AMERICAN PSYCHOLOGICAL ASSOCIATION DIVISION 10 Psychology of Aesthetics, Creativity, and the Arts

ISSN: 1931-3896

https://doi.org/10.1037/aca0000811

Relationships Between Surprise, Liking, and Error Perception in Musical Listening

Amira Korkor[1], Valdas Noreika[1], Caroline Di Bernardi Luft[2], and Marcus Pearce[3]
[1] School of Biological and Behavioural Sciences, Queen Mary University of London
[2] Department of Psychology, Brunel University London
[3] School of Electronic Engineering and Computer Science, Queen Mary University of London

A large part of musical enjoyment stems from the interplay between predictability and surprise that evolves throughout a melody, and yet more extreme violation of musical predictions results in perception of an error. The aim of the present research was to investigate the relationship between liking and predictability in music and establish whether a relationship exists between the degree of unpredictability of a pitch and perception of an error. Moreover, we investigated whether certain individual differences between participants, or the musical style of the stimuli, affect these relationships between predictability and liking, surprise, or error perception. In the series of three experiments, participants were evaluated for musical background, personality traits, and creativity. They were then presented with classical or jazz melodies comprising varying degrees of predictability and reported liking, surprise, and error perception after each melody. We manipulated the predictability of musical notes using a computational model as a way of introducing gradated unpredictability. The results showed that participants had a strong preference for the most predictable melodies. Very unpredictable melodies, as identified by the model, were more often perceived as containing errors, but error perception was more forgiving in jazz compared to classical melodies. There was no evidence that individual differences such as creativity and openness to experience had any association with liking, surprise, or error perception. These results suggest predictability is a strong factor in both liking and error perception in musical listening.

Keywords: music perception, unpredictability, surprise, liking, error perception

Listening to music can bring an immense amount of pleasure, resulting in people returning to familiar pieces to hear them again and again. An especially pleasurable aspect of musical experience is innovation on familiar music, such as an improvisation, variation on a theme, or an acoustic cover, which necessarily involves the introduction of unpredictable events. On the other hand, a single highly unexpected note (or "false" note in a musical context) can trigger an almost instantaneous reaction from a listener. Previous research has highlighted the importance of predictions based on statistical learning in music perception (Huron, 2006; Pearce, 2018) and shown that the pleasure of a listening experience is related to the predictability of the musical events (V. K. M. Cheung et al., 2019; Gold et al., 2019; Huron, 2006). However, these findings have not been explicitly related to error perception.

Prediction

Anticipatory psychological and neural processes such as prediction or expectation are thought to be fundamental to both perception and action (Clark, 2013; Friston, 2010). Predictive coding theory posits a hierarchical predictive system that learns the structure of the environment and generates top-down predictions that disambiguate sensory input. A discrepancy between predicted and actual input is known as a prediction error, which is generally thought to vary continuously. Prediction errors allow for efficient propagation of input through the hierarchy and provide a loss function for learning. Prediction is also used to guide action toward sensory objects, and these actions can, in turn, disambiguate sensory input. In music perception, expectations are thought to play an important role both in perception and in appreciation of music (Huron, 2006; Pearce, 2018). In particular, research has shown that perceptual expectations during musical listening are well simulated by information-theoretic measures of unpredictability such as information content (IC; V. Cheung et al., 2023; Hansen & Pearce, 2014; Hansen et al., 2016; Sears et al., 2019), which also predicts perception of complexity (Clemente et al., 2020; Sauvé & Pearce, 2019).

In the present context, we use IC (discussed in more detail below) as an indicator of the unpredictability of musical events. IC is compared to subjective ratings of surprise to assess perception of unpredictability, such as one might experience when hearing unexpected notes in a musical improvisation. Ultimately, we are interested in how this measure of unpredictability influences liking and error perception.

Liking

The relationship between liking and prediction has been the subject of debate. One of the most longstanding theories proposes an inverted U-shaped relationship (or Wundt curve) between hedonic response

Amy M. Belfi served as action editor.
Amira Korkor https://orcid.org/0000-0002-6313-6154
Correspondence concerning this article should be addressed to Amira Korkor, School of Biological and Behavioural Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom. Email: a.korkor@qmul.ac.uk

(liking) and perceived complexity (Berlyne, 1960). In other words, the theory asserts a general tendency in human aesthetic preference to prefer an optimal intermediate level of arousal (neither too arousing/stimulating nor too boring), associated with an intermediate level of a collative variable such as complexity, familiarity, uncertainty, interestingness, or ambiguity. Lower or higher levels result in reduced liking. A meta-analysis by Chmiel and Schubert (2017) found that out of 57 studies that measured liking for varying degrees of familiarity or complexity, 15 showed strong support in replicating the full inverted U, while 35 were compatible in showing increasing or decreasing linear relationships between liking and complexity/familiarity, thereby replicating one or other half of the full inverted U.

Evidence for the theory has focused on either familiarity or perceived complexity, with experimental studies using an objective measure of stimulus complexity less common (Crozier, 1974; McMullen & Arnold, 1976; Vitz, 1966). Recent experimental investigations of this theory have used IC as a measure of complexity, finding support for the Wundt curve (V. K. M. Cheung et al., 2019; Gold et al., 2019). For example, Gold et al. (2019) showed that participants exhibited a consistent preference for melodies with intermediate unpredictability but also found that this was modulated by predictive uncertainty (entropy) such that when the entropy of a melody was higher, meaning that the musical context was more uncertain, participants gave a higher liking rating to melodies with more predictable notes, and vice versa. Similar results were observed for chord sequences by V. K. M. Cheung et al. (2019).

In the present study, we collect subjective liking ratings for musical stimuli that are related to IC as a measure of the collative variable unpredictability.

Error Perception

Although the evidence reviewed above suggests that moderate degrees of unpredictability are experienced as pleasurable, highly unpredictable musical events are likely to be perceived as errors. Major violations of musical predictions result in perception of an error (Halpern et al., 2017; Huron, 2006). Huron (2006) proposed a theoretical framework to understand the different ways in which expectations can be violated when an unexpected note is perceived. Schematic expectations are violated when a note diverges from the stylistic norms or conventions of a piece of music, veridical expectations are violated when a previously known melody diverges from its expected path, and dynamic expectations are violated when a note diverges from short-term statistical patterns that a listener updates in real time while listening to a piece of music. A low probability note can become acceptable to a listener through manipulation of dynamic expectations by an improvising musician (Huron, 2006) or even simply repeated exposure to the same error, for example, in a recording (Sheldon, 2004). However, in many situations such as the performance of a well-known piece, a violation of a veridical expectation cannot be recovered. It remains unknown if there is a statistical threshold in the unpredictability of a musical note that determines whether or not it will be perceived as an error.

Repetition

When a listener recognizes a piece of music as having been heard before, veridical memory mechanisms are invoked as the present stimulus is compared to a memory representation of its previous occurrence(s) (Crowder, 1993). Research comparing experimental measures of schematic and veridical expectations has generally found that schematic expectations override veridical expectations (Justus & Bharucha, 2001; Tillmann & Bigand, 2010). On the other hand, musical stimuli that have been heard before are generally liked more than otherwise equivalent novel stimuli (Ara & Marco-Pallarés, 2021; Peretz et al., 1998). However, the fine-grained timescale and number of repetitions required are not clear, and the effects of repetition on liking have not been combined with measures of unpredictability and error perception.

Individual Differences

An aspect of musical prediction that is underrepresented in the literature is the impact of the personality and creative qualities of the listener. Research has shown that individuals who were better at learning a new music grammar displayed higher musical creativity using this grammar, suggesting that musical creativity relies on learning (Zioga et al., 2019, 2024). However, it could be that the individuals who learned better were more open to new experiences and more capable of overlooking previously held musical expectations from their own musical culture when learning a new style. Openness to experience is understood to be related to creativity (Mccrae, 1987; Puryear et al., 2017), which means that it may be an important factor in understanding the way that creativity interacts with perception of unpredictable musical events. Furthermore, openness to experience is related (albeit weakly) to a greater preference for musical styles categorized as sophisticated or intense (Schäfer & Mehlhorn, 2017; Vella & Mills, 2017).

Another potentially relevant dimension is musical expertise. Research on the effects of musical expertise on error perception is mixed. Studies have found musical expertise to have no effect on the identification of errors in recordings (Brand & Burnsed, 1981), whereas musical experience has also been found to increase participants' ability to classify low-probability notes as violations of specific musical rules (Besson & Faïta, 1995). Given evidence that musicians use their knowledge to generate sharper, more focused expectations (i.e., greater discrimination between expected and unexpected events), such that they find unpredictable notes more unexpected than nonmusicians (Hansen & Pearce, 2014; Hansen et al., 2016), we hypothesize that greater musical sophistication would result in a lower threshold for error perception.

Finally, we included a measure of musical hedonia, the Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013), as an indicator of the tendency to take pleasure in listening to music. Although often used to detect musically anhedonic individuals, the BMRQ has been used in nonmusically anhedonic individuals for whom it correlates with the tendency to immerse oneself or become absorbed in music (Cardona et al., 2022). Given this, we reasoned that individuals with a greater BMRQ score might be more sensitive to unpredictable notes that interrupted their listening (i.e., state of absorption), whereas individuals with lower BMRQ might be less sensitive to such interruptions.

Musical Style

As well as varying between individuals, the relationships between unpredictability, liking, and error perception may be affected by musical style. In other words, it may be that listeners have different conceptions of the degree of unpredictability that is appropriate in different styles, which could influence both their liking and their error perception. To our knowledge, this question has not been investigated directly. However, there is good evidence for framing effects, whereby contextual information affects perceptual interpretation and liking for visual and auditory stimuli. For example, it has been shown that liking for music is affected by whether it is introduced as human or computer-generated (Shank et al., 2023) or the desirability of the alternative option if a choice is given (Lopez-Persem et al., 2016), while similar effects have been observed for visual art (Kirk et al., 2009). Here, we compare responses to stimuli in the style of classical music (Experiments 1 and 2) with stimuli in the style of jazz (Experiment 3), hypothesizing that when listening to jazz music, where improvisatory variation is a regular characteristic of the musical style, participants would be more open to unpredictability, showing greater liking and a higher error perception threshold for unpredictable notes.

Experiments

In the following experiments, we build on existing research to attempt to establish a more detailed understanding of responses to unpredictability in music through a triangulation of its effects on liking, surprise, and error perception. We use information-theoretic unpredictability as an operational definition of the conformity of a note to the stylistic norms and conventions of Western tonal music. Specifically, a computational model known as Information Dynamics of Music (IDyOM; Conklin & Witten, 1995; Pearce, 2005, 2018) is used to characterize the unpredictability of each note in a melody based on its learning of the statistical structure of a corpus of training melodies (the long-term model [LTM]) and repeated structure learned dynamically within a given piece of music (the short-term model [STM]). IDyOM is a variable-order Markov model that computes the unpredictability of musical events in terms of IC, the negative log probability of the event conditional on the immediately preceding musical context, where the conditional probabilities are derived from the model's training.

For the present experiments, IDyOM's LTM was trained on a corpus of Western tonal music, consisting of 903 melodies including Nova Scotian folk songs (Creighton, 1966), German folk songs (Essen folk song collection, Schaffrath, 1995), and Bach chorales (Riemenschneider, 1941). This combination of melodies has been used in previous applications of IDyOM to construct a statistical model that has proved capable of simulating perception of listeners with Western musical enculturation (see Pearce, 2018, for a review) and was used here for consistency. Although two musical styles are used (classical in Experiments 1 and 2; jazz in Experiment 3), the listeners were sampled in exactly the same way across experiments without consideration of musical experience, so we used the same training set to provide a consistent and comparable simulation across experiments.

In the experiments, we introduce a paradigm in which participants are presented with variations of a set of melodies with one note (the target note) altered in pitch such that the set of variations reflects a gradated scale of pitch predictability for the target note. We refer to this as the progressive digression paradigm as it allows us to analyze progressive digression from the original melody in terms of increasing unpredictability of the target note. We use this paradigm to observe shifts in liking and surprise with increasing unpredictability, as well as establish the threshold of unpredictability at which a listener perceives the novel note as an error.

In all experiments, IC from IDyOM reflects both schematic (via the LTM) and dynamic (via the STM) influences on unpredictability. In Experiment 1, short classical melodies are presented to participants in an exposure phase to establish veridical memories in the subsequent experimental phase. In Experiment 2, half of the stimuli have prior exposure, as in Experiment 1, and half do not, to test the effects of veridical memory. Orthogonally, whereas in Experiment 1 the target note was always the most predictable option according to IDyOM, in Experiment 2, this is compared with a condition in which the target note is the second most predictable option, to assess a potential effect of the composer's original choice of note. Finally, Experiment 3 applies the same paradigm to bebop jazz melodies, some of which are repeated (either immediately or with delay) to examine whether perception of unpredictability and its relationship with liking and error perception are affected by musical style and repetition at a shorter timescale than in Experiment 1.

Experiment 1

This experiment was designed to familiarize participants with a set of melodies and then, by altering those melodies, examine how liking, surprise, and error perception are impacted by changing the pitch of a note, varying its predictability. We hypothesized that liking of the melodies would show an inverted U-shaped relationship with unpredictability, initially increasing to a peak and then decreasing with increasing unpredictability. With respect to the context dependency of musical preference (as observed in Gold et al., 2019), we predicted that our relatively simple stimuli should promote liking of more adventurous or unexpected melodic variations. We expected that participants who were more creative and open to experiences would be more likely to favor surprising pitches. Our hypothesis regarding error perception was that the relationship between unpredictability and error perception would be described by a logistic curve, meaning that we expected several possible note variants to sound acceptable, and that participants would stop liking the pitch variants and begin to perceive them as errors relatively quickly once they passed a certain threshold. Furthermore, we predicted that the midpoint of the logistic error perception curve would vary between individuals in a way that would correlate with participant characteristics, such that higher openness to experience and creativity would be associated with higher thresholds for error perception, while greater musical sophistication and BMRQ would be associated with lower thresholds for error perception.

Method

Participants

Data were collected from 220 participants who were recruited via Prolific (https://www.prolific.co). Only 182 of the participants (104 female, aged 18–65 years, $M = 33.08$, $SD = 11.44$) were retained in accordance with the exclusion criteria described below. The sample size was based on that found in the literature for similar experiment types, which have employed sample sizes of 44–48 participants (e.g., Clemente et al., 2021, 2022; Gold et al., 2019). Following recommendations for online studies (Sauter et al., 2020; Stewart et al., 2017), we doubled this sample size as our minimum for all experiments. A power analysis for a small effect size of 0.2 in a linear regression model with 2 *df* yields a sample size of 48 to achieve 80% power at an α level of .05. All participants were U.K. residents at the time of participation. All data were collected using the Gorilla Experiment Builder platform (https://www.gorilla.sc). Ethical approval for all experiments was given by the Joint Research Management Office for Barts Health National Health Service Trust and Queen Mary University of London. Participants provided informed consent and were paid at a rate of £7/hr for participation in the experiment.

Stimuli

Stimuli with varying predictability were created for this experiment using the MUST stimulus set (Clemente et al., 2020). This stimulus set consists of short melodies in a familiar Western tonal style, which were specially composed, thereby ensuring that they would be unfamiliar to participants. The stimuli are all stylistically stereotypical with some variation in internal complexity (length and variation of pitch and duration). This provides a concrete basis for introducing unpredictable notes. To achieve this, the IDyOM model (Pearce, 2005, 2018) was used to select note replacements varying continuously in predictability. IDyOM was configured to predict chromatic pitch using linked representations of chromatic pitch interval and scale degree, following previous research (Gold et al., 2019; Pearce, 2018; Quiroga-Martinez et al., 2019). Once trained, IDyOM calculated the conditional probabilities of each note in the 200 MUST melodies, using its BOTH configuration: combining LTMs (trained on the corpus) and STMs (trained individually on each stimulus).

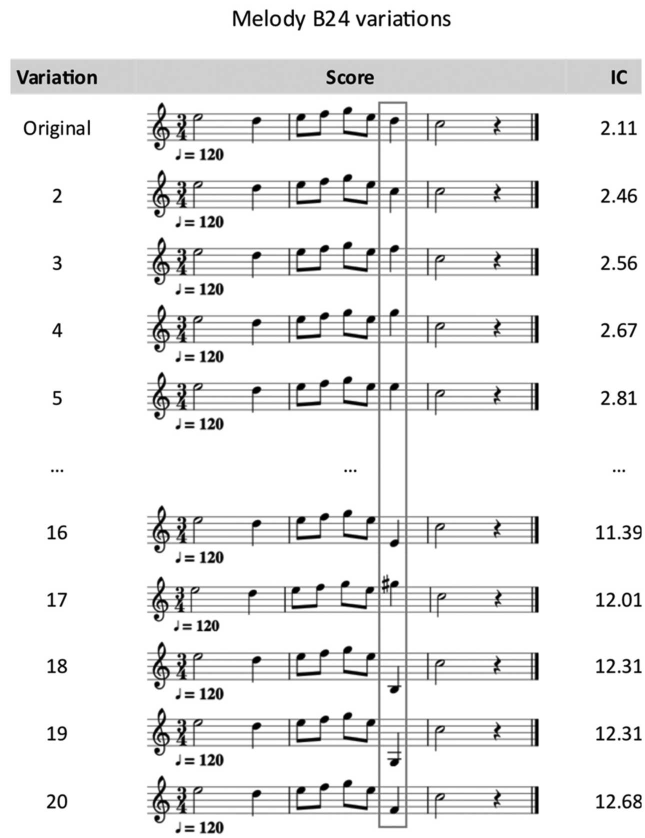
A target note was identified in each stimulus by selecting one pitch from the last half of the melody (excluding the last note) with an IC value closest to the median IC value of all the notes in the melody. At each target note location, IDyOM provided the probabilities of the 48 possible pitches, intended to cover the full range of predictability to reduce the likelihood that the resulting error thresholds might reflect floor or ceiling effects given a limited range of variation. Of the 200 melodies in the MUST set, melodies were selected for this experiment if the pitch of the target note corresponded with the most likely pitch at that location (e.g., if the target note was a middle C and IDyOM had identified middle C as the most likely note at that point in the melody). To standardize target note length across the stimuli, the selected melodies were further filtered to include only those in which the target note was a quarter note. To retain the shortest stimuli, the selected melodies were ordered with respect to note count and the 10 melodies with the fewest notes were retained for use in the experiment. This was to ensure that the melodies would be easy to commit to memory.

The 10 selected melodies were then reproduced with 20 variations each (including the original), resulting in 200 melodies with notes of varying probability. In each variation, the target note was replaced by a decreasingly probable pitch (see Figure 1). The resulting 200 melodies were broken up into two counterbalanced subsets with 10 IC balanced variants of each melody so that participants could be tested on 100 at a time. The stimuli were rendered from musical instrument digital interface to audio using a piano timbre.

Procedure

Participants completed a headphone screening test (Woods et al., 2017) assessing the ability to differentiate subtle changes in tones presented in phase and with a 180° phase difference. In each of the six trials, three tones were played in random order, and the participant

Figure 1
Stimulus Variation



Note. A sample of one of the MUST melodies used for the experiment (B24) as well as the variations produced by the IDyOM analysis. The first row shows the melody as it was originally composed. In the second row, the original target note pitch (D) has been replaced with the second most probable pitch (C). This continues to the last row in which the 20th most probable pitch has replaced the target note (F). IC = information content; IDyOM = Information Dynamics of Music.

clicked a button on screen to identify the quietest tone. Participants listening on speakers should not be able to differentiate between the tones, as the out-of-phase tones cancel each other out. Those who made two or more errors (threshold as recommended by the original authors) were offered a second chance to take the test. Participants who failed the test a second time were compensated for their time and did not proceed to the next stage.

Following the headphone test, participants completed two questionnaires. The first was the musical training subscale of the Goldsmiths Musical Sophistication Index (MSI; Müllensiefen et al., 2014). This survey assesses the participant's experience playing a musical instrument or studying music theory. The second was the openness to experience questionnaire from the International Personality Item Pool (https://ipip.ori.org/).

During the exposure stage of the experiment, participants listened to the 10 original melodies 3 times each, presented randomly. Eight memory tasks were interspersed among the melodies in pseudo-random order (every five melodies ± 2) in which participants signaled whether they recognized the exposure melodies. Three previously unheard melodies from the MUST collection were used as foils to ensure that both familiar and unfamiliar melodies appeared in the memory task. During this exposure period, participants also completed four free association tasks with the prompts "soap," "balloon," "lens," and "clock" (Beaty et al., 2019). In these tasks, participants read the prompt and then typed five associated words into a text box. Participants who responded correctly to at least six of the eight melody retention tasks continued to the next phase. Those participants who did not pass the exposure phase ($N = 38$) were compensated for their time and did not proceed. The motivation for the retention task was to ensure that changes to the melodies would elicit veridical surprise.

In the testing phase, participants were informed that they would be listening to the exposure melodies again; however, this time they would be performed by a "student improviser" and may contain errors. The instructions explained that there were no right answers to the questions in this part of the experiment, and that they should base their responses on their own feelings. They were then presented with one of the two subsets of 100 modified melodies and asked to rate their liking (1–100), their surprise (1–100), and tick "yes" or "no" to signal whether they thought the melody contained an error.

Analysis

The SemDis platform (https://semdis.wlu.psu.edu/) was used to analyze the free association task data. As recommended by Beaty and Johnson (2021), data were first checked for spelling errors and to make sure that participants did not put random, task-unrelated words. Word associations were analyzed using the GloVe semantic space (Pennington et al., 2014) as it has been shown to correlate strongly with human creativity ratings (Dumas et al., 2020). Gold MSI and openness to experience scores were calculated based on the validated rubrics.

The dependent variables collected in the experiment were the liking and surprise ratings as well as the binary report of error perception. The independent predictor variable was the target note IC of the melody variants. The variables that we expected to have a modulatory effect on relationships between independent and dependent variables were creativity (or divergent thinking, as measured by the free association task), openness to experience, and musical sophistication scores. The question of how liking and surprise are affected by predictability was addressed using multiple regression analyses, comparing linear and quadratic models. Ratings for liking and surprise were normalized separately for each participant to remove bias from participants who generally did or did not like the corpus as a whole and instead examine relationships between relative liking and surprise for each participant. Goodness of fit was evaluated with Akaike's information criterion and Bayesian information criterion.

The relationship between predictability and error perception was analyzed by computing a decision boundary: the unpredictability (IC) threshold at which a note had a 50% chance of being judged as an error, according to a generalized linear mixed model (GLMM) fitted to the data. Since the error data were binary (yes/no), the GLMM was configured to use logistic regression (i.e., binomial family with a logit link function) with participant included as a random effect (both slopes and intercepts). Error decision boundaries were estimated both at the group level and for individual participants, using the fixed and random effects, respectively. The overall model was compared for significance with a null model containing intercept-only fixed effects plus the random effect of participant.

Modulatory effects of creativity, openness, and musical sophistication were analyzed with Pearson correlations between the individual scores with individual slopes from linear mixed-effect models relating liking and surprise with unpredictability (IC) as well as the individual decision boundaries for error described above. Mixed-effects models for liking and surprise used raw scores for the dependent variables with random intercepts and slopes for participants and random intercepts for stimuli. p values were Bonferroni corrected for the number of correlations computed.

Results and Discussion

The questions we sought to answer were (a) how liking and surprise are influenced by pitch predictability; (b) how error perception is related to pitch predictability; and (c) whether liking, surprise, and error perception threshold are influenced by individual differences in openness to experience, musical background, or creativity.

Liking and Surprise Are Strongly Linked to Predictability

The comparison of linear and quadratic regression models (see Table 1) shows that quadratic regression models provided a better fit for the effect of IC on both liking and surprise. Significant relationships were found between liking and IC, $R^2 = .61$, $F(2, 197) = 154$, $p > .001$, and also between surprise and IC, $R^2 = .53$, $F(2, 197) = 111$, $p > .001$. The effect size for IC was large both for liking, $\eta^2 = .60$, and surprise, $\eta^2 = .48$, while the effect size for the quadratic effect of IC was small to moderate both for liking, $\eta^2 = .07$, and surprise, $\eta^2 = .16$.

Table 1
Model Comparisons

Effect	Linear regression coefficient	Quadratic regression coefficient
Liking ~ IC		
Intercept	.892***	1.27***
Variant IC	−.131***	−.27***
Variant IC (quadratic)		.0096***
R^2	.58	.61
AIC	181.54	168.39
BIC	191.44	181.59
Residual sum of square	28.17	26.11
Sum of square		2.05
F		15.5***
Surprise ~ IC		
Intercept	−.352***	−.658***
Variant IC	.0537***	.164***
Variant IC (quadratic)		−.008***
R^2	.44	.53
AIC	−62.6	−95.1
BIC	−52.7	−81.9
Residual sum of square	8.31	6.99
Sum of square		1.32
F		37.03***

Note. Regression coefficients and significance for liking and surprise responses. ANOVA comparing the significance of model fit have a grey background. IC = information content; AIC = Akaike's information criteria; BIC = Bayesian information criterion; ANOVA = analysis of variance.
* $p < .05$. ** $p < .01$. *** $p < .001$.

As hypothesized, surprise increased with increasing unpredictability, providing validation of the use of IC as an operational measure of perceived unpredictability. However, contrary to our hypothesis, these results do not suggest the presence of an inverted U curve with respect to the relationship between unpredictability and liking. The average liking and surprise of the exposure melodies (see Figure 2a and 2b, gray points) show that the original melodies experienced in the exposure phase were most liked and, among the melody variants (black points), greater deviation in unpredictability from the original melodies resulted in greater surprise and reduced liking.

A limitation of these findings is that because the original melodies were used in the exposure phase of the experiment to acclimate participants to the style and set a primary expectation, we are unable to tell whether liking of the original melody is due to veridical familiarity per se or because it is the most probable note as appearing in the original melody (see Figure 1). To address this, Experiment 2 varies the presence of prior familiarization with the original melody and whether the selected target note is the most probable note.

Error Perception Is Related to Predictability

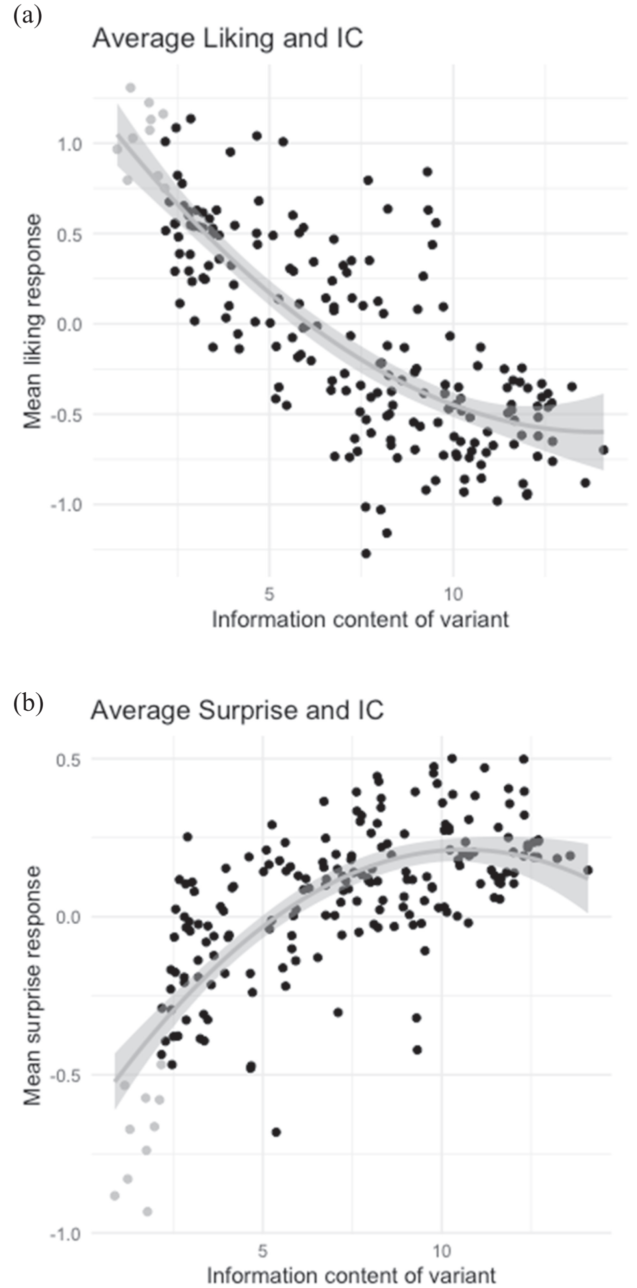
Logistic regression showed a significant relationship between unpredictability (IC) and error perception, $\chi^2(1) = 380.34$, $p < .01$, with coefficients of determination of .23 (marginal, fixed effects only) and .38 (conditional, fixed, and random effects). The overall odds ratio for the model is 1.46, 95% confidence interval (CI) [1.43, 1.49]. Figure 3 plots the individual curves as well as the fitted fixed effect, which produced a decision boundary of 6.83. The significant variation across participants shows differences both in terms of sensitivity to unpredictable notes (i.e., slope) but also in terms of the threshold of IC at which they became more likely to perceive an error (decision boundary), with some participants showing greater tolerance of unpredictability than others.

Individual Differences in Creativity, Openness, and Musical Sophistication Do Not Correlate With Liking, Surprise, or Error Perception

None of the individual characteristics (creativity, openness, and musical sophistication) were found to have any significant modulatory effect on individual error decision boundary or individual relationships between IC and liking or surprise (see Figure 4). The only significant relationship that emerged was a positive correlation between decision boundary and liking/IC slope, suggesting that participants who showed greater liking for more unpredictable melodies also showed a greater threshold of unpredictability for error perception.

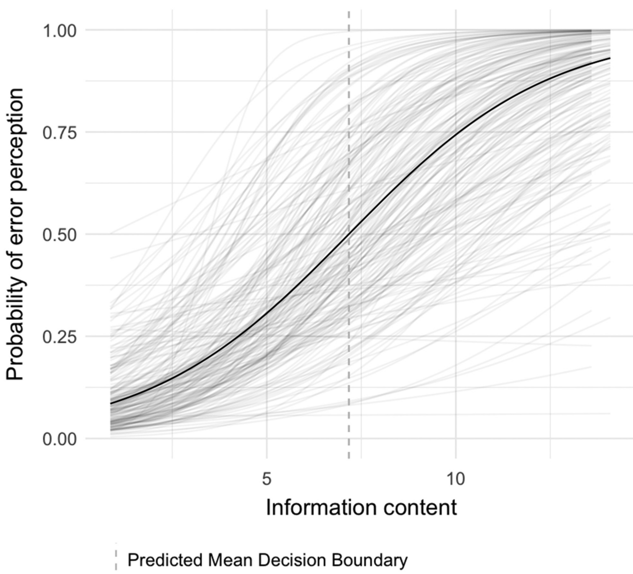
While this suggests a consistent variation in attitude toward unpredictability in terms of both liking and error perception across participants, the individual characteristics associated with these differences remain unclear. This motivates further investigation into the individual characteristics that may underlie these relationships between unpredictability, liking, and error perception. Furthermore, it raises questions about the methods of assessment of creativity used. The free association task may have been suboptimal in this context for measuring comfort or fluidity with extreme divergence as it stops at five words per prompt. It is also limited in its word choice as it provides prompts, limiting participants' ability to exercise originality. We address these limitations in Experiment 2.

Figure 2
Liking and Surprise



Note. Quadratic modelling of normalized liking and surprise ratings as a function of IC. Gray points represent the originally composed melodies which were used in training and testing, black points represent variant melodies used only in testing, confidence intervals are represented in gray. (a) Average normalized liking scores are plotted against the IC of the variant note. A quadratic regression suggests that enjoyment dropped rapidly with the increase in variant note IC. (b) Average normalized surprise is plotted against the IC of the variant note. Surprise increased with variant note IC. IC = information content.

Figure 3
Error Perception Across All Participants



Note. The fine gray lines represent the error perception of individual participants. Across all participants the tendency to perceive an error increased with IC. The black line shows the GLMM of error perception as a function of IC. The model suggests that the predicted mean decision boundary (or 50% likelihood threshold) is crossed at an IC value of 6.83 (dotted line). IC = information content; GLMM = generalized linear mixed model.

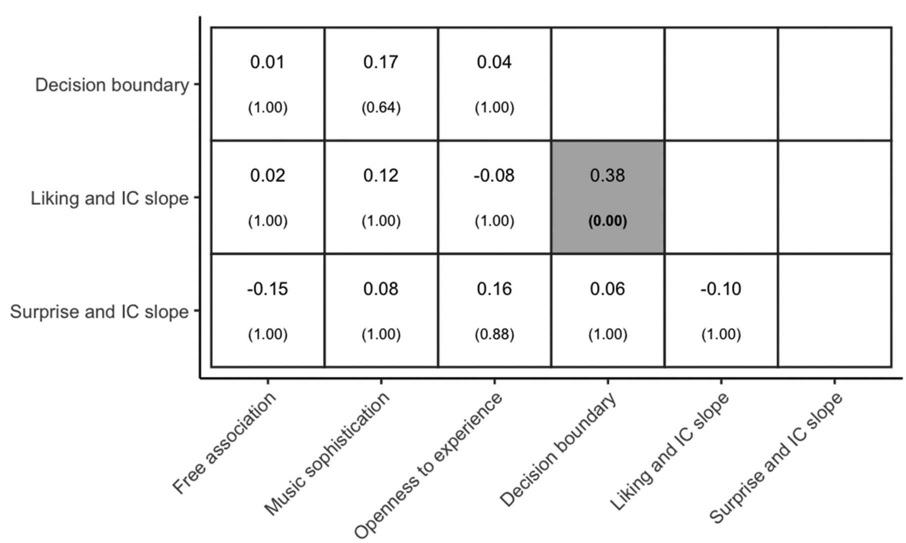
Experiment 2

The results of Experiment 1 left unanswered questions regarding veridical familiarity, note predictability, and modulatory effects of creativity. The second experiment was designed to investigate why the original melodies in Experiment 1 reliably scored so much higher in liking and lower in surprise than the new variants, even those variants that were not particularly unpredictable. In this experiment, the stimuli were reorganized, and new stimuli were introduced to understand the following questions. First, whether we would see a difference in liking between variants of melodies that were veridically familiar and those that were unfamiliar. Second, whether relative predictability of the original target note would have an impact on the liking and surprise of the melody variants. Third, whether error perception would be (more) affected by veridical familiarity or relative predictability. Fourth, whether more creative participants would have a different perception of the melodies than other participants, using a different measure of creativity.

We hypothesized that when melodies were unfamiliar (not presented in the exposure phase), the most probable notes would still be considered enjoyable as there would be no veridical expectation for specific pitches based on previous experience of the melody. We also hypothesized that melodies in which the composer-chosen note was considered to be more unpredictable by IDyOM (see the Method section), meaning that the original melody should be slightly less predictable overall, would encourage participants to be more open to unpredictability, and therefore that relatively unpredictable melody variants would be liked more by listeners. In the same vein, we expected that the decision boundary for perception of error would happen at a higher IC value for less predictable (according to IDyOM) new melodies than for the melodies used in the exposure phase, which should be more veridically familiar.

Given the nonsignificant modulatory effects in Experiment 1, we expanded the battery of individual differences measures and creativity tasks to further investigate their effects on the relationship between unpredictability and surprise, liking and error perception. Specifically, we added the divergent association task (DAT), which is a less goal-directed measure of creativity than the free association task. We also replaced openness to experience with a measure of tendency to engage

Figure 4
Correlations Between Task Responses and Individual Differences



Note. Bonferroni corrected Pearson's correlations between individual characteristics with correlation coefficients and p values (in parentheses). Shading indicates $p < .05$. IC = information content.

in and enjoy thinking (need for cognition) and a measure of tendency to find music rewarding to investigate whether these might underlie the relationships between unpredictability, liking, and error perception observed in Experiment 1.

Method

Participants

Data were collected via Prolific from 96 participants (71 female, nine prefer not to say, aged 19–81, $M = 35.4$, $SD = 12.89$). Three participants were excluded from the analysis because they gave the same answers for all trials. Prolific screening ensured that participants for this experiment had not participated in Experiment 1.

Stimuli

The stimuli used in the experiment were developed using the same process as in Experiment 1. However, for the present experiment, only four randomly selected melodies from the earlier experiment were reserved in which the target note was the highest probability note (1HP) according to IDyOM, and four more were added, which introduced new variables into the experiment. Selection of the new melodies was based on their having the target note chosen by the composer coincide with the second highest probability (2HP) note according to IDyOM (see Figure 5). Variants of these melodies were then generated in the same way as in Experiment 1 such that each of the eight melodies had 20 variants. Half of the melodies were used in the exposure phase of the experiment and half were not. The melodies were evenly split among the conditions, resulting in four groups: new/1HP, new/2HP, exposure/1HP, and exposure/2HP.

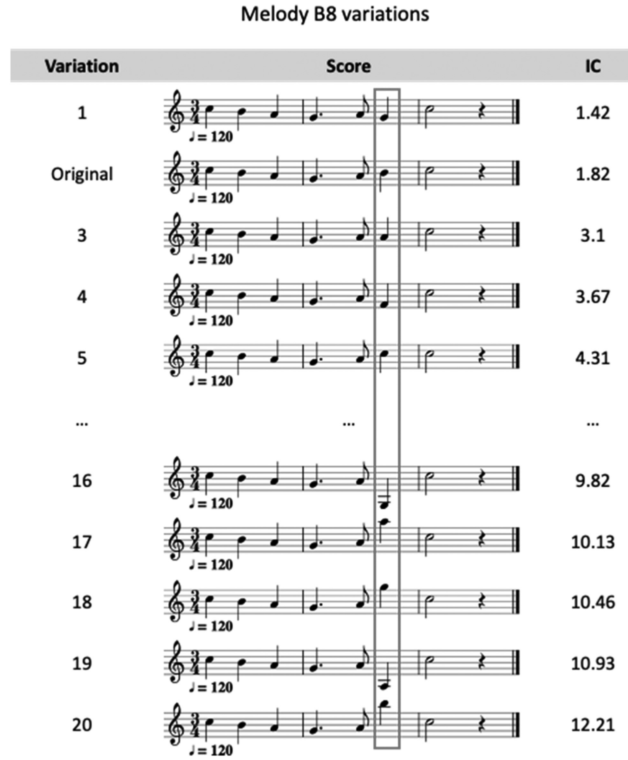
As melody entropy has been shown to have an effect on perception (Gold et al., 2019; Hansen & Pearce, 2014), we sought to avoid potential confounding factors in the interpretation of the results by controlling IC and entropy across 1HP and 2HP melodies with regard to mean IC and entropy values across all notes of the melody as well as those of the target note. IC and entropy values of the original melody versions and each of the target notes were extracted and compared (see Table 2a). Two-sample t tests of IC and entropy for all notes across the two melodies demonstrated that there was no significant difference between the two groups in either IC or entropy (see Table 2b).

Procedure

As in Experiment 1, participants completed a headphone test and responded to three questionnaires and one new task. The questionnaires included the BMRQ (Mas-Herrero et al., 2013), the Need for Cognition Scale (Cacioppo & Petty, 1982), and the same subset of the Goldsmiths MSI (Müllensiefen et al., 2014). A new creativity task was introduced after the exposure and before the testing phase. In the DAT, participants produce 10 words that are as different from each other as possible (Olson et al., 2021). The words must be single words, not proper nouns or technical terms, and participants are encouraged to think of the words on their own as opposed to looking around their environment. The free association task was administered throughout the exposure phase as in the previous experiment.

In the exposure phase, participants were presented with four different 4-s MUST set melodies (Clemente et al., 2020). These original melodies were played 4 times each, and a memory test was administered as in Experiment 1. The main testing phase of the experiment was conducted as with Experiment 1, melodies were played to the participants, and they rated their liking (1–100), surprise (1–100), and error perception (yes/no).

Figure 5
New Stimulus Selection Method



Note. Stimuli for the second experiment included four melodies in which the composer chosen note was the 2HP note in the target position according to IDyOM. An example of one of these 2HP melodies is B8. The gray box signifies the target note and its variants. Note that there is one note (Variation 1) which is more predictable than the note selected by the composer. IC = information content; 2HP = second highest probability; IDyOM = Information Dynamics of Music.

Analysis

Analysis was as for Experiment 1. The online data analysis portal available at https://www.datcreativity.com was used to analyse the DAT data following Olson et al. (2021). In this analysis, seven of the 10 words were conserved (words that do not fit the rules, i.e., nouns, are removed). The words were then analyzed using the GloVe semantic space. The divergent association score represents the semantic distance between the possible pairs of the seven remaining words, which were then averaged and multiplied by 100 following standard practice.

Effects of familiarity and composer-chosen note predictability (IHP vs. 2HP) were analyzed by adding predictors to the regression models for surprise and liking and the GLMM for error perception. These experimental conditions were also included in the mixed-effects models (as random effects with stimulus nested within condition) used to extract individual liking/IC and surprise/IC slopes for the individual differences analysis.

Table 2
Melody Characteristics

Melody	Condition	Target note IC	Target note Entropy	Melody Mean IC	Melody Mean entropy
(a) Comparison of IC and entropy for target note and across full melody					
C30	Untrained, 1HP	1.75	2.98	2.7	3.15
B15	Untrained, 1HP	1.14	2.41	2.81	3.04
C43	Trained, 1HP	1.97	3.03	2.49	3.08
B24	Trained, 1HP	2.11	2.83	2.36	3.03
K16	Untrained, 2HP	2.28	3.06	3.07	2.65
B8	Untrained, 2HP	1.82	2.56	2.22	2.87
B45	Trained, 2HP	2.7	2.63	3.55	2.82
C31	Trained, 2HP	2.01	2.64	2.15	2.97

Measure	1HP *M*	1HP *SD*	2HP *M*	2HP *SD*	*t*	*df*	*p*
(b) Two-sample *t* tests of IC and entropy for all note values across 1HP and 2HP melodies							
Mean IC	2.59	1.64	2.86	1.98	−0.64	67.91	.53
Mean entropy	0.07	1.056	2.8	1.05	1.05	62.62	.3
Target IC	0.74	0.43	2.20	0.38	−1.6	5.92	.16
Target entropy	2.81	0.28	2.72	0.23	0.51	5.76	.63

Note. (a) The stimuli used in the second experiment are presented here with the mean entropy and IC of each melody and each target note in its original configuration. (b) The results of two-sample *t* tests on IC and entropy for all notes across the 1HP and 1HP stimuli melodies showing no statistically significant differences between the two groups. IC = information content; 1HP = highest probability; 2HP = second highest probability.

Table 3
Model Comparisons With Conditions

Effect	Familiarity Linear	Familiarity Quadratic	Predictability Linear	Predictability Quadratic
Liking ~ IC				
Intercept	.632***	.73***	.6016***	.701***
Variant IC	−.093***	−.13***	−.093***	−.129***
Variant IC (quadratic)		.003		.003
Condition	.02	.032	.082	.086
R^2	.512	.515	.52	.523
AIC	90.4	91.2	87.8	88.6
BIC	102.7	106.6	100.1	103.9
RSS	15.678	15.562	15.424	15.307
Sum of square		0.116		0.117
F		1.162		1.196
Surprise ~ IC				
Intercept	−.133**	−.235**	−.279***	−.334***
Variant IC	.023***	.062*	.024***	.045
Variant IC (quadratic)		−.00275		−.001
Condition	−.043	−.05598	.225***	.223***
R^2	.107	.119	.293	.297
AIC	8.1	7.9	−29.3	−28.1
BIC	20.4	23.3	−17.0	−12.7
RSS	9.3759	9.245	7.419	7.382
Sum of square		0.131		0.036
F		2.208		0.769

Note. Linear and quadratic models for each condition are displayed with the results of ANOVAs confirming that the linear models are best fits for all conditions (gray background). The only significant difference between conditions is surprise for the predictability condition. Variations of 2HP melodies were consistently perceived as more surprising by participants than 1HP melodies. IC = information content; AIC = Akaike's information criteria; BIC = Bayesian information criterion; RSS = residual sum of squares; ANOVA = analysis of variance; 2HP = second highest probability; 1HP = highest probability.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Results and Discussion

The key questions we sought to answer were (a) whether liking for a melody would change if the melody were familiar or unfamiliar, (b) whether there would be a difference in liking and surprise between 1HP and 2HP melodies, (c) whether familiarity or composer-chosen note predictability (1HP/2HP) would influence error perception, and (d) whether using new measures of individual differences might provide insight into whether creative participants differ in their perception of the melodies.

No Effect of Veridical Familiarity but 2HP Melodies Are More Surprising Overall

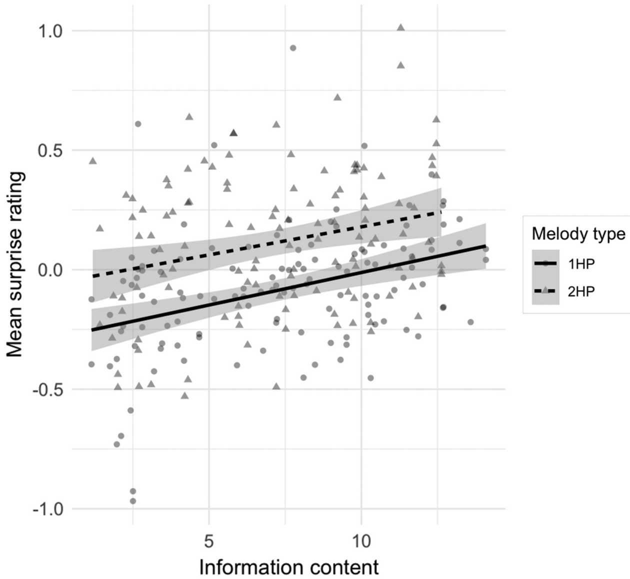
The comparison of linear and quadratic models for each of the response and condition groups (liking/surprise, familiarity/target probability) shows that linear IC provided a better fit than the model containing an additional quadratic IC predictor (see Table 3). Effect sizes for IC were large for liking, $\eta^2 = .47$, and moderate for surprise, $\eta^2 = .09$, whereas those for the quadratic IC predictor were small, $\eta^2 < .01$. As in Experiment 1, liking was significantly negatively correlated with IC while surprise was positively correlated with IC across all conditions (see Table 3). There were no significant effects of any experimental conditions (see Table 3; $\eta^2 < .01$) except for an effect of composer-chosen note predictability (1HP vs. 2HP) on surprise (with a moderate effect size, $\eta^2 = .12$). Specifically, 2HP melodies were associated with greater surprise than 1HP melodies across the entire IC range of pitches, as shown in Figure 6, collapsed across familiarity.

These results suggest that the greater liking and lower surprise observed in Experiment 1 were not due to the veridical familiarity arising from hearing them in the exposure session. Rather, it seems that the effect was due to the lower unpredictability (IC) of the original notes and, as shown in the present experiment, their lower surprise relative to the 2HP notes. Although this did not translate into an effect on liking in the present experiment, it is possible that the difference in unpredictability (IC) between 1HP and 2HP was too subtle. In any case, the results suggest that the exposure session is unnecessary, so it is dropped in Experiment 3, which extends the approach to stimuli in an improvisatory musical style and repetitions of material over a shorter timescale within a session.

Veridical Familiarity and Composer-Chosen Note Predictability Do Not Impact on Error Perception

Logistic regression showed a significant increasing relationship between error perception and unpredictability (IC), $\chi^2(1) = 198.13$, $p < .01$, such that more unpredictable stimuli were more likely to be perceived as containing an error, with coefficients of determination of .18 (marginal, fixed effects only) and .30 (conditional, fixed and random effects). The overall odds ratio for the model is 1.38 (95% CI [1.35, 1.42]), and the decision boundary (i.e., the IC value at which the fitted logistic curve reaches a 50% chance of a boundary being perceived across participants) is 8.58. This is slightly higher than the decision boundary from Experiment 1 (6.83). Adding to the basic logistic regression model (dependent variable: error perception; independent variable: IC) the experimental conditions familiarity (old, new) and

Figure 6
Surprise and Relative Predictability



Note. Modelling of normalized surprise ratings as a function of IC for melody conditions collapsed across familiarity, the lines represent 1HP and 2HP melody conditions and confidence intervals are in grey. 1HP = highest probability; 2HP = second highest probability; IC = information content.

composer-chosen note predictability (1HP, 2HP) as interactive fixed effects improved the fit of the model, $\chi^2(3) = 11.01$, $p = .01$, with a significant effect of the latter, $z = 3.18$, $p < .01$. However, a 2 × 2 analysis of variance on the individual decision boundaries in each condition (see Table 4) showed no main effect of familiarity, $F(1, 392) = 1.36$, $p = .24$, $\eta^2 < .01$, composer-chosen note predictability, $F(1, 392) = 0.29$, $p = .59$, $\eta^2 < .01$, or the interaction between the two, $F(1, 392) = 0.14$, $p = .71$, $\eta^2 < .01$. These results suggest that familiarity had no effect on the relationship between unpredictability (IC) and error perception while composer-chosen note predictability did influence the shape of the fitted logistic curve relating error perception with IC but the effect was too subtle to produce significantly different decision boundaries between 1HP and 2HP melodies.

DAT Performance Correlates With Surprise/IC Slope

Figure 7 shows correlations of participant scores on the individual differences and creativity measures with individual liking/IC slopes, surprise/IC slopes, and decision boundaries. As in Experiment 1, the individual liking/IC slope was significantly correlated with the decision boundary, reinforcing the finding that participants who showed greater liking for improbable melodies had a higher threshold for error perception. Additionally, performance on the DAT was positively correlated with the surprise/IC slope such that more creative individuals showed greater surprise for improbable melodies. That this task produced a significant correlation where the free association did not may reflect the fact that it is less task-oriented and more reflective of internal associative ability since target words are not supplied. Experiment 3 examines all the individual differences investigated in the present experiment along with openness to experience from Experiment 1 with an extension to stimuli in a musical style that may be more associated with creativity.

Table 4
Comparison of Decision Boundaries Across Conditions

Condition	Grand average, M_{DB}	Individual M_{DB}	Individual *SD*	Individual Minimum	Individual Maximum
Distribution of decision boundaries across conditions					
Unfamiliar, 2HP	8.75	8.66	1.88	3.87	13.13
Unfamiliar, 1HP	8.02	8.61	3	3.87	20.29
Familiar, 2HP	9	9.1	3.31	0.18	18.03
Familiar, 1HP	8.36	8.84	2.91	3.84	19.26

Note. Characteristics of error perception for each condition. DB = decision boundary; 2HP = second highest probability; 1HP = highest probability.

Experiment 3

It is possible that the negative relationship between unpredictability and liking found in Experiments 1 and 2 reflects listeners' implicit association of classical music with relatively low levels of unpredictability. Experiment 3 applies the same experimental paradigm to a different style of music, bebop jazz, to understand the ways that expectation, liking, and error perception operate in music where improvisatory variation is a regular characteristic of the musical style. We hypothesized that when listening to jazz music, participants would be more open to unpredictable musical events, showing greater liking for unpredictable notes and a higher threshold for error. We also investigated whether repetition would affect the relationships between unpredictability, surprise, liking, and error perception given the adage: "Once is a mistake, twice is jazz." We hypothesized that repetition would decrease surprise and increase liking.

Method

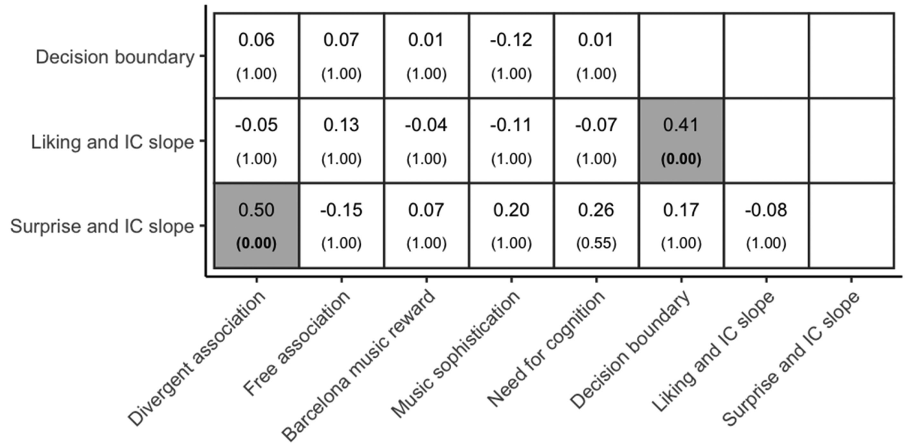
Participants

Prolific was used to collect data from 107 participants (50 female, aged 19–70, $M = 39.91$, $SD = 12.71$). In line with the exclusion criteria outlined below, two participants were removed from the data analysis ($N = 105$). Prolific screening ensured that participants for this experiment had not participated in Experiments 1 or 2.

Stimuli

The stimuli have been taken from the Charlie Parker data set, composed of bebop-style improvisations encoded by Niels Chr. Hansen (Hansen et al., 2016). The original melodies are several minutes in total, and therefore, segments were selected for the experiment. Each melody in the data set was analyzed to select segments that matched the same criteria used in the previous experiments: segments must contain a quarter note in the second half of the melody, and each melody had to represent a musical phrase that was less than 6 s long. The manual selection process resulted in 110 melody segments in total. All segments were characterized by IDyOM using the same configuration and training schedule as the previous experiments. Given the IDyOM analysis, melodies were selected for use in the experiment if the note closest to the median IC in the second half of the melody (excluding the last note) was a quarter note. This was selected as the target note. The resulting segments were ordered

Figure 7
Individual Characteristics and Task Performance



Note. Bonferroni corrected Pearson's correlations between individual characteristics (scores on creativity tasks and personality questionnaires) and comparison with individual slopes of liking/IC, and surprise/IC, as well as individual decision boundaries. Shading indicates $p < .05$. IC = information content.

in terms of the target note's position in increasing order of IC according to IDyOM. Nine segments were selected in total, with IDyOM predicting the target note as the first ($N = 1$), second ($N = 3$), third ($N = 2$), or fourth choice ($N = 3$). For each of these nine base stimuli, 20 variants were created with the target note replaced using IDyOM, reflecting the same process and a similar range of IC as in the previous experiments. The resulting stimuli were organized into three repetition conditions (not repeated, repeated sequentially, repeated nonsequentially—see details below), evenly distributed between target note IC and IDyOM choice. Given the shortness of the stimuli, they were rendered from musical instrument digital interface files to audio in a saxophone timbre to further imply the jazz style.

Procedure

The Woods et al.'s (2017) headphone test was administered in the first instance. During the exposure phase, participants were presented with 15 melody samples. After each sample, they were asked if they recognized it from outside of the experiment. Participants who recognized one or more melodies were excluded automatically from the experiment and paid for their time. Four trials of the free association task were administered randomly during the exposure phase with the prompts "soap," "balloon," "lens," and "clock" (Beaty et al., 2019). The DAT was administered at the end of the exposure phase. Following the testing phase, the participants completed the same subset of the Gold MSI as used in the previous experiments as well as the Barcelona Music Reward, Need for Cognition, and Openness to Experience questionnaires.

The instructions for the testing phase of the experiment informed the participants that they would hear multiple melodies and that some may contain errors. After each melody, participants rated their liking (1–100), surprise (1–100), and error judgement (yes/no). To understand whether perception of melodies changed with repetition, stimuli fell into three categories. First, where melodies were not repeated, second, where melodies were repeated 3 times sequentially, and third, where melodies were presented three different times but distributed randomly among the other stimuli. Given the larger number of stimuli compared to Experiments 1 and 2, participants could optionally take a self-paced break every 22 trials.

Statistical Analysis

Analyses were prepared and conducted in the same way as the first and second experiments.

Results and Discussion

In this experiment we sought to understand: (a) whether style would influence liking and surprise of variant melodies by comparing responses between Experiments 2 (classical) and 3 (Jazz) and especially whether predictable (low IC) notes would elicit lower liking ratings in a Wundt effect for this more improvisatory musical style; (b) whether repetition would increase liking; (c) whether the decision boundary would be higher for jazz (Experiment 3) than classical (Experiment 2), signaling that listeners are more tolerant in making error judgements in a style where improvisation and unpredictability is expected; and finally (d) whether the individual difference measures from Experiments 1 and 2 would correlate with effects of IC on liking, surprise, and error perception for this more improvisatory style.

Liking and Surprise Are Not Significantly Impacted by Style

The comparison of linear and quadratic models for each of the response and condition groups (liking/surprise, familiarity/relative probability group) shows that linear IC predictor provided a better fit than the quadratic predictor (see Table 5). The effect sizes for IC were large for both liking, $\eta^2 = .27$, and surprise, $\eta^2 = .17$, while the effect sizes for the quadratic effect of IC were small, $\eta^2 \leq .01$.

Regression analysis was used to compare the effect of unpredictability on liking and surprise between Experiments 2 and 3 to evaluate differences between jazz and classical music stimuli. Responses

Table 5

Comparison of Linear and Quadratic Regressions for Liking/IC and Surprise/IC in Experiment 3

Effect	Linear regression coefficient	Quadratic regression coefficient
Liking ~ IC		
Intercept	.45774***	.4551***
Variant IC	−.07163***	−.0707
Variant IC (quadratic)		−.0006
R^2	.265	.265
AIC	157.7	159.7
BIC	167.3	172.5
Residual sum of square	24.487	24.487
Sum of square		.00007
F		.00005
Surprise ~ IC		
Intercept	−.2232***	−.3481***
Variant IC	.0375***	.0828**
Variant IC (quadratic)		−.0033
R^2	.171	.182
AIC	26.8	26.3
BIC	36.4	39.1
Residual sum of square	11.83	11.669
Sum of square		.161
F		2.45

Note. A comparison of linear and quadratic regression models in predicting liking and, separately, surprise as a function of IC. For both response types (liking and surprise), linear models provided the best fit. IC = information content; AIC = Akaike's information criteria; BIC = Bayesian information criterion.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 6

Results of Linear Regressions Testing for Liking and Surprise in Jazz and Classical Music

Effect	Liking ~ IC, regression coefficient	Surprise ~ IC, regression coefficient
Intercept	.645***	−.159**
IC	−.093***	.024***
Style	−.161	−.045
Style: IC	.020	.017*
R^2	.384	.156
F	70.87***	20.88***

Note. This analysis identifies significant differences between participant responses to novel IC variants in classical and jazz music genres. IC has a significant effect on all conditions; however the interaction between genre and IC was only significant in the surprise condition. IC = information content.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

to repeated melodies from Experiment 3 were removed to replicate the conditions of Experiment 2. As shown in Table 6, the results demonstrate a significant main effect of IC as in the previous experiments (with a large effect size for both liking, $\eta^2 = .37$, and surprise, $\eta^2 = .13$), no main effect of style but a significant interaction for surprise only (with a small effect size, $\eta^2 = .01$), such that participants were more surprised by high IC notes in jazz than in classical music (see Figure 8). These results are generally consistent with the relationships observed in Experiments 1 and 2: surprise increases and liking decreases with increasing unpredictability, while the effect of unpredictability on surprise is stronger for the jazz stimuli in Experiment 3 than the classical stimuli in Experiment 2.

Repetition Reduced Surprise but Did Not Affect Liking

Liking and surprise were analyzed for each repetition condition (one, two, or three occurrences, immediate or delayed). Regression analyses were conducted separately for liking and surprise in the immediate and delayed conditions with IC and repetition (three levels) as the predictors. The results are shown in Table 7. All conditions showed significant effects of IC as expected. In the immediate repetition condition, there was a significant effect of repetition on surprise (with a medium to large effect size, $\eta^2 < .15$) but not liking, such that surprise decreased for each repetition (see Figure 9). In the delayed repetition group, neither liking nor surprise showed any significant effect of repetition.

Contrary to our expectations and findings in the literature (Madison & Schiölde, 2017; Margulis, 2013), repetition did not increase liking in spite of the fact that surprise did decrease with immediate repetition. It is possible that effects on liking might depend on the repetition (perhaps with some variation) occurring in the context of the thematic development of a melody rather than exact repetition on the next trial. This might also account for the null effects of delayed repetition, which also may have introduced too long a time interval for a strong memory representation to form and allow subsequent recognition. Alternatively, it is possible that a greater number of repetitions would be required for a significant effect to be observed. Finally, it is possible that effects of repetition on liking are critically dependent on conditions such as veridical familiarity, style preferences, or current emotional state that were not replicated in the experimental context.

Error Perception Threshold Increased in Comparison to Classical Stimuli but Was Not Impacted by Repetition

Logistic regression showed a significant relationship between unpredictability (IC) and error perception, $\chi^2(1) = 172.4$, $p < .01$, with coefficients of determination of .06 (marginal, fixed effects only) and .22 (conditional, fixed and random effects). The overall odds ratio for IC is 1.2 (95% CI [1.18, 1.22]), and the decision boundary is 10.49. A two-sample *t* test between the decision boundaries for jazz (no repetition condition only) and classical music across individuals revealed a significant difference, $t(191.65) = 2.04$, $p = .042$, Cohen's $d = 0.28$, such that jazz music had a higher mean decision boundary ($M = 10.14$) than classical music ($M = 8.58$) as illustrated in Figure 10. This suggests that individuals are less likely to judge an unpredictable note as an error in jazz than in classical stimuli. In other words, they are more tolerant of unpredictability when listening to the jazz stimuli in Experiment 3 than the classical stimuli in Experiment 2.

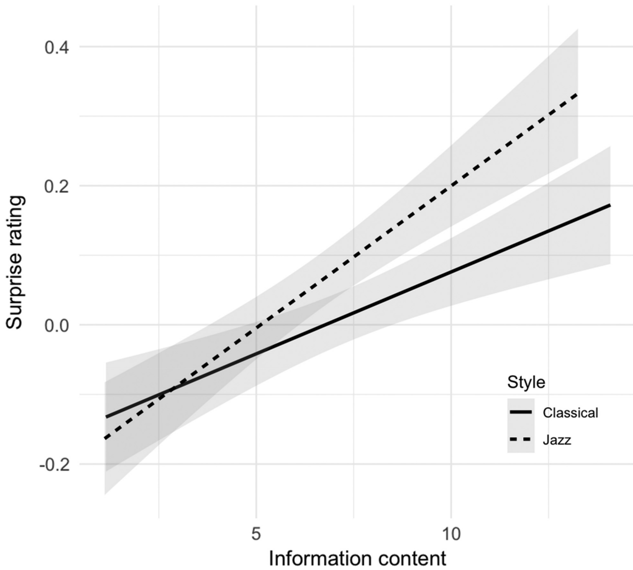
One-way analyses of variance on the individual decision boundaries in each condition showed no main effect of repetition for either the immediate, $F(2, 312) = 0.17$, $p = .84$, $\eta^2 < .01$, or delayed repetition conditions, $F(2, 312) = 0.19$, $p = .83$, $\eta^2 < .01$.

No Impact of Creativity, Musical Experience, or Personality

Figure 11 shows correlations of participant scores on the individual differences and creativity measures with individual liking/IC slopes, surprise/IC slopes, and decision boundaries. None of the

Figure 8

Comparisons of Liking and Surprise Ratings for Jazz and Classical Music



Note. Modelling of normalized surprise ratings as a function of IC for classical and jazz music with confidence intervals in gray. Unpredictable notes in jazz music elicit greater surprise than classical music. IC = information content.

significant relationships from Experiments 1 and 2 were replicated here. However, there was a significant negative correlation between liking/IC slope and surprise/IC slope, suggesting that participants whose surprise was less positively correlated with IC showed a more positive (or less negative) relationship between liking and IC. This might be interpreted as indicating that participants who found unpredictable notes less surprising also tended to like them more. In contrast with both previous experiments, there was no significant correlation between slope of liking/IC and decision boundary.

Table 7

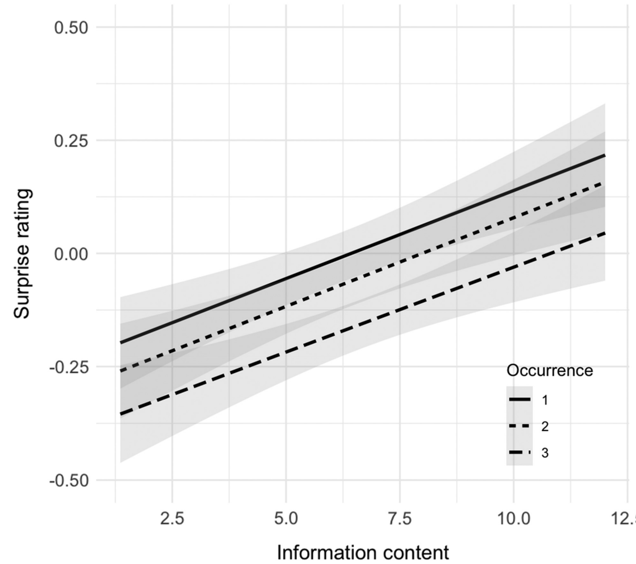
Results of Linear Regressions Testing for Liking and Surprise Across Repetition Conditions

Effect	Liking ~ IC Regression coefficient	Surprise ~ IC Regression coefficient
Immediate repetition		
Intercept	.63***	−.136*
IC	−.09***	.052***
Repetition	−.025	−.123***
R^2	.376	.383
F	53.32***	54.9***
Delayed repetition		
Intercept	.445***	−.18***
IC	−.077***	.037***
Repetition	.006	−.007
R^2	.297	
F	39.32	

Note. Linear regression outcomes showed that only surprise and IC in the immediate repetition condition was significant. IC = information content.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 9

Liking and Surprise Across Conditions



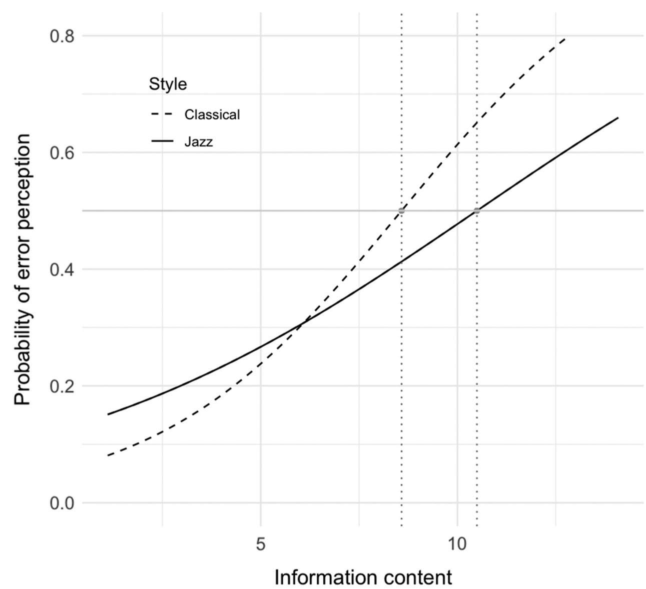
Note. Linear modelling of normalized surprise as a function of IC in the immediate repetition condition. IC = information content.

Relationships Between Individual Difference Measures Across Experiments

Across the three experiments, the results in general show a striking lack of significant relationships between the experimental measures and individual differences in creativity, musical sophistication,

Figure 10

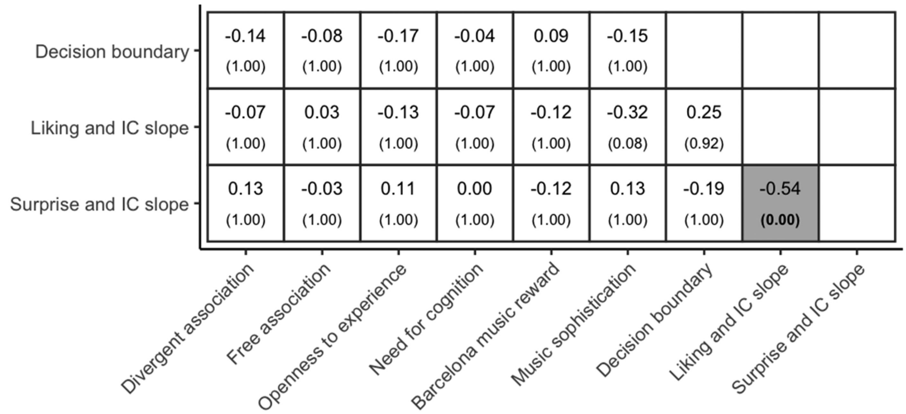
Error Perception Between Genres and Repetition Conditions



Note. GLMMs of error perception as a function of IC for classical (solid) and jazz music (dashed). The mean predicted decision boundaries for each genre are highlighted with dotted lines. GLMM = generalized linear mixed model; IC = information content.

Figure 11

Individual Differences and Responses to Jazz Music Stimuli



Note. Bonferroni corrected Pearson's correlation between individual characteristics and responses to the testing portion of the task including the correlation coefficient and p value in parentheses. Shading indicates $p < .05$. IC = information content.

musical reward, need for cognition, and openness to experience. These data also permit an exploratory investigation of the relationships between these individual differences themselves. Thus, the individual characteristics data (Openness to Experience, Need for Cognition, Barcelona Music Reward, Gold MSI, Free Association Task, and DAT) were compiled across the three experiments. A further 50 participants were added from a pilot version of Experiment 2 in which the participants engaged in the same surveys, resulting in a sample of 453 participants. Not all participants had been presented with every survey or creativity task since they had participated in slightly different experiments. However, the overall results provide a relatively large-scale picture of how these different metrics relate to one another. A correlation plot for this analysis is shown in Figure 12.

Across all participants in all experiments, significant correlations were found between all of openness to experience, music sophistication, and need for cognition. Additionally, music reward correlated significantly with music sophistication and openness to experience but not need for cognition. Divergent association and free association showed no significant correlations with any other variable, including each other, which is surprising given that both are measures of creativity.

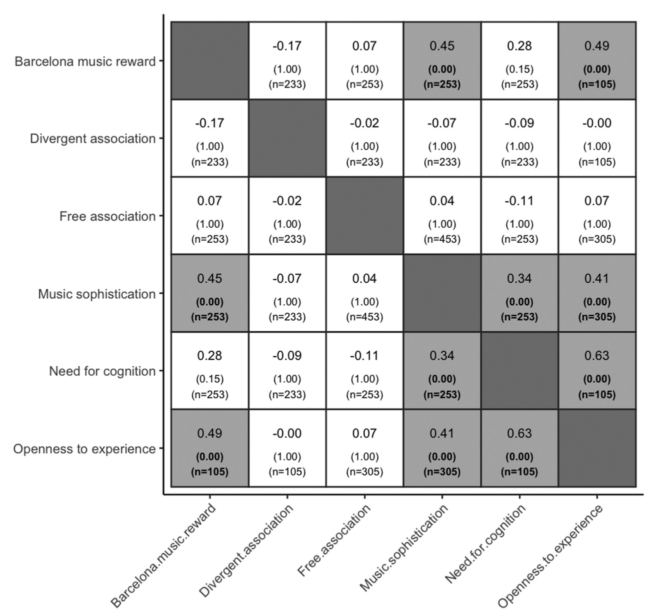
Correlations between openness to experience and need for cognition have been previously reported (Woo et al., 2007). In fact, openness to experience is generally correlated with increased interest in music, which might explain why it is also correlated with music reward and music sophistication (Schäfer & Mehlhorn, 2017). The surprising absence of a correlation between the divergent and free association tasks suggests that these probe different dimensions of creativity, possibly based on task-based application of creativity (Beaty & Kenett, 2023), which is more strongly emphasized in the free association task.

General Discussion

Through the progressive digression paradigm, these experiments established a spectrum of unpredictability in melodies, allowing observation of the resulting changes in listeners' perception of liking, surprise, and error. In addition, several experimental conditions were compared to examine the effects of veridical familiarity, composer-chosen note, style, and repetition. Throughout the series of experiments, modulatory effects of individual creativity, personality (openness to experience), and musical experience (sophistication

Figure 12

Correlations of Individual Differences Measures Across Experiments



Note. Bonferroni corrected Pearson's correlations between individual characteristics for all participants. Significant correlations were found among several variables with correlation coefficients and p values (in parentheses) displayed. Because these data were collected across multiple experiments, the number of participants for each correlation is displayed below the p value. Light gray shading indicates $p < .05$.

and reward sensitivity) were assessed to understand individual factors influencing perception of events varying in unpredictability.

Liking and Surprise

In these experiments, reported surprise was very reliably and consistently correlated with IC, which confirmed our expectation that increasingly unpredictable notes would elicit corresponding levels of increasing surprise. This finding serves to validate the use of IDyOM in this novel experimental paradigm. However, liking showed a consistent negative linear relationship with IC, conflicting with the hypothesized Wundt curve reported by Gold et al. (2019), V. K. M. Cheung et al. (2019) as well as other empirical results (see Chmiel & Schubert, 2017, for a review). One explanation for this, offered by Chmiel and Schubert, is that the stimuli are too complex, so that the results only reveal the right-hand descending portion of the inverted U. However, it seems unlikely that these short melodies are sufficiently complex to cover only that portion of the inverted U. In fact, the converse may be true; it may be that the stimuli are sufficiently simple that the variations on the target note stand out even when unpredictability is increased only by a small degree. In the second experiment, we tested whether the negative association between liking and IC observed in the first experiment resulted from the use of familiar melodies where the pitch of the target note (and thus the exposure note) was also the most predictable option for that note. Our hypothesis that introducing more unfamiliar and unpredictable melodies would increase the liking of higher IC variants was not supported.

Another explanation is that the stimuli are too abstract or artificial, or the manipulations too limited. Generally, within a real musical experience, listeners would encounter changes in several collative variables (e.g., complexity, tension, uncertainty), across several musical parameters (e.g., melody, rhythm, harmony) simultaneously and repeatedly, which might provide the conditions necessary for an inverted U curve to emerge (Chmiel & Schubert, 2017, 2019). One way to address this in the future would be to use longer, more complex real-world stimuli and change multiple notes instead of just one, thereby introducing a more complex version of the paradigm, which could also be extended to rhythm, harmony, and other musical parameters. Also, if the experience of listening to music involves processing multiple collative variables at once, the interaction of these variables may be required to observe the inverted U-shaped relationship. Finally, it is possible that this finding reflects contamination between ratings, which were presented on a single page (e.g., a rating for high surprise may have negatively influenced the liking rating). While this seems unlikely to have had such a strong effect, it is a possibility that future research should take steps to avoid.

Overall, the results suggest that regardless of exposure, for these short and relatively simple melodies, listeners consistently gravitate toward highly predictable pitches, regarding deviations from these as surprising and displeasurable.

Error

The results strongly suggest a relationship between error perception and unpredictability, following a logistic function such that meaningful decision boundaries could be identified. The results of Experiments 1 and 2 also highlighted a relationship between liking and error perception such that participants who showed greater liking for unpredictable melodies also showed a greater threshold for error perception. It would be an interesting extension to investigate whether these boundaries correspond to principles of musical composition or analysis.

Another striking finding with respect to error perception was the difference between decision boundaries in classical and jazz music. This result suggests that the participants attuned their expectations to the musical style, being more restrictive when judging the MUST stimuli in terms of Western tonal functional syntax and less restrictive when judging the jazz stimuli. This is consistent with framing effects previously observed, whereby liking is sensitive to contextual information surrounding a piece of music or visual art rather than the work itself (Ara & Marco-Pallarés, 2021; Kirk et al., 2009; Shank et al., 2023). While the empirical literature generally focuses on observing expectation violations in isolation (Koelsch & Jentschke, 2010; Ruiz et al., 2009), Huron (2006) suggested that listeners form different internal predictive models for the musical styles with which they are familiar, much as bilingual speakers do for natural language, and choose an appropriate model depending on sensory input. The present results suggest that this idea should be extended to style-specific framing effects.

If such differences can be observed between styles, this begs the question of what other factors may modulate error perception. Cultural difference seems an interesting path for investigation (Klarlund et al., 2023). Further research could apply the tasks in these experiments to jazz musicians or participants who listen to contemporary noise styles or participants from non-Western musical cultures and observe variations in error thresholds for different musical styles and cultures. Investigations along these lines have already shown that jazz musicians have different expectations when listening to jazz music than do classical musicians or nonmusicians (Hansen et al., 2016) but have not been conducted for error perception. Given observed effects of predictive entropy (uncertainty) on both perception (Hansen et al., 2021) and liking (V. K. M. Cheung et al., 2019; Gold et al., 2019), future research should examine the effects of entropy on musical error perception, both in isolation and in interaction with IC.

There are some limitations to the present investigation of musical error perception. First, it remains possible that genre differences in error perception reflected lower stylistic familiarity and hence confidence in classifying errors for jazz than classical stimuli. Future research should assess participants' stylistic familiarity with the musical styles involved and take confidence ratings for the error perception task. Second, although a difference in error decision boundary was observed between jazz and classical stimuli, the effect size was small and a within-participant style comparison would be more powerful. Finally, error decision boundaries were computed by fitting a GLMM and taking the IC value corresponding to a 50% chance of an error being perceived. While this method proved useful, it does not take into account the slope, which, given the nonlinear relationship between error perception and IC, differs at different levels of IC. A more sophisticated measure might also consider the slope or changes in slope as well as the absolute threshold.

Creativity

Based on the literature, we hypothesized that individual characteristics related to creativity, personality (openness to experience and need for cognition), and musical experience (musical reward sensitivity and musical sophistication) would provide modulatory influences on the relationship between liking and unpredictability (IC), but the results of these experiments did not, in general, support this hypothesis. While there was a correlation between the divergent association test of creativity in Experiment 2, this was not replicated in Experiment 3, warranting a cautious assessment of its reliability.

We did however find that the measurements of those characteristics correlated with each other in ways that were consistent with the literature. For example, studies have shown that openness to experience is often correlated with other musical characteristics (Schäfer & Mehlhorn, 2017), and we found this to be reliable across experiments. However, the lack of correlation between the free association and DATs was surprising. This may be because the two measures assess different aspects of creativity, with the free association task being more task-directed in providing prompt words, whereas the DAT is less task-directed in asking for spontaneous production of words that are as different from each other as possible (see Beaty & Kenett, 2023).

Testing creativity is difficult and can be done in many ways, and the present results may reflect in some part the creativity measures employed. Neither of the semantic association tasks showed reliable correlations with task performance across all three experiments, which may have been due to several factors. First, we did not ask participants to "be creative" in their answers, as literature has shown that this type of instruction can lead to a significant increase in the desired behavior (Christensen et al., 1957). Second, there are arguably many varieties of creativity, and it may be that the linguistic tasks that we used did not account for a more musical or auditory kind of creativity. Third, the free association task was shortened to five responses per cue in order to fit within the experimental constraints. Considering that the original version of this task asks for 19 associations, it could be that we did not capture a sufficient number of associations. Finally, it may be that perception of unpredictability and its effect on liking and error perception are not in fact mediated by individual creativity, which would be consistent with the failure of openness to experience to predict task performance in any way, since the two measures are somewhat related (Jauk et al., 2014; Puryear et al., 2017). Considering these limitations, future studies could adopt more comprehensive and multidimensional measures of creativity to clarify whether the perceptual factors associated with perception of unpredictability in music are in fact associated with creative production.

Conclusion

In conclusion, this series of experiments sheds light on the complicated relationship between perception of musical unpredictability, liking, error perception, and individual factors (creativity, personality, and musical experience). The findings demonstrate a meaningful relationship between error perception and unpredictability, question the notion that creativity influences error perception, and highlight the significance of contextual factors such as music style. This research introduces a useful paradigm for systematically investigating the effects of unpredictability on error perception and paves the way for further investigations into the role of expectation in aesthetic experience of music.

References

Ara, A., & Marco-Pallarés, J. (2021). Different theta connectivity patterns underlie pleasantness evoked by familiar and unfamiliar music. *Scientific Reports*, *11*(1), Article 18523. https://doi.org/10.1038/s41598-021-98033-5

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, *53*(2), 757–780. https://doi.org/10.3758/s13428-020-01453-w

Beaty, R. E., & Kenett, Y. N. (2023). Associative thinking at the core of creativity. *Trends in Cognitive Sciences*, *27*(7), 671–683. https://doi.org/10.1016/j.tics.2023.04.004

Beaty, R. E., Kenett, Y. N., & Hass, R. W. (2019). *Fanning creative thought: Semantic richness impacts divergent thinking*. Proceedings of the 41st Annual Meeting of the Cognitive Science Society (pp. 126–131).

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. McGraw-Hill. https://doi.org/10.1037/11164-000

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1278–1296. https://doi.org/10.1037//0096-1523.21.6.1278

Brand, M., & Burnsed, V. (1981). Music abilities and experiences as predictors of error-detection skill. *Journal of Research in Music Education*, *29*(2), 91–96. https://doi.org/10.2307/3345017

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cardona, G., Ferreri, L., Lorenzo-Seva, U., Russo, F. A., & Rodriguez-Fornells, A. (2022). The forgotten role of absorption in music reward. *Annals of the New York Academy of Sciences*, *1514*(1), 142–154. https://doi.org/10.1111/nyas.14790

Cheung, V., Harrison, P., Koelsch, S., Pearce, M., Friederici, A., & Meyer, L. (2023). Cognitive and sensory expectations independently shape musical expectancy and pleasure. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1895), Article 20220420. https://doi.org/10.1098/rstb.2022.0420

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology: CB*, *29*(23), 4084–4092.e4. https://doi.org/10.1016/j.cub.2019.09.067

Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, *45*(6), 886–909. https://doi.org/10.1177/0305735617697507

Chmiel, A., & Schubert, E. (2019). Unusualness as a predictor of music preference. *Musicae Scientiae*, *23*(4), 426–441. https://doi.org/10.1177/1029864917752545

Christensen, P. R., Guilford, J. P., & Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *Journal of Experimental Psychology*, *53*(2), 82–88. https://doi.org/10.1037/h0045461

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clemente, A., Pearce, M. T., & Nadal, M. (2022). Musical aesthetic sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, *16*(1), 58–73. https://doi.org/10.1037/aca0000381

Clemente, A., Pearce, M. T., Skov, M., & Nadal, M. (2021). Evaluative judgment across domains: Liking balance, contour, symmetry and complexity in melodies and visual designs. *Brain and Cognition*, *151*, Article 105729. https://doi.org/10.1016/j.bandc.2021.105729

Clemente, A., Vila-Vidal, M., Pearce, M. T., Aguiló, G., Corradi, G., & Nadal, M. (2020). A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: Behavioral and computational assessments. *Behavior Research Methods*, *52*(4), 1491–1509. https://doi.org/10.3758/s13428-019-01329-8

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73. https://doi.org/10.1080/09298219508570672

Creighton, H. (1966). *Songs and ballads from Nova Scotia*. Dover.

Crowder, R. G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 111–145). Oxford University Press.

Crozier, J. B. (1974). Verbal and exploratory responses to sound sequences varying in uncertainty level. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps towards an objective psychology of aesthetic appreciation* (pp. 27–90). Hemisphere Publishing Co.

Dumas, D., Organisciak, P., & Doherty, M. (2020). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(4), 645–663. https://doi.org/10.1037/aca0000319

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(47), 9397–9409. https://doi.org/10.1523/JNEUROSCI.0428-19.2019

Halpern, A. R., Zioga, I., Shankleman, M., Lindsen, J., Pearce, M. T., & Bhattarcharya, J. (2017). That note sounds wrong! Age-related effects in processing of musical expectation. *Brain and Cognition*, *113*, 1–9. https://doi.org/10.1016/j.bandc.2016.12.006

Hansen, N. C., Kragness, H., Vuust, P., Trainor, L., & Pearce, M. T. (2021). Predictive uncertainty underlies auditory boundary perception. *Psychological Science*, *32*(9), 1416–1425. https://doi.org/10.1177/0956797621997349

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*(SEP), Article 1052. https://doi.org/10.3389/fpsyg.2014.01052

Hansen, N. C., Vuust, P., & Pearce, M. (2016). "If you have to ask, you'll never know": Effects of specialised stylistic expertise on predictive processing of music. *PLOS ONE*, *11*(10), Article e0163584. https://doi.org/10.1371/journal.pone.0163584

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press. https://doi.org/10.7551/mitpress/6575.001.0001

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European Journal of Personality*, *28*(1), 95–105. https://doi.org/10.1002/per.1941

Justus, T. C., & Bharucha, J. J. (2001). Modularity in musical processing: The automaticity of harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 1000–1011. https://doi.org/10.1037/0096-1523.27.4.1000

Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *Neuroimage*, *44*(3), 1125–1132. https://doi.org/10.1016/j.neuroimage.2008.10.009

Klarlund, M., Brattico, E., Pearce, M. T., Wu, Y., Vuust, P., Overgaard, M., & Du, Y. (2023). Worlds apart? Testing the cultural distance hypothesis in music perception of Chinese and Western listeners. *Cognition*, *235*, Article 105405. https://doi.org/10.1016/j.cognition.2023.105405

Koelsch, S., & Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *Journal of Cognitive Neuroscience*, *22*(10), 2251–2262. https://doi.org/10.1162/jocn.2009.21338

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *eLife*, *5*, Article e20317. https://doi.org/10.7554/eLife.20317

Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. *Frontiers in Neuroscience*, *11*(MAR), Article 147. https://doi.org/10.3389/fnins.2017.00147

Margulis, E. (2013). Aesthetic responses to repetition in unfamiliar music. *Empirical Studies of the Arts*, *31*(1), 45–57. https://doi.org/10.2190/EM.31.1.c

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Perception*, *31*(2), 118–138. https://doi.org/10.1525/mp.2013.31.2.118

Mccrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, *52*(6), 1258–1265. https://doi.org/10.1037/0022-3514.52.6.1258

McMullen, P. T., & Arnold, M. J. (1976). Preference and interest as functions of distributional redundancy in rhythmic sequences. *Journal of Research in Music Education*, *24*(1), 22–31. https://doi.org/10.2307/3345063

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, *9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(25), Article e2022340118. https://doi.org/10.1073/pnas.2022340118

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition* [Doctor of Philosophy]. Department of Computing City University.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global vectors for word representation*. In EMNLP 2014—2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. https://doi.org/10.3115/v1/d14-1162

Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*(5), 884–902. https://doi.org/10.3758/BF03201171

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(1), 59–68. https://doi.org/10.1037/aca0000079

Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., & Vuust, P. (2019). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex*, *120*, 181–200. https://doi.org/10.1016/j.cortex.2019.06.010

Riemenschneider, A. (1941). *371 Harmonised chorales and 69 chorale melodies with figured bass*. G. Schirmer.

Ruiz, M. H., Koelsch, S., & Bhattacharya, J. (2009). Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human Brain Mapping*, *30*(4), 1207–1225. https://doi.org/10.1002/hbm.20584

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), Article 251. https://doi.org/10.3390/brainsci10040251

Sauvé, S., & Pearce, M. T. (2019). Information-theoretic modelling of perceived musical complexity. *Music Perception*, *37*(2), 165–178. https://doi.org/10.1525/mp.2019.37.2.165

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, *116*, 265–273. https://doi.org/10.1016/J.PAID.2017.04.061

Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide [computer database]*. CCARH.

Sears, D. R. W., Pearce, M. T., Spitzer, J., Caplin, W. E., & McAdams, S. (2019). Expectations for tonal cadences: Sensory and cognitive priming effects. *Quarterly Journal of Experimental Psychology*, *72*(6), 1422–1438. https://doi.org/10.1177/1747021818814472

Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2023). AI Composer bias: Listeners like music less when they think it was composed by an AI. *Journal of Experimental Psychology: Applied*, *29*(3), 676–692. https://doi.org/10.1037/xap0000447

Sheldon, D. A. (2004). Effects of multiple listenings on error-detection acuity in multivoice, multitimbral musical examples. *Journal of Research in Music Education*, *52*(2), 102–115. https://doi.org/10.2307/3345433

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748. https://doi.org/10.1016/j.tics.2017.06.007

Tillmann, B., & Bigand, E. (2010). Musical structure processing after repeated listening: Schematic expectations resist veridical expectations. *Musicae Scientiae*, *14*(2_Suppl), 33–47. https://doi.org/10.1177/10298649100140S204

Vella, E. J., & Mills, G. (2017). Personality, uses of music, and music preference: The influence of openness to experience and extraversion. *Psychology of Music*, *45*(3), 338–354. https://doi.org/10.1177/0305735616658957

Vitz, P. C. (1966). Affect as a function of stimulus variation. *Journal of Experimental Psychology*, *71*(1), 74–79. https://doi.org/10.1037/h0022619

Woo, S. E., Harms, P. D., & Kuncel, N. R. (2007). Integrating personality and intelligence: Typical intellectual engagement and need for cognition. *Personality and Individual Differences*, *43*(6), 1635–1639. https://doi.org/10.1016/J.PAID.2007.04.022

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Zioga, I., Harrison, P. M. C., Pearce, M. T., Bhattacharya, J., & Luft, C. D. B. (2019). From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. *NeuroImage*, *206*, Article 116311. https://doi.org/10.1016/j.neuroimage.2019.116311

Zioga, I., Harrison, P. M. C., Pearce, M. T., Bhattacharya, J., & Luft, C. D. B. (2024). The association between liking, learning and creativity in music. *Scientific Reports*, *14*, Article 19048. https://doi.org/10.1038/s41598-024-70027-z

Received June 21, 2023
Revision received July 29, 2025
Accepted July 30, 2025 ■