisco, CA: W. H. Freeman.

77. Manning, C.D. & H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

78. Chen, S.F. & J. Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*. **13:** 310–318.

79. Hansen, N.C. & M.T. Pearce. 2014. Predictive uncertainty in auditory sequence processing. *Front. Psychol.* **5:** 1–17.

80. Pearce, M.T., M.H. Ruiz, S. Kapasi, *et al.* 2010. Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *Neuroimage* **50:** 302–313.

81. Omigie, D., M.T. Pearce & L. Stewart. 2012. Tracking of pitch probabilities in congenital amusia. *Neuropsychologia* **50:** 1483–1493.

82. Egermann, H., M.T. Pearce, G.A. Wiggins, *et al.* 2013. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cogn. Affect. Behav. Neurosci.* **13:** 533–553.

83. Sauvé, S., A. Sayad, R.T. Dean, *et al.* 2018. Effects of pitch and timing expectancy on musical emotion. *Psychomusicology* https://arxiv.org/ftp/arxiv/papers/1708/1708.03687.pdf.

84. Manzara, L.C., I.H. Witten & M. James. 1992. On the entropy of music: an experiment with Bach chorale melodies. *Leonardo* **2:** 81–88.

85. Besson, M. & F. Faïta. 1995. An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* **21:** 1278–1296.

86. Paller, K., G. McCarthy & C. Wood. 1992. Event-related potentials elicited by deviant endings to melodies. *Psychophysiology* **29:** 202–206.

87. Verleger, R. 1990. P3-evoking wrong notes: unexpected, awaited or arousing? *Int. J. Neurosci.* **55:** 171–179.

88. Koelsch, S. & S. Jentschke. 2010. Differences in electric brain responses to melodies and chords. *J. Cogn. Neurosci.* **22:** 2251–2262.

89. Loui, P., E.H. Wu, D.L. Wessel, *et al.* 2009. A generalized mechanism for perception of pitch patterns. *J. Neurosci.* **29:** 454–459.

90. Besson, M. & F. Macar. 1987. An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology* **24:** 14–25.

91. Omigie, D., M.T. Pearce, V.J. Williamson, *et al.* 2013. Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia* **51:** 1749–1762.

92. Vuust, P., L. Ostergaard, K.J. Pallesen, *et al.* 2009. Predictive coding of music—brain responses to rhythmic incongruity. *Cortex* **45:** 80–92.

93. Juslin, P.N. 2013. From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* **10:** 235–266.

94. Hanslick, E. 1854. *On the Musically Beautiful.* G. Payzant, Ed.: Indianapolis, IN: Hackett Publishing Company.

95. Meyer, L.B. 1956. *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

96. Steinbeis, N., S. Koelsch & J.A. Sloboda. 2006. The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological and neural responses. *J. Cogn. Neurosci.* **18:** 1380–1393.

97. Koelsch, S., S. Kilches, N. Steinbeis, *et al.* 2008. Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS One* **3:** e2631.

98. Berlyne, D.E. 1974. The new experimental aesthetics. In *Studies in the New Experimental Aesthetics: Steps Towards an Objective Psychology of Aesthetic Appreciation.* D.E. Berlyne, Ed.: 1–25. Washington, DC: Hemisphere Publishing Co.

99. Witek, M.A.G., E.F. Clarke, M. Wallentin, *et al.* 2014. Syncopation, body-movement and pleasure in groove music. *PLoS One* **9:** e94446.

100. Orr, M.G. & S. Ohlsson. 2005. Relationship between complexity and liking as a function of expertise. *Music Percept.* **22:** 583–611.

101. Cohen, A.J., L.L. Cuddy & D.J.K. Mewhort. 1977. Recognition of transposed tone sequences. *J. Acoust. Soc. Am.* **61:** 87–88.

102. Bartlett, J.C., A.R. Halpern & W.J. Dowling. 1995. Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Mem. Cognit.* **23:** 531–546.

103. Bartlett, J.C. & W.J. Dowling. 1980. Recognition of transposed melodies: a key-distance effect in developmental perspective. *J. Exp. Psychol. Hum. Percept. Perform.* **6:** 501–515.

104. Cuddy, L.L. & H.I. Lyons. 1981. Musical pattern recognition: a comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology* **1:** 15–33.

105. Müllensiefen, D. & A.R. Halpern. 2014. The role of features and context in recognition of novel melodies. *Music Percept.* **4:** 695–701.

106. Eerola, T. 2016. Expectancy-violation and information-theoretic models of melodic complexity. *Empir. Musicol. Rev.* **11:** 2–17.

107. Loui, P. & D. Wessel. 2008. Learning and liking an artificial musical system: effects of set size and repeated exposure. *Music Sci.* **12:** 207–230.

108. Agres, K., S. Abdallah & M.T. Pearce. 2018. Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cogn. Sci.* **42:** 43–76.

109. Shepard, R.N. 1986. Toward a universal law of generalization for psychological science. *Science* **237:** 1317–1323.

110. Hodgetts, C.J., U. Hahn & N. Chater. 2009. Transformation and alignment in similarity. *Cognition* **113:** 62–79.

111. Hahn, U. & N. Chater. 1998. Understanding similarity: a joint project for psychology, case-based reasoning, and law. *Artif. Intell. Rev.* **12:** 393–427.

112. Hahn, U., N. Chater & L.B. Richardson. 2003. Similarity as transformation. *Cognition* **87:** 1–32.

113. Chater, N. & P.M.B. Vitányi. 2001. The generalized universal law of generalization. *J. Math. Psychol.* **47:** 346–369.

114. Li, M., X. Chen, X. Li, *et al.* 2004. The similarity metric. *IEEE Trans. Inf. Theory* **50:** 3250–3264.

115. Pearce, M. & D. Müllensiefen. 2017. Compression-based modelling of musical similarity perception. *J. New Music Res.* **46:** 135–155.

116. Typke, R., M. Den Hoed, J. De Nooijer, *et al.* 2005. A ground truth for half a million musical incipits. *J. Digit. Inf. Manag.* **3:** 34–39.

117. Typke, R., F. Wiering & R. Veltkamp. 2005. Evaluating the earth mover's distance for measuring symbolic melodic similarity. In *Paper Presented at the Annual Music Information Retrieval Evaluation Exchange (MIREX) as Part of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, Queen Mary University of London.

118. Brent, M.R. 1999. Speech segmentation and word discovery: a computational perspective. *Trends Cogn. Sci.* **3:** 294–301.

119. Elman, J.L. 1990. Finding structure in time. *Cogn. Sci.* **14:** 179–211.

120. Cohen, P.R., N. Adams & B. Heeringa. 2007. Voting experts: an unsupervised algorithm for segmenting sequences. *Intell. Data Anal.* **11:** 607–625.

121. Saffran, J.R., E.K. Johnson, R.N. Aslin, *et al.* 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* **70:** 27–52.

122. Pearce, M.T., D. Müllensiefen & G.A. Wiggins. 2010. The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. *Perception* **39:** 1367–1391.

123. Cambouropoulos, E. 2001. The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference*, San Francisco, ICMA: 17–22.

124. Ayari, M. & S. McAdams. 2003. Aural analysis of Arabic improvised instrumental music (Taqsim). *Music Percept.* **21:** 159–216.

125. Hannon, E.E. & S.E. Trehub. 2005. Metrical categories in infancy and adulthood. *Psychol. Sci.* **16:** 48–55.

126. Hannon, E.E. & S.E. Trehub. 2005. Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl. Acad. Sci. USA* **102:** 12639–12643.

127. Hannon, E.E., G. Soley & S. Ullal. 2012. Familiarity overrides complexity in rhythm perception: a cross-cultural comparison of American and Turkish listeners. *J. Exp. Psychol. Hum. Percept. Perform.* **38:** 543–548.

128. Cameron, D.J., J. Bentley & J.A. Grahn. 2015. Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. *Front. Psychol.* **6:** 1–11.

129. Polak, R., J. London & N. Jacoby. 2016. Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: a corpus study of malian jembe drumming. *Front. Neurosci.* **10:** 1–11.

130. Mcauley, J.D. 1995. Perception of time as phase: toward an adaptive oscillator model of rhythmic pattern processing. PhD dissertation. University of Indiana, Bloomington, USA.

131. Large, E.W. & M.R. Jones. 1999. The dynamic of attending: how people track time-varying events. *Psychol. Rev.* **106:** 119–159.

132. Van der Weij, B., M.T. Pearce & H. Honing. 2017. A probabilistic model of meter perception: simulating enculturation. *Front. Psychol.* **8:** 824.

133. Temperley, D. 2007. *Music and Probability.* Cambridge, MA: MIT Press.

134. Temperley, D. 2009. A unified probabilistic model for polyphonic music analysis. *J. New Music Res.* **38:** 3–18.

135. Schaffrath, H. 1995. The Essen folksong collection. In *Database Containing 6,255 Folksong Transcriptions in the Kern Format and a 34-Page Research Guide [Computer Database].* D. Huron, Ed. Menlo Park, CA: CCARH.

136. Schaffrath, H. 1994. The ESAC electronic songbooks. *Comput. Musicol.* **9:** 78.

137. Schaffrath, H. 1992. The ESAC databases and MAPPET software. *Comput. Musicol.* **8:** 66.

138. Demorest, S. & S.J. Morrison. 2016. Quantifying culture: the cultural distance hypothesis of melodic expectancy. In *The Oxford Handbook of Cultural Neuroscience.* J.Y. Chiao, S.-C. Li, R. Seligman, *et al.*, Eds.: 183–194. Oxford, UK: Oxford University Press.

139. Demorest, S.M., S.J. Morrison, D. Jungbluth, *et al.* 2008. Lost in translation: an enculturation effect in music memory performance. *Music Percept.* **25:** 213–233.

140. Demorest, S.M., S.J. Morrison, V.Q. Nguyen, *et al.* 2016. The influence of contextual cues on cultural bias in music memory. *Music Percept.* **33:** 590–600.

141. Morrison, S.J., S.M. Demorest & L.A. Stambaugh. 2008. Enculturation effects in music cognition: the role of age and music complexity. *J. Res. Music Educ.* **56:** 118–129.

142. Demorest, S.M., S.J. Morrison, L.A. Stambaugh, *et al.* 2010. An fMRI investigation of the cultural specificity of music memory. *Soc. Cogn. Affect. Neurosci.* **5:** 282–291.

143. Nan, Y., T.R. Knösche & A.D. Friederici. 2009. Nonmusicians' perception of phrase boundaries in music: a cross-cultural ERP study. *Biol. Psychol.* **82:** 70–81.

144. Nan, Y., T.R. Knösche & A.D. Friederici. 2006. The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res.* **1094:** 179–191.

145. Hannon, E.E., C.M. Vanden Bosch Der Nederlanden & P. Tichko. 2012. Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Ann. N.Y. Acad. Sci.* **1252:** 92–99.

146. Soley, G. & E.E. Hannon. 2010. Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Dev. Psychol.* **46:** 286–292.

147. Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* San Mateo, CA, Morgan Kaufmann: 1137–1145.

148. Cross, I. 2012. Cognitive science and the cultural nature of music. *Top. Cogn. Sci.* **4:** 668–677.

149. Cross, I. 2014. Music and communication in music psychology. *Psychol. Music* **42:** 809–819.

150. Hansen, N.C., P. Vuust & M.T. Pearce. 2016. "If You Have to Ask, You'll Never Know": effects of specialised stylistic expertise on predictive processing of music. *PLoS One* **11:** e0163584.

151. Hannon, E.E. & L.J. Trainor. 2007. Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* **11:** 466–472.

152. Wong, P.C.M., A.H.D. Chan & E.H. Margulis. 2012. Effects of mono- and bicultural experiences on auditory perception. *Ann. N.Y. Acad. Sci.* **1252:** 158–162.

153. Wong, P.C.M., A.H.D. Chan, A. Roy, *et al.* 2011. The bimusical brain is not two monomusical brains in one: evidence from musical affective processing. *J. Cogn. Neurosci.* **23:** 4082–4089.

154. Miranda, R.A. & M.T. Ullman. 2007. Double dissociation between rules and memory in music: an event-related potential study. *Neuroimage* **38:** 331–345.

155. Leino, S., E. Brattico, M. Tervaniemi, *et al.* 2005. Representation of harmony rules in the human brain: further evidence from event-related potentials. *Brain Res.* **1142:** 169–177.