The rewards of Muzak: elevator music as a tool for the quantitative characterization of emotion and preference

Ellie Bean Abrams[1,2,3], David Poeppel[1,2], Pablo Ripollés[1,2,3]

[1] Department of Psychology, New York University
[2] Center for Language, Music, and Emotion (CLaME), New York University
[3] Music and Audio Research Laboratory (MARL), New York University

ellie.abrams@nyu.edu

Introduction

Emotion – the conscious awareness that something of psychological or biological importance is affecting us (Ledoux and Brown, 2017) – is central to the human experience, interacting with multiple cognitive domains. Emotion influences our thoughts, behavior, and physiological responses from birth, with the general population experiencing at least some emotion 90% of the time during an average day (Trempe et al., 2015). In addition to experiencing emotions as a result of external or involuntary changes, humans also use external stimuli to modulate their emotions voluntarily. In this vein, there is a stimulus –specific to humans – that is consistently used across individuals and cultures tool for emotion regulation: music (Juslin & Vastfjall, 2008; Goethem & Sloboda, 2011; Mas-Herrero et al., 2020). Indeed, music is known to induce a complex range of emotions cross-culturally (Cowen et al., 202) and engages brain networks involved in emotion and reward (Koelsch 2014; 2020; Mas-Herrero et al., 2020). In fact, music plays a crucial role in overall mood regulation and stress management (Juslin & Vastfjall, 2008; Goethem & Sloboda, 2011; Ferreri et al., 2021; Mas-Herrero et al., 2020). A key neural mechanism which allows music to directly and immediately induce emotional or mood change is through the human reward system. Specifically, music is well-known to activate the reward system and to induce the release of dopamine (Ferreri et al., 2019; Mas-Herrero et al., 2018; Blood and Zatorre, 2001; Salimpoor et al., 2011). Unlike other primary or secondary reinforcers (e.g., money) which can downregulate the reward system over time (Murayama et al., 2010), music – an abstract reward –can induce maximal reward and dopaminergic responses repeatedly (Martinez-Molina et al., 2016), making it suitable for spontaneous and/or direct mood regulation (Hennessy et al., 2021; Mas-Herrero et al., 2020; Stewart et al., 2019). Music is also naturalistic and instantly accessible, and it is known to induce reward- and emotion-related responses almost universally (Mehr et al., 2019; Mas-Herrero et al., 2014).

While research show that music is widely used to regulate mood, individual differences play a major role in modulating this process, especially across two dimensions: preference and reward sensitivity. Regarding preference, individuals show vastly different musical tastes even within one specific musical culture (Rentfrow et al., 2011). This variability is due to several psychological and social factors, but studies show that music preference develops according to a range of factors including early exposure and experience, and social associations (Greenberg et al., 2016). Regarding the human reward system, it has recently been shown that sensitivity to *musical* reward specifically exists on a scale which is separate from general sensitivity to reward (Mas-Herrero et al., 2013). The spectrum of musical reward ranges from specific musical anhedonia, in which people may experience reward from other stimuli but not music, to music hyperhedonia, in which people are highly affected and motivated by music across multiple dimensions, including social reward, emotion evocation, and mood regulation (Mas-Herrero et al., 2013, 2014; Martinez-Molina et al., 2016, 2019).

These individual differences make music an interesting and useful stimulus not only for the study of how humans upregulate and downregulate their mood, but also for the study of the cognitive neuroscience of emotion and reward: participants self-select music to maximize their experience of reward, thus allowing us to tailor each individual's pleasure response. Note that humans experience reward not only from primary or secondary reinforcers (e.g., food, money), but also from abstract stimuli (e.g., paintings, music) and that recent research shows that there are differences in the way primary and abstract rewards are represented at the neural level (Mas-Herrero et al., 2020). If a full understanding of how the human reward system is to be reached, research programs cannot neglect the study of abstract rewards, and music is perfectly suited for this endeavor. However, using music as a tool to unveil the neural mechanisms governing human reward and emotional related processes also presents a challenge regarding the nature of the control conditions to be used when doing so: if large individual differences exists in the type of abstract rewards people find pleasurable within a particular enculturation (e.g., there is no universal rewarding musical stimulus for Western listeners), it may be unlikely that there is also a universal neutral musical stimulus within the same particular musical enculturation.

However, among Western musical genres, there is one that was specifically designed to be emotionally neutral. Elevator music, originally termed "Muzak" by its founder, George Owen Squier, was brought to use in the early 1920s as automated elevators became widespread. The installation of elevators everywhere coincided with the development of new technology allowing for the transmission of audio signals via electrical wires as opposed to radio. Thus, music was utilized to keep hesitant new passengers calm and increase elevator use. In the decades to come, elevator or background music was largely produced by Muzak Holdings, who recorded specially formed orchestras playing original pieces or rerecording popular songs of the time (Lanza, 2004). Importantly, the music was tightly controlled in pace, style, and dynamics to induce a soothing effect and unassumingly occupy periods of inactivity across public settings. This resulted in a large body of music designed to induce as neutral a response as possible, with the intended goal of not drawing too much attention to itself.

Elevator music thus presents an interesting and potentially enlightening phenomenon: while preferences of Western music listeners vary greatly from person to person, Muzak was specifically tailored to induce a neutral response *across* listeners with the same Western musical enculturation. Importantly, *Muzak* pulls from several styles of music. Many elevator music pieces have instrumental solos with jazz rhythms and instruments but are stripped of harmonic complexity or creativity. Structurally, typical elevator music borrows from pop music (ABAB, ABACB), with repeated stepwise melodies which rarely veer from an expected chord progression (often, they were simply re-recordings of popular music of the time). In terms of instrumentation, songs typically feature up to 30 instruments, including strings and brass, often resembling a classical orchestra. However, instruments take few expressive liberties, with consistent tempi and dynamics. The intention behind its composition, as well as its structural aspects, make elevator music an opportunity to explore the musical elements of neutrality across listeners with a Western enculturation, as well as to validate a new stimulus which may be used as a control for self-selected pleasurable music in the study of the neural mechanism of human reward and emotional-related processes.

Here, we first characterize elevator music using online behavioral data collection, by recording ratings of pleasure, emotional valence, familiarity, and recognition in comparison with other popular genres of music (Experiments 1 and 2). We then turn to surprisal, a measure of unexpectedness, to assess the perceptual dimensions across which elevator music may be neutral (Experiment 3). We hypothesize that elevator music will elicit less pleasure than other well-known genres, and that pleasure ratings will be consistently neutral. Building off research showing this link between surprisal and liking, we expect elevator music to contain less overall dynamic change in the sense of surprise both behaviorally and computationally.

General Methods

General Experimental Design

All experiments were developed using *oTree*, a Python-based framework for the development of controlled online behavioral experiments (Chen et al., 2016). The online experiments coded using oTree were presented to participants as an HTML webpage running in the Google Chrome web browser (the task was designed to work only in Google Chrome to increase the comparability of the results across participants). Participants were first presented with a few lines summarizing the task and then with an informed consent page. Upon acceptance, detailed instructions for each experiment were presented. In all experiments, we implemented a number of attention checks to enforce participants' compliance. First, after the presentation of each musical excerpt and after providing overall ratings, participants heard 5 seconds of a musical piece. Half of the time, the presented clip was part of the song they just heard and the other half the clip was extracted from another song presented during the experiment. The 5 second clips were always extracted from seconds 30 to 35 of the original 1 minute long musical pieces presented. After hearing the musical clip for 5 seconds, participants had to indicate whether the clip was part of the song they just heard or not, using a two-alternative forced choice design (i.e., yes or no). From these responses, we calculated a *dprime* score that serves as a measure of task compliance (that is, as a measure of whether they were actually paying attention to the songs presented or not). In addition, all participants completed online versions of the Barcelona Music Reward Questionnaire (BMRQ, measuring sensitivity to musical reward; Mas-Herrero et al., 2013) and of the Goldsmiths Musical Sophistication Index (Gold-MSI, a self-report measuring musical skills; Müllensiefen et al., 2015). Of use in our analyses was the musical training subscore of the Gold-MSI, which falls between 0 and 1. In the two questionnaires, we also included a question that served as an attentional check (e.g., *Please, select the option "Agree"*). Finally, the experiment was coded so that participants could not advance to the next page unless the song had been played for the full minute (in the listening trials) or all answers had been provided (in the response trials). Participants had no control over the music player, so that they could not skip to the end of the musical piece.

Participants

Participants were recruited using Amazon Mechanical Turk (AMT), a crowdsourcing platform for the acquisition of large online behavioral datasets. Recent research has replicated a number of tasks from experimental psychology using participants recruited via AMT (e.g., Stroop, Flanker, statistical learning, among others; Crump et al., 2013; Assaneo et al., 2019). Participants were recruited from the United States and were required to have 99% of previous submissions approved on AMT. From the pool of participants who completed the experiments, we excluded: i) those with specific musical anhedonia (i.e., those with a score lower than 66 in the Barcelona Music Reward Questionnaire; Mas-Herrero et al., 2013, 2014); ii) those with a *d-prime* lower than 1 in the attention task for the songs presented iii) those failing both attentional checks when completing the BMRQ and Gold-MSI; and iv) those who did not provide continuous answers while listening to the songs, in the experiments in which continuous responses were required. We collected data from a total of 474 participants, of which 131 were excluded (final sample, adding participants from all experiments was 343, 206 male, 135 female, 2 non-binary, 38.6 ± 10.5 years of age, 80.2 ± 8.01 BMRQ score, 0.58 ± 0.15 Gold-MSI, and 0.42 ± 0.2 Gold-MSI musical training scores).

Statistical Analysis

Given the repeated measures design of all the proposed experiments, we avoided simply averaging values by using linear mixed modeling (LMM). This also allowed us to include random intercepts to account for individual differences in internal scales of pleasure and other subjective scores, and to add control parameters that might account for variance in the data. In this vein, we expected, based on previous research (Mas-Herrero et al., 2013, 2014), that sensitivity to musical reward (i.e. musical hedonia) would modulate music pleasure ratings specifically, and that musical training, as measured by the Gold-MSI, might modulate subjective ratings related to music as well. We thus included both the BMRQ and Gold-MSI in all our models (except for the empty ones) as fixed factors. The use of LMM allowed us to include both scores as continuous variables, instead of using other less elegant solutions such as splitting the population into high and low musical hedonics (as in Ferreri et al., 2021). We performed generalized linear mixed modeling in *R* (version 4.0.2) and *RStudio* (version 1.3.959) using the *lmer4* package.

In each analysis, we generated first an empty model, which contained only random intercepts for participants. Next we added in the main contrast conditions (e.g., musical genre) within the experiment as fixed factors, as well as BMRQ and/or Gold-MSI to generate a minimal model. We then proceeded to generate models that included all possible interactions between the main fixed factors of interest. For each generalized linear mixed models analysis, we selected the model which was the best candidate to explain the variance using the Akaike information criterion (AIC). We considered a model different from another if the difference in AIC was greater than 2, as recommended in the literature (Simmons and Moussali., 2011). This criterion was used in order to balance complexity and goodness of fit. If models were separated by less than two AIC, we selected the model with fewer factors, as this explains the same amount of variance in the data (see Fererri et al., 2021 for a similar approach). Then, the effects of the different predictors and their interactions were assessed using likelihood ratio tests (LRT) using *mixed()* from the *afex* package in R. Contrasts were carried out using the *emmeans* package in R.

For all models, we checked for collinearity using R's *vif* function and confirmed that all factors in all models had values of less than 3. We ran post-hoc power analyses for each main result using the *simr* package with 1000 simulations, all of which showed that the power to find each effect was more than 80% (the minimum power for our results was 89% with 95% confidence interval: 81.17 - 94.38%). This proves that our experiments were sufficiently powered.

Experiment 1

Methods

Participants

We collected data from a total of 51 participants, of which 9 were excluded, for a final total of 42 participants (17 females, mean age = 39.23 ± 12.7 years). Average BMRQ score was 79.17 ± 7.12 and average musical training score as measured by the Gold-MSI was 0.42 ± 0.19.

Experimental Design

Experiment 1 consisted of two musical conditions that differed in the type of stimuli presented. Participants listened to elevator music (*Elevator* condition) or to rock/pop songs (*Reward* condition). The musical pieces were presented following a blocked design with a counterbalanced order of the two conditions. That is, roughly half of the participants (N=20) completed the experiment first rating the songs of the *Elevator* condition and then those of the *Reward* condition, and roughly half (N=22) completed the task in the reverse order. Each block consisted of ten one-minute long musical excerpts, randomized within condition and participant. Participants were instructed to plug in headphones or to use a loudspeaker. Each trial started with the presentation of a musical excerpt, and the participant began each trial at their own pace by clicking a start button. Once the musical piece had been played until the end, participants were allowed to move to the next page were they provided ratings across four measures using a 9 point Likert scale (-4 to 4): **pleasure** (-4 = no pleasure, 0 = some pleasure, 4 = intense pleasure), **emotional valence** (-4 = very sad, 0 = neutral, 4 = very happy)**, recognition** (-4 = I have never heard this song, 0 = I am unsure, 4 = I have heard this song many times, and **familiarity** (-4 = the elements of this song are unfamiliar, 0 = neutral familiarity, 4 = the elements of this song are highly familiar). Note that here we differentiate between exact recognition ("Have you heard this exact song before?") and familiarity ("How familiar does the melody, structure, and instrumentation sound regardless of whether you have heard this specific composition before?") in order to have a more fine-grained measure of novelty for each song in both conditions. This was to address the fact that the rock/pop songs of the *Reward* condition were most likely going to be highly recognizable to our participant pool and also to evaluate how familiar in style and structure elevator music excerpts were. The ratings and rating scales were presented all on the same page. After providing scores for all scales, participants completed the attentional check in which they listened to a five second clip of music and were asked to confirm whether or not it was part of the song they just heard. After this attention task, the next trial began.

Stimuli
The *Elevator* stimuli for Experiments 1 and 2 (Supplementary Tables 1 and 2) were selected from a range of sources, initially including songs from Muzak Orchestra's Stimulus Progression albums, as this specific style was the prototype for the genre of elevator music. *Reward* stimuli consisted of rock/pop music selected from a combination of top charts, including TimeOut's top 40 pop songs of all time, Rolling Stone's top 500 songs of all time, and Billboard's Hot 100 ranging from 1967 to 2014. To maximize the number of pieces played in the experiment, participants heard only one minute of each piece. We sought to choose the most interesting minute of each piece, particularly in the elevator condition, so that any results showing lower ratings for elevator music would not simply be due to choosing a particularly unexciting minute of the piece. It has been shown that increases in surprisal is one of the driving forces behind musical preference (Gold et al., 2019; Cheung et al., 2019). Thus, we chose to select the minute with highest musical surprisal within a song. To extract this minute, we turned to a new perceptual model which collects statistical regularities of over time: Skerritt-Davis & Elhilali's Dynamic Regularity Extraction model (D-REX; Skerritt-Davis & Elhilali, 2018, 2019). The model calculates musical surprisal using a Bayesian framework to model prediction error across various acoustically salient features such as spectral energy, pitch value, and temporal modulation (15 features total; see Skerritt-Davis & Elhilali, 2018, 2019 for details). In addition to surprisal over time for each feature, the model outputs a joint surprisal which collapses across features for a summary measure of overall surprisal. We used the **joint surprisal** as the computational measure of surprise to complement and validate behavioral data, as well as to compare the degree of surprise between music genres. Because the model outputs surprisal as a vector over time, we calculated two summary measures per piece: i) the *mean*, calculated as the average joint surprisal over time; and the *accumulated* surprisal, calculated as the sum of all positive changes in surprisal over time.

All excerpts were normalized to 70dB using Praat and python's AudioSegment package, and the sound faded 3s in 3s out. Because participants completed the task online, we were not able to control whether auditory stimuli were presented using a headset, nor how the loudness of the excerpts was adjusted throughout the experiment. Note that, as mentioned before, we implemented multiple attentional checks to enforce task compliance and that we excluded participants who were not engaged with the task.

Results

Pleasure

Our minimal model contained *Trial Type* (*Rock/Pop* or *Elevator*), as well as BMRQ score and/or musical training as measured by the Gold-MSI. We generated two new models by adding *Order* as a fixed effect (*Elevator* first or *Rock/Pop* first) in one and by modelling the interaction between *Trial Type* and *Order* in the other. Model selection using the AIC criterion showed the following model to be the best, taking into account that there was a less than two difference in AIC between the top two models: *order * trialtype + BMRQ* (Table 1). This model was selected for subsequent analysis. Importantly, the model shows a significant interaction between Order and Trial Type (p = 0.001; See Table 2, first column). Regarding the latter, *emmeans* shows that participants rated Elevator music as pleasurable as Rock/Pop when listening to Elevator music first (p=0.87) but that they experienced more pleasure from Rock/Pop than Elevator music when they listened to Rock/Pop songs first ($p < 0.0001$; See Figure 1A). There was also a main effect of BMRQ (see Figure 2B), that is, as expected, participants with a higher musical hedonia tended to provide higher pleasure ratings.

Model: Pleasure	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
order*trialtype+BMRQ+GoldMusicTraining	8	3232.66	0.00	0.64	-1608.24
order*trialtype+BMRQ	7	3233.82	1.16	0.36	-1609.84
order+trialtype+BMRQ+GoldMusicTraining	7	3242.49	9.83	0.00	-1614.18
order+trialtype+BMRQ	6	3243.65	10.99	0.00	-1615.77
trialtype+BMRQ	5	3247.16	14.50	0.00	-1618.54
trialtype+BMRQ+GoldMusicTraining	6	3247.46	14.80	0.00	-1617.68
null	3	3261.80	29.14	0.00	-1627.89

Table 1. Candidate models for pleasure. All models included random intercepts for participants. ∗ indicates an interaction. K_i = the number of estimated parameters for model i. AIC_{ci} = corrected Akaike information criterion. $_i(AIC_c)$ = difference between AIC_c for model i and best model's AIC_c. $w_i(AIC_c)$ = the Akaike weight measuring the level of support in favor of model i being the most parsimonious among the candidate model set. $log(L_i)$ = natural logarithm of the maximum likelihood for model i.

	Pleasure			*Valence*			*Familiarity*			*Recognition*		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	-2.84	-5.76 – 0.08	0.064	-0.46	-3.01 – 2.10	0.727	-2.04	-5.95 – 1.86	0.312	0.75	-2.49 – 3.99	0.653
Trial Type	-0.20	-0.31 – -0.09	**<0.001**	0.29	0.17 – 0.41	**<0.001**	-0.91	-1.05 – -0.77	**<0.001**	-1.65	-1.82 – -1.47	**<0.001**
BMRQ	0.05	0.01 – 0.09	**0.011**	0.01	-0.02 – 0.04	0.573	0.03	-0.02 – 0.07	0.325	-0.01	-0.05 – 0.03	0.547
Order	0.32	0.06 – 0.58	**0.019**							0.35	0.06 – 0.64	**0.022**
Order*Trial Type	0.19	0.08 – 0.29	**0.001**									
GoldMusicTraining				1.46	0.25 – 2.67	**0.023**	1.64	-0.22 – 3.49	0.091			

Table 2. Summary of selected linear mixed regression models for four rating measures: valence, familiarity, recognition, and pleasure.

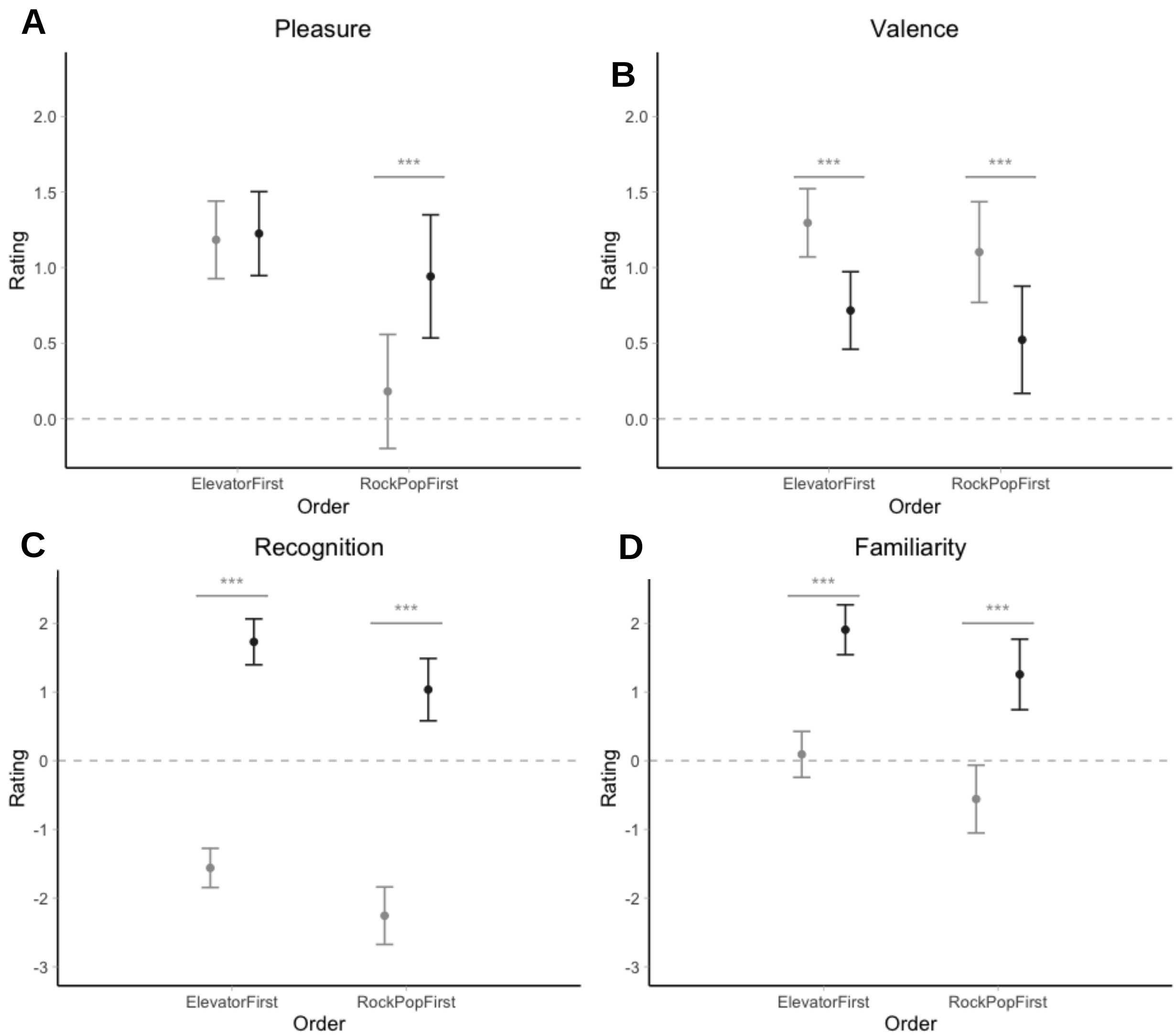


Figure 1. Results for the best models showing an interaction between Order and Trial Type for pleasure **(A)**, a main effect of Trial Type for valence **(B)**, a main effect of order and Trial Type for recognition **(C)**, and a main effect of Trial Type for familiarity **(D)**. Predictions for rock/pop music are shown in dark grey, and predictions for classical music are shown in light grey. Shaded areas indicate 95% confidence intervals.

Valence

Model selection showed no effect of order on valence ratings, with the minimal model emerging as the best candidate to explain the variance: *trialtype* + *BMRQ* + *GoldMusicTraining* (Table 3). Model predictions show that Elevator music was rated as having higher valence than Rock/Pop music($p < 0.001$; see Table 2, second column, and Figure 1B). The results also show a positive relationship between Gold-MSI musical training and valence ratings regardless of Trial Type ($p = 0.02$; Figure 2A). No significant effects for BMRQ were found. This suggests that that musical expertise, but not sensitivity to musical reward, selectively modulated valence ratings.

Model: Valence	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	3434.76	0.00	0.49	-1711.33
order+trialtype+BMRQ+GoldMusicTraining	7	3436.12	1.36	0.25	-1710.99
trialtype+BMRQ	5	3437.95	3.19	0.10	-1713.94
order*trialtype+BMRQ+GoldMusicTraining	8	3438.00	3.25	0.10	-1710.92
order+trialtype+BMRQ	6	3439.78	5.02	0.04	-1713.84
order*trialtype+BMRQ	7	3441.66	6.90	0.02	-1713.76
null	3	3456.49	21.74	0.00	-1725.23

Table 3. Candidate models for valence.

Familiarity

Model selection showed the best candidate model to be *trialtype* + *BMRQ* + *GoldMusicTraining,* with a main effect of *Trial Type* (Table 4; Table 2, fourth column). Model predictions show that ratings of familiarity were higher for Rock/Pop than Elevator music ($p < 0.0001$; Figure 1D). Familiarity ratings were related to Gold-MSI music training, with overall higher familiarity ratings for higher musical training (Figure 2A).

Model: Familiarity	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
order+trialtype+BMRQ+GoldMusicTraining	7	3714.23	0.00	0.34	-1850.05
order*trialtype+BMRQ+GoldMusicTraining	8	3715.42	1.20	0.18	-1849.62
trialtype+BMRQ+GoldMusicTraining	6	3715.55	1.33	0.17	-1851.73
order+trialtype+BMRQ	6	3716.19	1.96	0.13	-1852.04
trialtype+BMRQ	5	3716.41	2.19	0.11	-1853.17
order*trialtype+BMRQ	7	3717.38	3.15	0.07	-1851.62
Null	3	3857.04	142.82	0.00	-1925.51

Table 4. Candidate models for familiarity.

Recognition

Model selection showed the best candidate model to be *trialtype* + *order* + *BMRQ,* with a main effect of order (Table 5). Model predictions show that recognition ratings were higher for Rock/Pop music than Elevator music, with both ratings of familiarity higher when Elevator music was presented first ($p < 0.0001$; See Table 2, third column). Ratings of recognition were not modulated by Gold-MSI music training.

Model: Recognition	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
order*trialtype+BMRQ+GoldMusicTraining	8	4025.93	0.00	0.30	-2004.88
order*trialtype+BMRQ	7	4026.10	0.17	0.27	-2005.98
order+trialtype+BMRQ+GoldMusicTraining	7	4026.75	0.83	0.20	-2006.31
order+trialtype+BMRQ	6	4026.93	1.00	0.18	-2007.41
trialtype+BMRQ	5	4030.18	4.25	0.04	-2010.05
trialtype+BMRQ+GoldMusicTraining	6	4031.09	5.17	0.02	-2009.50
null	3	4311.95	286.02	0.00	-2152.96

Table 5. Candidate models for recognition.

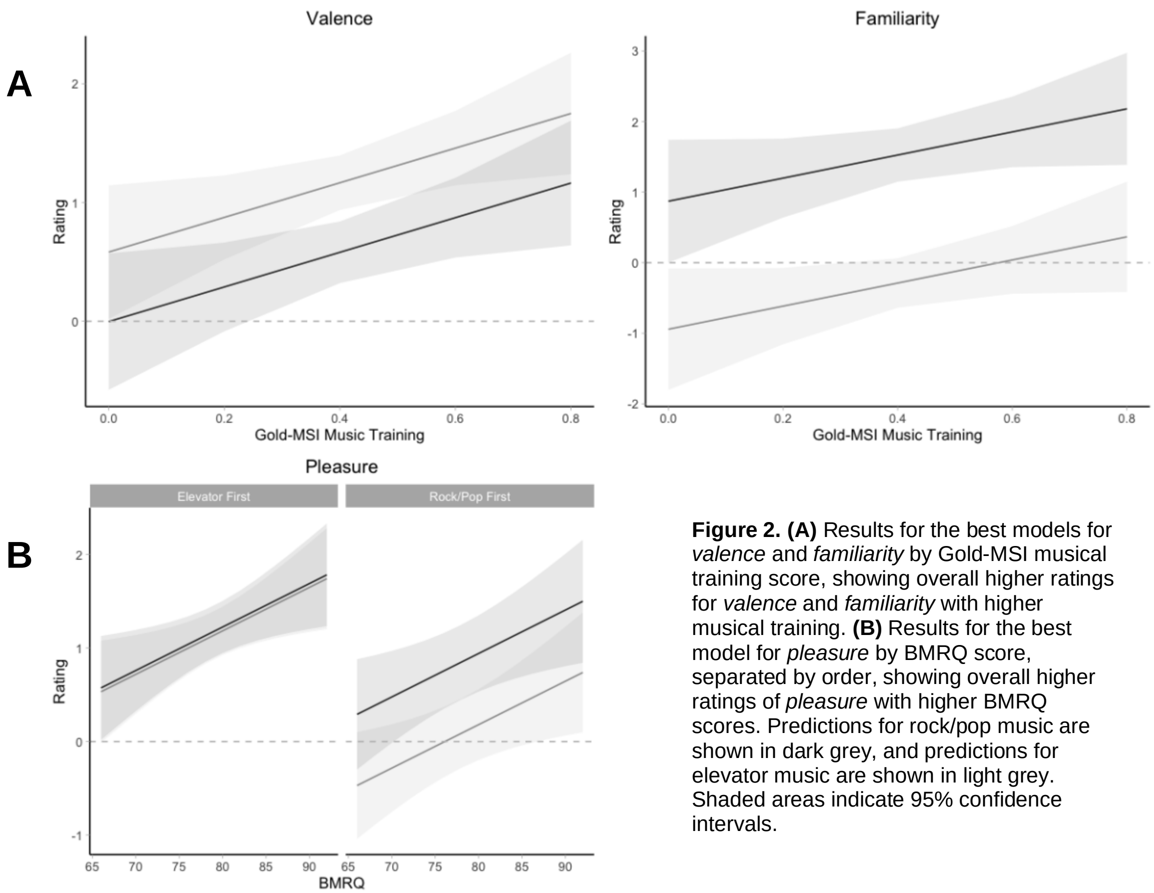


Figure 2. (A) Results for the best models for *valence* and *familiarity* by Gold-MSI musical training score, showing overall higher ratings for *valence* and *familiarity* with higher musical training. **(B)** Results for the best model for *pleasure* by BMRQ score, separated by order, showing overall higher ratings of *pleasure* with higher BMRQ scores. Predictions for rock/pop music are shown in dark grey, and predictions for elevator music are shown in light grey. Shaded areas indicate 95% confidence intervals.

Discussion for Experiment 1

The results of Experiment 1 suggest that the pleasure responses for the songs of the Reward condition were stable regardless of the context of the other music (i.e., the order), while the pleasure reported for elevator music varied. It is thus possible that participants may have defined their internal scale for rating pleasure based on the first block. That is, the baseline pleasure of hearing neutral-sounding music may have led participants to rate elevator music as more pleasurable upon hearing it first. Interestingly, there were no between group differences in mean pleasure ratings for rock/pop music, suggesting that, in general, participants provide more consistent ratings for well-known musical pieces.

Our valence analysis showed a main effect of musical training, with higher valence ratings associated with higher musical training. We attribute this effect to musicians' more complex relationship between musical elements and positive or negative value. Musicians may have a more developed set of representations connecting structural aspects of music with positive or negative valence. In a similar vein, those with little to no musical training may only attribute large or positive *lyrical* content with higher valence, as opposed to musical elements.

The main effect of order for our recognition analysis could also be explained by participants setting the baseline of recognition using the first genre they listen to: if elevator music is first, the participant is not aware that they may hear songs that they have heard before, so they may set the baseline higher. Lastly, it is possible that those with higher musical expertise have greater familiarity across genres, but do not necessarily recognize rock/pop music more than those with lower musical training.

The only measure which showed a main effect of BMRQ was pleasure, which validates previous results showing that that individual sensitivity to musical reward modulates ratings of musical pleasure (Mas-Herrero et al., 2014). Our results strengthen the specificity of this relationship, showing that individual musical hedonia does not systematically influence measures of valence, familiarity, or recognition.

Experiment 2

Experimental Design & Participants

Due to the order effects in Experiment 1, the *Reward* and *Elevator* stimuli were presented in a non-blocked, randomized design in Experiment 2. Additionally, it is possible that the difference in pleasure ratings may have been due to the presence of lyrics in rock/pop and absence of lyrics in elevator music. Thus, we created a second version of the experiment in which the Reward condition was composed of classical music with no lyrics instead of rock/pop songs. Therefore, Experiment 2 was completed by two different groups of participants who were presented with the same elevator music stimuli but listened to either classical music or rock/pop songs. The experimental procedure was exactly the same as in Experiment 1, with an additional question about musical preference which categorized participants as generally preferring or not preferring the genre of music opposite elevator music ("What kinds of music do you listen to? Select ONLY the ones you listen to on a regular basis."). This more precise measure of preference allowed us to add preference (preferred or not preferred) for the reward musical stimuli as a fixed factor in the computed models.

Stimuli

The same elevator stimuli were used in Experiment 2 as in Experiment 1 (Supplementary Table 2). Classical music stimuli were taken from a list of 82 one-minute long musical excerpts rated for by 65 participants for pleasantness in a previous study (Martinez-Molina et al., 2016). The top ten excerpts were selected. All new excerpts were normalized to 70dB using Praat and python's AudioSegment package, and the sound faded 3s in 3s out.

Participants

For the Rock/Pop version of the task, we collected data from a total of 68 participants, of which 11 were excluded, for a final total of 57 participants (22 females, mean age = 37.81 ± 9.63). Average BMRQ score was 76.15 ± 12.27, average musical training score as measured by the Gold-MSI was 0.39 ± 0.2. Regarding preference, 10 participants did not list pop or rock music as a preferred genre, and 47 listed pop or rock music as a preferred genre. Given the low number of participants reporting rock/pop as a preferred genre, this factor was not included in the model for this group.

For the Classical version of the task, we collected data from a total of 164 participants, of which 38 were excluded, for a final total of 126 participants (56 females, 1 non-binary, mean age = 38.39 ± 10.01 years). Average BMRQ score was 81.59 ± 8.71, and average musical training score as measured by the Gold-MSI was 0.42 ± 0.2. Regarding preference, 62 participants did not list classical music as a preferred genre, and 64 participants listed classical music as a preferred genre.

Results: Rock/pop Music

Our minimal models contained Trial Type (Rock/Pop or Elevator), as well as just BMRQ score or BMRQ and musical training as measured by the Gold-MSI.

Pleasure

Model selection using AIC showed the following model to be the best: *trialtype + BMRQ* (Table 6). This model was selected for subsequent analysis (Table 7, first column). The model shows a main effect of *Trial Type* ($p < 0.001$), with Rock/Pop rated more highly pleasurable than Elevator music, and no significant effects of *BMRQ* (Figure 3A).

Model: Pleasure	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$\log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	4562.80	0.00	0.58	-2275.36
trialtype+BMRQ	5	4563.41	0.61	0.42	-2276.68
null	3	4632.94	70.14	0.00	-2313.46

Table 6. Candidate models for pleasure.

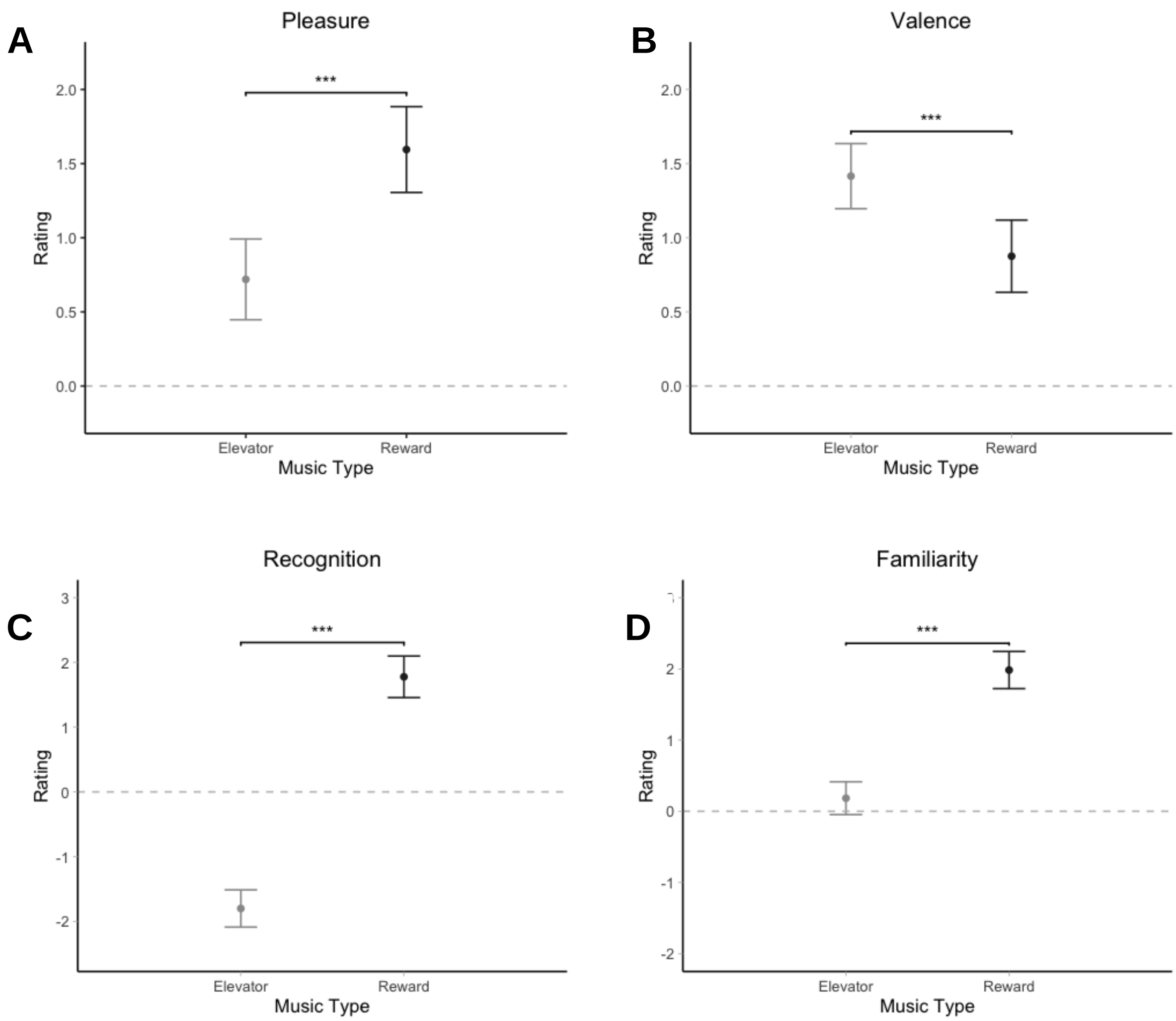


Figure 3. Results for the best models showing a main effect of Trial Type for pleasure **(A)**, valence **(B)**, recognition **(C)**, and familiarity **(D)**. Predictions for rock/pop music are shown in dark grey, and predictions for elevator music are shown in light grey. Shaded areas indicate 95% confidence intervals.

	Pleasure			*Valence*			*Familiarity*			*Recognition*		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
(Intercept)	1.10	-1.52 – 3.71	0.415	1.01	-1.08 – 3.10	0.346	-0.01	-2.26 – 2.24	0.994	-0.79	-3.59 – 2.02	0.585
Trial Type	-0.44	-0.54 – -0.34	**<0.001**	0.28	0.17 – 0.38	**<0.001**	-0.93	-1.05 – -0.81	**<0.001**	-1.81	-1.95 – -1.66	**<0.001**
BMRQ	-0.00	-0.04 – 0.03	0.789	0.00	-0.02 – 0.03	0.899	0.00	-0.03 – 0.03	0.825	-0.00	-0.04 – 0.03	0.878
GoldMusicTraining	1.18	-0.27 – 2.63	0.116				2.18	0.94 – 3.43	**0.001**	2.66	1.11 – 4.21	**0.001**

Table 7. Summary of selected linear mixed regression models for four rating measures: *pleasure, valence, familiarity,* and *recognition*.

Valence

The Akaike weight showed equal weights for each minimal model (0.5), so the model with fewer fixed factors was selected: *trialtype + BMRQ* (Table 8). This model shows a main effect of *Trial Type*, with elevator music rated as having a higher valence than the reward condition ($p < 0.001$). This effect was not modulated by BMRQ.

Model: Valence	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	4684.49	0.00	0.50	-2336.21
trialtype+BMRQ	5	4684.52	0.03	0.50	-2337.23
null	3	4706.64	22.15	0.00	-2350.31

Table 8. Candidate models for valence.

Familiarity

Model selection showed the best model to be the minimal model including both BMRQ and musical training: *trialtype + BMRQ + GoldMusicTraining* (Table 9). Subsequent analysis showed a main effect of *Trial Type* ($p < 0.001$; Table 7, third column) and GoldMusicTraining ($p = 0.001$). Plotting mode predictions showed participants rated rock/pop music as more highly familiar than elevator music (Figure 3D). Familiarity ratings were related to Gold-MSI music training, with overall higher familiarity ratings with higher musical training (Figure 4A).

Model: Familiarity	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	5010.90	0.00	0.99	-2499.41
trialtype+BMRQ	5	5020.12	9.22	0.01	-2505.04
null	3	5221.47	210.57	0.00	-2607.72

Table 9. Candidate models for familiarity.

Recognition

Model selection showed the best model to be the minimal model including both BMRQ and musical training: *trialtype + BMRQ + GoldMusicTraining* (Table 10). Subsequent analysis showed a main effect of *Trial Type* ($p < 0.001$; Table 7, fourth column) and GoldMusicTraining (($p = 0.001$). Plotting model predictions showed recognition ratings were higher for rock/pop music than elevator music (Figure 3C). Recognition ratings were also related to Gold-MSI music training, with overall higher recognition ratings with higher musical training (Figure 4B).

Model: Recognition	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	5375.62	0.00	0.99	-2681.78

trialtype+BMRQ	5	5384.44	8.81	0.01	-2687.19
null	3	5868.79	493.17	0.00	-2931.39

Table 10. Candidate models for recognition.

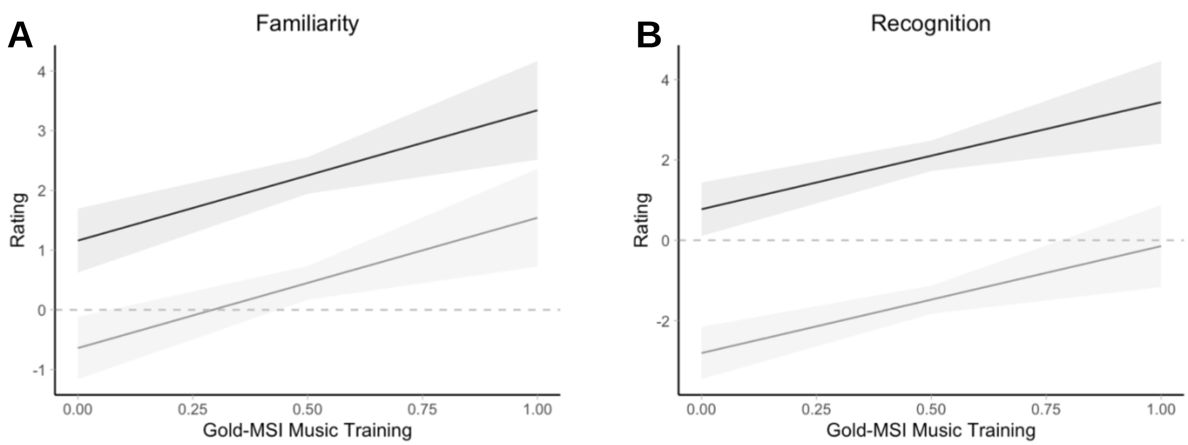


Figure 4. Results for the best models of *familiarity* **(A)** and *recognition* **(B)** plotted by Gold-MSI musical training score, showing that for both measures higher musical training predicted higher ratings. Predictions for classical music are shown in dark grey, and predictions for elevator music are shown in light grey. Shaded areas indicate 95% confidence intervals.

Results: Classical Music

Our minimal models contained Trial Type (classical or elevator), as well as BMRQ score or BMRQ and musical training as measured by the Gold-MSI. Because we had roughly equal participants with a preference or lack of preference for the Reward condition (classical music), we were able to add preference into our models as a fixed factor.

Pleasure

Model selection using AIC showed the following model to be the best: *preference * trialtype + BMRQ* (Table 11). This model was selected for subsequent analysis. The model shows a main effect of BMRQ ($p = 0.023$), and an interaction between preference and Trial Type ($p < 0.001$; See Table 12, first column). Plotting model predictions for pleasure separated by preference and Trial Type shows, as expected, that participants rated Classical music as more highly pleasurable if they had a preference for classical music (Figure 5A).

Model: Pleasure	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
preference*trialtype+BMRQ+GoldMusicTraining	8	9345.43	0.00	0.56	-4664.69
preference*trialtype+BMRQ	7	9345.94	0.51	0.44	-4665.95
preference+trialtype+BMRQ+GoldMusicTraining	7	9360.80	15.37	0.00	-4673.38
preference+trialtype+BMRQ	6	9361.31	15.87	0.00	-4674.64
trialtype+BMRQ+GoldMusicTraining	6	9361.86	16.43	0.00	-4674.91

trialtype+BMRQ	5	9365.70	20.27	0.00	-4677.84
null	3	9446.04	100.61	0.00	-4720.02

Table 11. Candidate models for pleasure.

Importantly, a closer investigation using emmeans showed that there was still a significant difference between conditions for participants with no preference for classical music, with higher pleasure ratings for Classical music than for Elevator music ($p = 0.002$), even if this difference was smaller than that of participants who had a preference for classical music. Additionally, there was no difference between Elevator ratings across preference types (p=0.31). As expected, participants provided higher pleasure ratings for Classical music when it was their preferred musical genre as compared when it was not ($p = 0.0002$).

	Pleasure			*Valence*			*Familiarity*			*Recognition*		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	-0.72	-2.24 – 0.80	0.353	-0.42	-1.93 – 1.10	0.591	1.68	-0.37 – 3.74	0.111	1.45	-0.73 – 3.63	0.196
Preference	-0.16	-0.33 – 0.02	0.080									
Trial Type	-0.25	-0.31 – -0.20	**<0.001**	0.40	0.33 – 0.46	**<0.001**	-0.77	-0.85 – -0.69	**<0.001**	-1.10	-1.20 – -1.01	**<0.001**
BMRQ	0.02	0.00 – 0.04	**0.023**	0.02	0.00 – 0.04	**0.026**	-0.02	-0.05 – 0.00	0.089	-0.03	-0.06 – 0.00	**0.030**
GoldMusicTraining	0.75	-0.17 – 1.67	0.113				2.52	1.37 – 3.67	**<0.001**	2.66	1.44 – 3.88	**<0.001**
Preference*Trial Type	0.12	0.06 – 0.18	**<0.001**									

Table 12. Summary of selected linear mixed regression models for four rating measures: pleasure, valence, familiarity, and recognition.

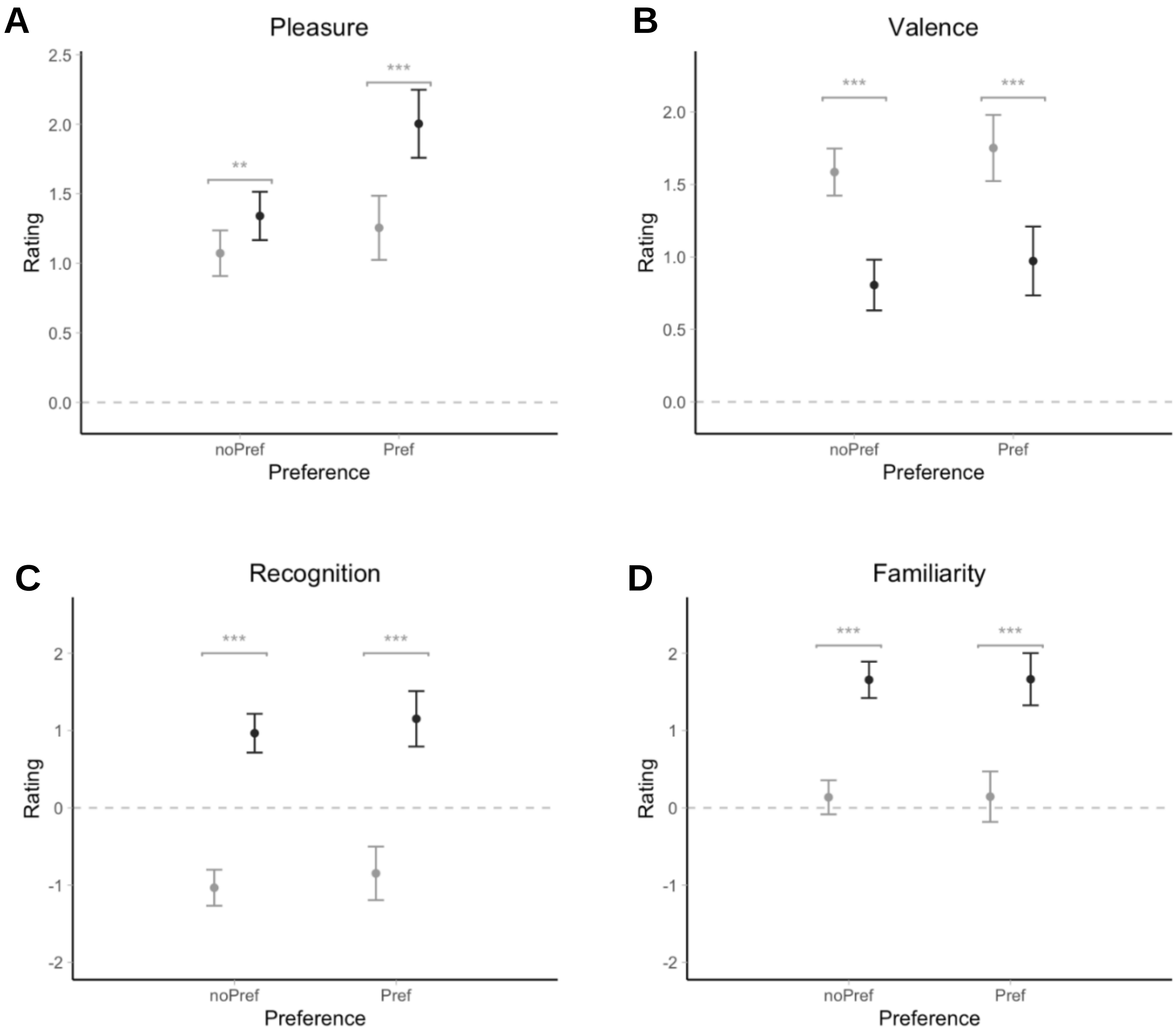


Figure 5. Predictions for each model. For legibility, all models are plotted with *preference* for Elevator (light grey) vs. Classical (dark grey). Model predictions show an interaction between Trial Type and preference for pleasure **(A)**, a main effect of Trial Type for valence **(B)**, a main effect of Trial Type for recognition **(C)**, and a main effect of Trial Type for familiarity **(D)**. Predictions for classical music are shown in dark grey, and predictions for elevator music are shown in light grey. Error bars indicate 95% confidence intervals.

Model predictions also showed a replication of Experiment 1's results, that is, pleasure ratings were related to the participant's level of sensitivity to musical reward (Figure 6A).

Valence

Model selection showed no effect of preference on valence ratings, with the minimal model (without Gold-MSI musical training) emerging as the best candidate to explain the variance: *trialtype + BMRQ* (Table 13). Model predictions show that Elevator music was rated as having higher valence than Classical music ($p < 0.001$; See Table 12, second column; Figure 5B). There was also an effect of BMRQ ($p = 0.026$), which is a departure from our previous model of valence ratings in Rock/Pop excerpts, as BMRQ emerges as a main effect as opposed to Gold-MSI (Figure 6A).

Model: Valence	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ	5	9984.52	0.00	0.26	-4987.25
preference*trialtype+BMRQ	7	9984.62	0.09	0.25	-4985.29
preference+trialtype+BMRQ	6	9985.50	0.98	0.16	-4986.73
trialtype+BMRQ+GoldMusicTraining	6	9985.64	1.11	0.15	-4986.80
preference*trialtype+BMRQ+GoldMusicTraining	8	9986.27	1.74	0.11	-4985.11
preference+trialtype+BMRQ+GoldMusicTraining	7	9987.15	2.63	0.07	-4986.55
null	3	10124.91	140.39	0.00	-5059.45

Table 13. Candidate models for valence.

Familiarity

Model selection showed the best candidate model to be *trialtype* + *BMRQ* + *GoldMusicTraining* (Table 14), with a main effect of Trial Type ($p < 0.001$; See Table 12, third column) and GoldMusicTraining ($p < 0.001$). Plotting model predictions showed that Classical music was rated as more familiar than Elevator music (Figure 5D). Familiarity ratings were modulated by Gold-MSI music training, with overall higher familiarity ratings with higher musical training (Figure 6B).

Model: Familiarity	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	11162.27	0.00	0.45	-5575.12
preference*trialtype+BMRQ+GoldMusicTraining	8	11162.61	0.35	0.38	-5573.28
preference+trialtype+BMRQ+GoldMusicTraining	7	11164.28	2.01	0.17	-5575.12
preference*trialtype+BMRQ	7	11175.33	13.06	0.00	-5580.64
preference+trialtype+BMRQ	6	11177.00	14.73	0.00	-5582.48
trialtype+BMRQ	5	11177.55	15.28	0.00	-5583.76
null	3	11486.68	324.41	0.00	-5740.33

Table 14. Candidate models for familiarity.

Recognition

Model selection showed the best candidate model to be *trialtype* + *BMRQ* + *GoldMusicTraining* (Table 15) with main effects of trialtype ($p < 0.001$), BMRQ ($p = 0.03$; See Table 12, fourth column), and GoldMusicTraining ($p < 0.001$). Model predictions sho wed higher recognition ratings for Classical music than Elevator music (Figure 5C). This was modulated by BMRQ, with lower recognition ratings predicted for those with a higher sensitivity to musical reward (Figure 6A). Additionally, higher musical training predicted higher recognition ratings (Figure 6B).

Model: Recognition	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
trialtype+BMRQ+GoldMusicTraining	6	11726.12	0.00	0.42	-5857.04
preference*trialtype+BMRQ+GoldMusicTraining	8	11726.37	0.25	0.37	-5855.16
preference+trialtype+BMRQ+GoldMusicTraining	7	11727.59	1.48	0.20	-5856.77
preference*trialtype+BMRQ	7	11737.12	11.00	0.00	-5861.54
preference+trialtype+BMRQ	6	11738.34	12.23	0.00	-5863.16
trialtype+BMRQ	5	11741.21	15.09	0.00	-5865.59
null	3	12233.42	507.30	0.00	-6113.70

Table 15. Candidate models for recognition.

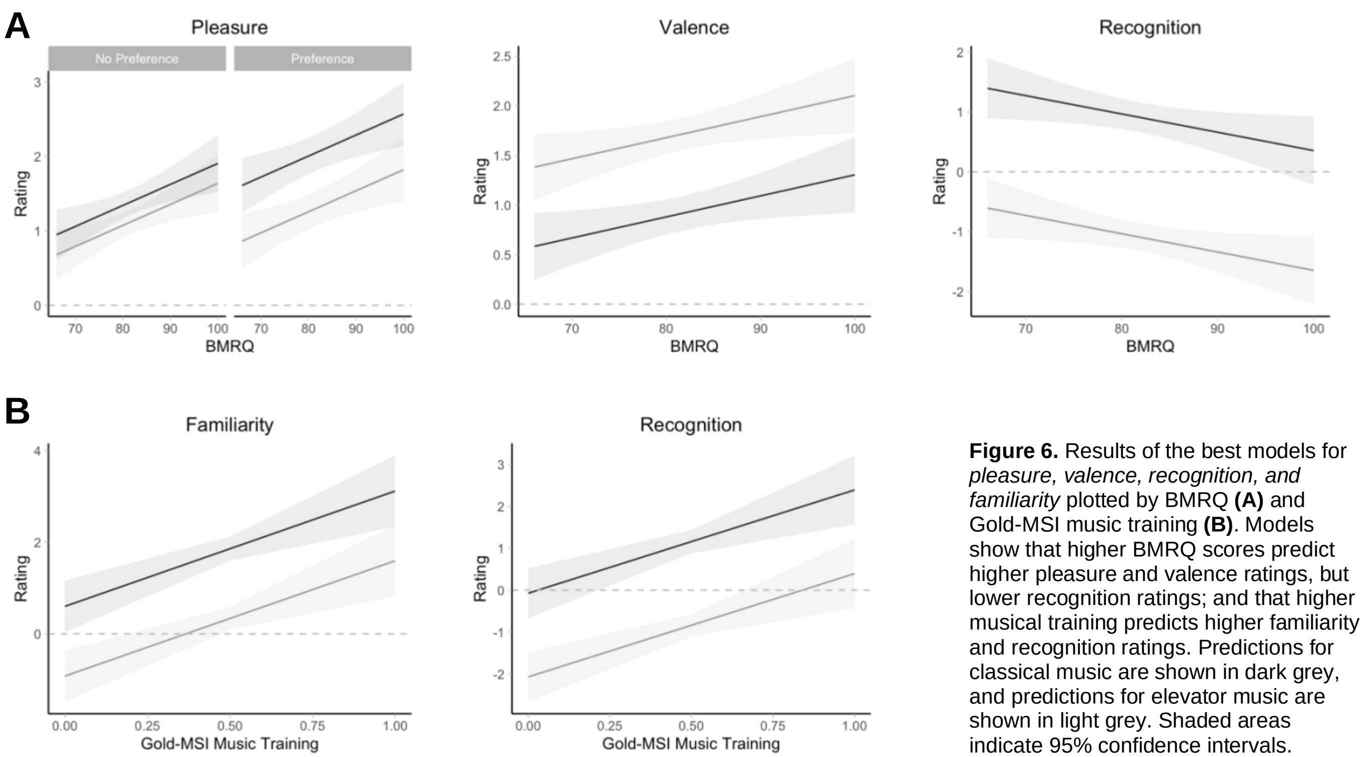


Figure 6. Results of the best models for *pleasure, valence, recognition, and familiarity* plotted by BMRQ **(A)** and Gold-MSI music training **(B)**. Models show that higher BMRQ scores predict higher pleasure and valence ratings, but lower recognition ratings; and that higher musical training predicts higher familiarity and recognition ratings. Predictions for classical music are shown in dark grey, and predictions for elevator music are shown in light grey. Shaded areas indicate 95% confidence intervals.

Discussion for Experiment 2

The results of Experiment 2 show that participants report elevator music as being less pleasurable even in comparison with genres for which participants do not have a preference (see Figure 5, showing that participants who do not like classical music still like it more than the elevator one). In addition, mean pleasure ratings for elevator music are close to 0 (Pop/Rock group: M = 0.72 ± 0.35, Preferred Classical group: M =1.31 ± 0.28, Non-Preferred Classical group: M = 1.10 ± 0.39) suggesting that participants rate elevator music very close to a “neutral” level of pleasure.

Another interesting result is the main effect of BMRQ on valence ratings as opposed to the main effect of musical training on valence ratings in the previous experiment. This may reflect the differing genres of reward music, and presence or lack of lyrics. In the rock/pop condition, participants with low musical training may be more attuned to the lyrics of each piece, which may (contrary to the musical aspects of the piece) have a lower valence in content (i.e. about heartbreak, loss, anger). Those with higher musical training may be more focused on the musical, non-lyrical, aspects of the excerpts which tend to be upbeat and catchy, thus having a higher valence. In the case of classical music, all pieces lacked lyrics and had a more complex structure. Thus, perhaps a higher sensitivity to music reward was necessary to provide an increased valence rating, as this would have necessitated an emotional response to the piece.

Musical training had a main effect for both familiarity and recognition, with more musical training corresponding to higher ratings. This is presumably due to those with musical expertise having been exposed to more music generally, thus leading to higher familiarity across genres. We also found that BMRQ had a main effect for recognition, with BMRQ scores corresponding to lower ratings. This may be due to those with higher sensitivity to musical reward having a more accurate representation of whether they have heard the song before. Thus, they may be more stringent in their ratings of exact recognition.

Experiment 3

Experimental Design & Participants

The motivation for Experiment 3 came from several areas of inquiry. and surprisal to tease apart different aspects in which elevator music may systematically differ from other genres of music. Due to the higher number of stimuli, forty for each condition, participants completed each version of the experiment separately, with order counterbalanced across participants. Using AMT, we were able to call back participants for the second session with a low attrition rate.

As discussed in general methods, surprisal is thought to be an objective measure which, broadly, corresponds to how unexpected a stimulus is. Given the initial purpose for elevator music, surprisal may be a crucial factor by which elevator music differs from other genres. Several computational frameworks have been developed to model musical surprisal across time, operating on both symbolic or machine-readable musical data (i.e. MIDI) and naturalistic audio signals. Thus, for Experiment 3 we were able to not only gather a large body of surprisal ratings, but also validate these against surprisal extracted from the same musical pieces using a novel computational model, Dynamic Regularity Extraction (D-REX). Using a Bayesian framework, D-REX calculates prediction error at each timepoint, resulting in a continuous output of surprisal across multiple acoustic features. To model this in behavior, participants rated surprisal continuously throughout each musical excerpt by dragging a slider (20 - 80) left and right while listening. They were told to only drag left into 1 – 20 or right into 80 – 100 if the surprisal they were experiencing fell below or surpassed their previous limits (only in extreme cases of low or high surprisal). They were then asked to provide an overall rating of surprisal on a 1-100 scale on the following page accompanying all other ratings.

For a statistical analysis of continuous surprisal ratings, we created summary measures to add to our linear mixed effects models (accumulated and mean). Accumulated surprisal was calculated by summing all *positive changes in surprisal* over time. Mean surprisal was calculated by averaging surprisal ratings at all timepoints.

Stimuli

Thirty new excerpts were added to each stimuli group of ten excerpts (Supplementary Tables 3 and 4). Additional classical music stimuli were selected from the same list from Experiment 2, which were one-minute segments rated as highly pleasurable by 65 participants in a previous study (Martinez-Molina et al., 2016). Additional elevator music stimuli were selected from other Muzak albums (The Sound Heard Round the World: Stimulus Progression Number Four, 1972; Muzak: Stimulus Progression 5, 1973; Muzak Stimulus Progression 1974, 1976; Stimulus Progression 6 – Muzak: 40 Years, More Than a Name, 1974) as well as more contemporary composed elevator music. All new excerpts were normalized to 70dB using Praat and python's AudioSegment package, and the sound faded 3s in 3s out.

Participants

For counterbalancing purposes, we published batches for both the Elevator and Classical music on AMT. Participants randomly completed one of the versions and were then called back to complete the other. Thus, participants were subject to exclusion not only for our *dprime* and BMRQ score, but also depending on whether they returned. Taking into account the established exclusion criteria, as well as attrition rate on AMT, we had a final total of 67 participants complete both versions of the experiment (29 female, mean age = 40.20 ± 11.16 years). Average BMRQ score was 80.79 ± 7.09, and average musical training score as measured by the Gold-MSI was 0.42 ± 0.2. Similarly to the distribution of musical genre preferences in Experiment 2, 29 of participants (43%) had a preference for classical music, and 38 did not. Thus, we again were able to analyze the modulatory role of preference in our data.

Results

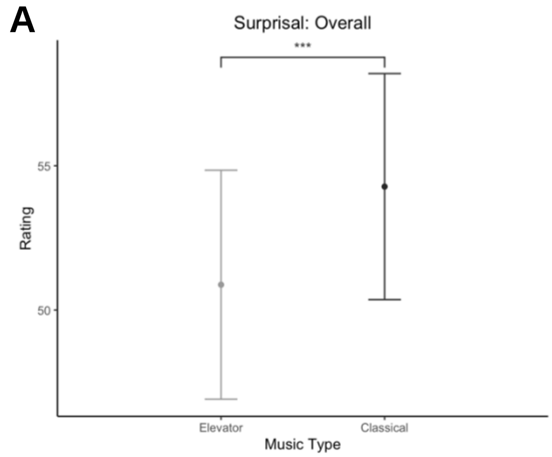
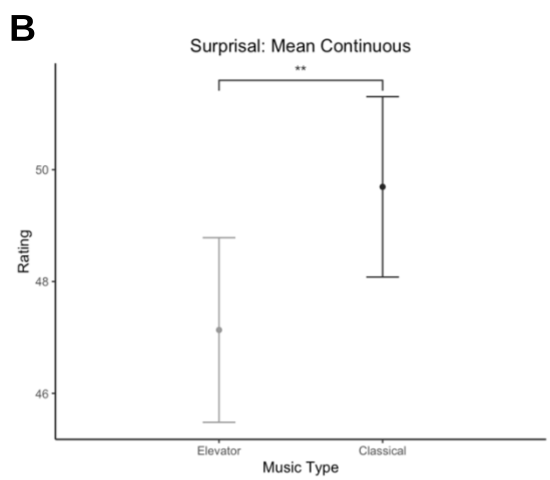
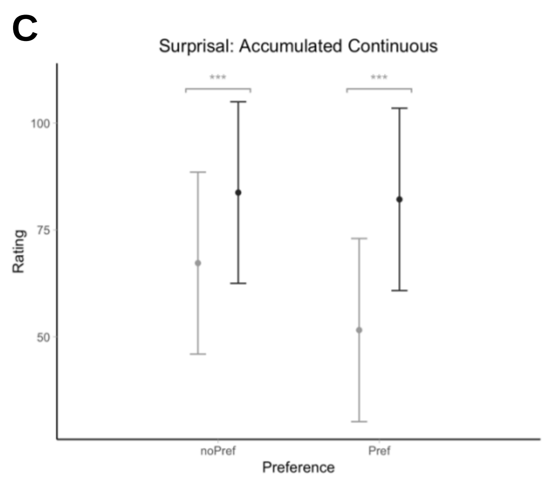


Figure 7. Results for the best models showing a main effect of Music Type for overall surprisal **(A)**, a main effect of Music Type for mean continuous surprisal **(B)**, an interaction between preference and Music Type for accumulated continuous surprisal **(C)**. Predictions for classical music are shown in dark grey, and predictions for elevator music are shown in light grey. Error bars indicate 95% confidence intervals.





Our minimal models in this case included both BMRQ and Gold-MSI musical training. Given that surprisal is a higher-level musical concept, as hypothesized that musical training would play a role in modulating surprisal ratings.

	Accumulated Continuous			*Overall*			*Mean Continuous*		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	254.45	191.41 – 317.50	**<0.001**	41.58	19.02 – 64.13	**<0.001**	35.62	23.86 – 47.38	**<0.001**
Preference	4.32	1.02 – 7.62	**0.010**						
MusicType	7.33	5.75 – 8.92	**<0.001**	1.77	1.14 – 2.40	**<0.001**	1.34	0.98 – 1.71	**<0.001**
BMRQ	-1.42	-2.14 – -0.71	**<0.001**	0.26	-0.01 – 0.53	0.063	0.19	0.04 – 0.33	**0.011**
GoldMusicTraining	-165.20	-207.32 – -123.09	**<0.001**	-23.77	-37.97 – -9.58	**0.001**	-6.06	-12.80 – 0.68	0.081

Preference*MusicType	-3.52	-5.25 – -1.79	**<0.001**						

Table 16. Summary of selected linear mixed regression models for three surprisal measures: *accumulated continuous, overall, and mean continuous.*

Surprisal: overall ratings

The minimal model contained Music Type (classical or elevator), BMRQ, and Gold-MSI musical training. We added in order and preference and created models which tested for all possible interactions. Model selection using AIC showed the minimal model as the best candidate model: *Music Type + BMRQ + GoldMusicTraining* (Table 17)*,* with a main effect of Music Type ($p < 0.001$; See Table 16, second column) and musical training ($p = 0.001$), and a trending main effect of BMRQ ($p = 0.06$). Plotting model predictions shows higher predicted surprisal ratings for classical and elevator music (Figure 7A). Overall surprisal ratings were modulated by musical training, with higher musical training corresponding to lower surprisal ratings (Figure 8A). The modulatory effect of BMRQ on overall surprisal ratings was trending towards significance, with the opposite relationship: higher sensitivity to musical reward was associated with higher surprisal ratings.

Model: Overall Surprisal	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
MusicType+BMRQ+GoldMusicTraining	6	49039.60	0.00	0.31	-24513.79
preference+MusicType+BMRQ+GoldMusicTraining	7	49039.70	0.10	0.29	-24512.84
order+preference+MusicType+BMRQ+GoldMusicTraining	8	49041.16	1.57	0.14	-24512.57
preference*MusicType+BMRQ+GoldMusicTraining	8	49041.34	1.74	0.13	-24512.66
order*preference+MusicType+BMRQ+GoldMusicTraining	9	49042.82	3.22	0.06	-24512.39
order+preference*MusicType+BMRQ+GoldMusicTraining	9	49042.91	3.31	0.06	-24512.44
order* preference *MusicType+BMRQ+GoldMusicTraining	12	49046.77	7.17	0.01	-24511.36
null	3	49071.61	32.01	0.00	-24532.80

Table 17. Candidate models for overall surprisal.

Surprisal: continuous ratings

The same models were fit, this time predicting the first summary measure of continuous surprisal ratings: Mean Surprisal. The model which emerged as the best candidate model was the minimal model: *Music Type + BMRQ + GoldMusicTraining* (Table 18); this model was selected for further analysis, with a main effect of Music Type ($p < 0.001$; See Table 16, third column) and BMRQ ($p = 0.01$). Plotting model predictions showed that mean surprisal was higher for classical music than elevator (Figure 7B). Model predictions by BMRQ score showed that higher BMRQ score predicted higher mean surprisal ratings (Figure 8C).

Model: Mean Continuous Surprisal	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
order+preference*MusicType+BMRQ+GoldMusicTraining	9	30659.91	0.00	0.31	-15320.93
MusicType+BMRQ+GoldMusicTraining	6	30659.93	0.02	0.30	-15323.95
preference+MusicType+BMRQ+GoldMusicTraining	7	30661.12	1.21	0.17	-15323.54
order+preference+MusicType+BMRQ+GoldMusicTraining	8	30662.70	2.79	0.08	-15323.33
preference*MusicType+BMRQ+GoldMusicTraining	8	30662.90	2.99	0.07	-15323.43
order* preference *MusicType+BMRQ+GoldMusicTraining	12	30663.64	3.73	0.05	-15319.78
order*preference+MusicType+BMRQ+GoldMusicTraining	9	30664.49	4.58	0.03	-15323.22
null	3	30708.74	48.83	0.00	-15351.36

Table 18. Candidate models for mean continuous surprisal.

Interestingly, when using accumulated positive surprisal as a summary measure of the continuous ratings, preference played a role in listeners' responses. The model which emerged as the best candidate model was *preference * Music Type + BMRQ + GoldMusicTraining* (Table 19); this model was selected for further analysis (See Table 16, first column). Plotting model predictions showed that accumulated surprisal was higher for classical music than elevator in both preference conditions ($p < 0.0001$, confirmed using *emmeans;* Figure 7C). Further analysis showed that accumulated surprisal was the same for classical music regardless of preference ($p = 0.67$) but significantly lower for elevator music when participants had a preference for classical music ($p < .0001$). Both Gold-MSI musical training and BMRQ emerged as modulatory variables for accumulated surprisal ratings. Higher musical training predicting lower accumulated surprisal (Figure 8B), and higher BMRQ predicted higher accumulated surprisal (Figure 8D).

Model: Accumulated Continuous Surprisal	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
order* preference *MusicType+BMRQ+GoldMusicTraining	12	41551.68	0.00	0.62	-20763.80
preference*MusicType+BMRQ+GoldMusicTraining	8	41553.45	1.77	0.26	-20768.71
order*preference+MusicType+BMRQ+GoldMusicTraining	9	41554.93	3.24	0.12	-20768.44
preference+MusicType+BMRQ+GoldMusicTraining	7	41567.37	15.69	0.00	-20776.67
order+preference+MusicType+BMRQ+GoldMusicTraining	8	41568.80	17.11	0.00	-20776.38
order+preference*MusicType+BMRQ+GoldMusicTraining	9	41570.69	19.01	0.00	-20776.32
MusicType+BMRQ+GoldMusicTraining	6	41570.89	19.21	0.00	-20779.43
Null	3	41718.14	166.45	0.00	-20856.07

Table 19. Candidate models for accumulated continuous surprisal.

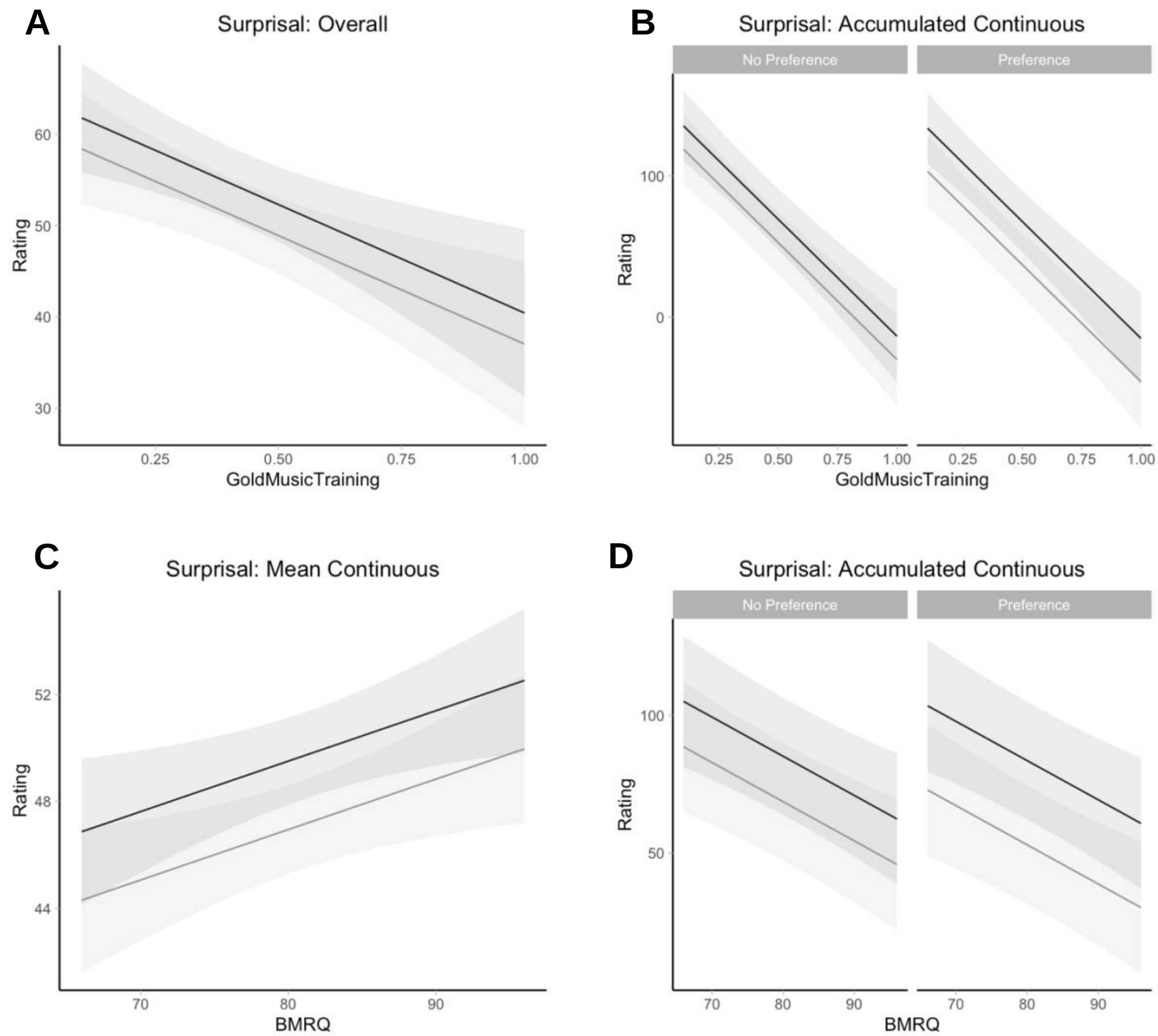


Figure 8. (A) Results for the best models for *overall surprisal* Gold-MSI musical training score, showing overall lower ratings for surprisal with higher musical training. **(B)** Results for the best model for *accumulated continuous surprisal* by Gold-MSI musical training score, showing lower accumulated ratings of surprisal with higher musical training. **(C)** Results for the best model for *mean continuous surprisal* showing higher ratings of surprisal with higher BMRQ scores. **(D)** Results for the best model for *accumulated continuous surprisal* by BMRQ score, separated by preference, showing overall lower ratings of surprisal with higher BMRQ scores.Predictions for classical music are shown in dark grey, and predictions for elevator music are shown in light grey. Shaded areas indicate 95% confidence intervals.

Surprisal: relationship between overall and continuous ratings

In order to assess the correspondence between participants' overall surprisal ratings with their continuous surprisal measures, we generated models using each summary measure for participants' continuous ratings (mean and accumulated) to predict overall surprisal ratings. Model selection showed that mean continuous surprisal is the better model to explain the variance in overall surprisal ratings, but both summary measures had a main effect in their respective models ($p < 0.001$; Table 21; Figure 9).

Model: Surprisal (Overall ~ Continuous Measures)	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$\log(L_i)$
MeanSurprisal	4	32577.94	0.00	1	-16284.97
AccumSurprisal	4	36215.88	3637.94	0	-18103.94
null	5	37846.75	5268.81	0	-18918.37

Table 20. Candidate models for overall surprisal using continuous ratings of surprisal showing that mean surprisal better explains overall surprisal ratings.

	Overall Surprisal ~ Mean			*Overall Surprisal ~ Accum*		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	-34.74	-37.41 – -32.07	**<0.001**	35.50	31.73 – 39.28	**<0.001**
MeanSurprisal	1.72	1.68 – 1.76	**<0.001**			
AccumSurprisal				0.18	0.17 – 0.20	**<0.001**

Table 21. Summary of linear mixed regression models for overall surprisal using mean continuous and accumulated continuous surprisal as fixed factors.

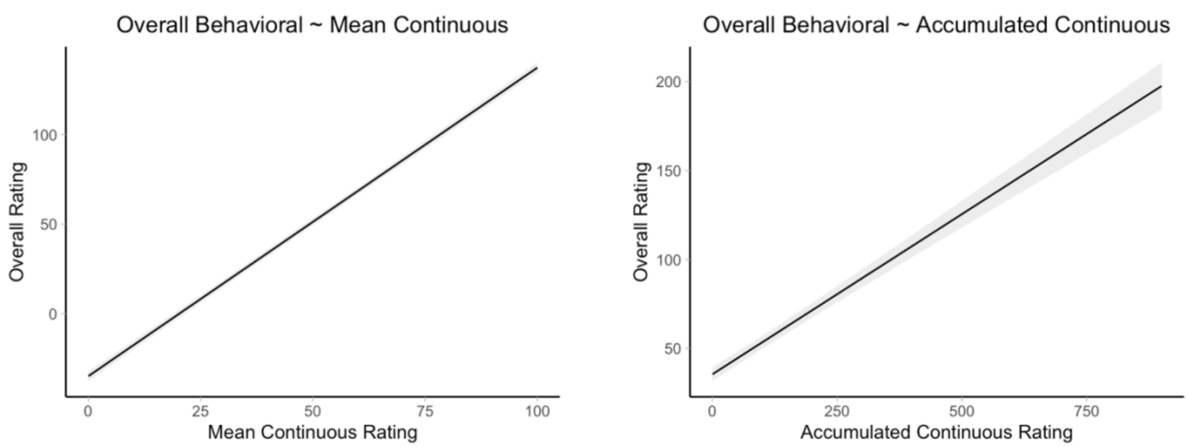


Figure 9. Results from models using participants' continuous ratings to predict their overall surprisal ratings showing both mean and accumulated surprisal significantly modulated participants' overall ratings, with mean surprisal emerging as a better candidate.

D-REX Computational Model of Surprise

Comparing D-REX across Music Type

As a final validation of our participants' continuous surprisal ratings, we compared summary statistics for D-REX's joint surprisal measures (mean and accumulated) across Music Type. Two-sample t-tests were performed to compare D-REX's output for joint mean surprisal and joint accumulated surprisal in the 40 classical music excerpts vs. the 40 elevator music excerpts. There was a significant difference in mean surprisal between Classical (M = 10.04 ± 2.84) and Elevator (M = 6.40 ± 2.74); t(78) = 5.84, p < 0.001, with Classical music excerpts having significantly higher mean surprisal (Figure 10A). There was also a significant difference in accumulated surprisal between Classical (M = 93.39

± 24.53) and Elevator (M = 81.61 ± 14.60); t(78) = 2.61, p = 0.01, with Classical music excerpts having significantly higher accumulated surprisal (Figure 10B).

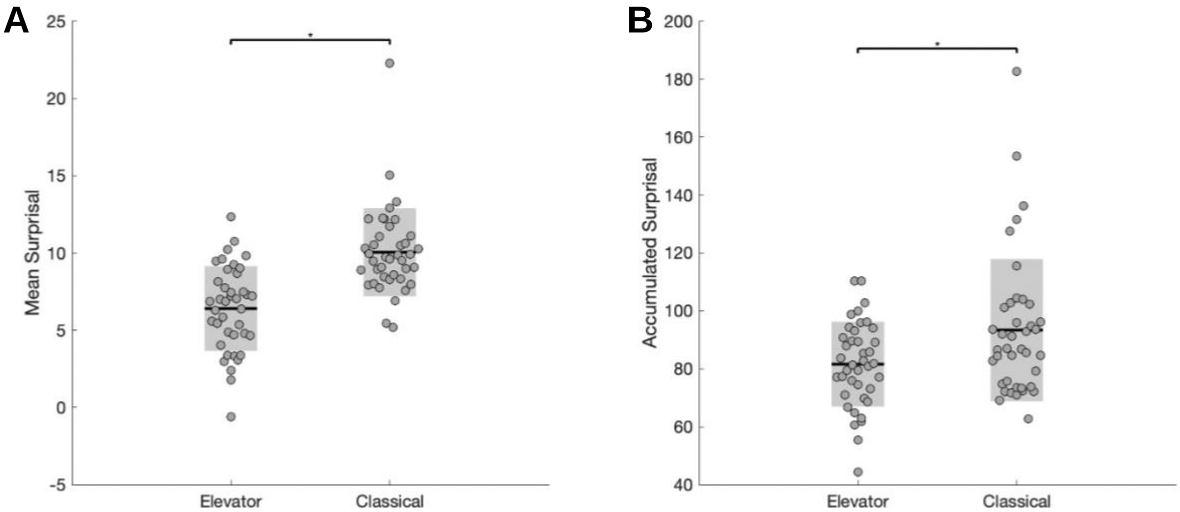


Figure 10. Bar plots showing data points, mean, and standard deviation for D-REX mean joint surprisal **(A)** and accumulated joint surprisal **(B)** for *elevator* vs. *classical* music.

Linking Behavioral and Computational measures of surprise

In order to assess the consistency between the measures of surprisal provided by the participants and those generated by the computational model, First, we generated two models to predict Overall Behavioral Surprisal, using each of the summary measures from D-REX: joint mean surprisal and joint accumulated surprisal. Model selection showed that *D-REX Accumulated* better explained the variance in the data (Table 22). Both models showed a significant relationship between behavioral and computational measures, with higher *D-REX Accumulated* or higher *D-REX Mean* predicting higher behavioral overall surprisal (p < 0.0001, p <0.0001; Table 23; Figure 11A).

Model: Surprisal (Overall ~ D-REX Summary Measures)	K_i	AIC_{ci}	$\Delta_i(AIC_c)$	$w_i(AIC_c)$	$log(L_i)$
drexAccumModel	4	49038.86	0.00	0.93	-24515.42
drexMeanModel	4	49044.03	5.17	0.07	-24518.01
emptyModel	3	49071.61	32.75	0.00	-24532.80

Table 22. Candidate models for overall surprisal using continuous ratings of surprisal showing that mean surprisal better explains overall surprisal ratings.

	Overall Surprisal ~ Mean			Overall Surprisal ~ Accum		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	48.28	44.15 – 52.41	**<0.001**	44.74	40.12 – 49.37	**<0.001**
DREXMeanSurprisal	0.52	0.33 – 0.70	**<0.001**			
DREXAccumSurprisal				0.09	0.06 – 0.12	**<0.001**

Table 23. Summary of linear mixed regression models for overall surprisal using mean continuous and accumulated continuous surprisal as fixed factors.

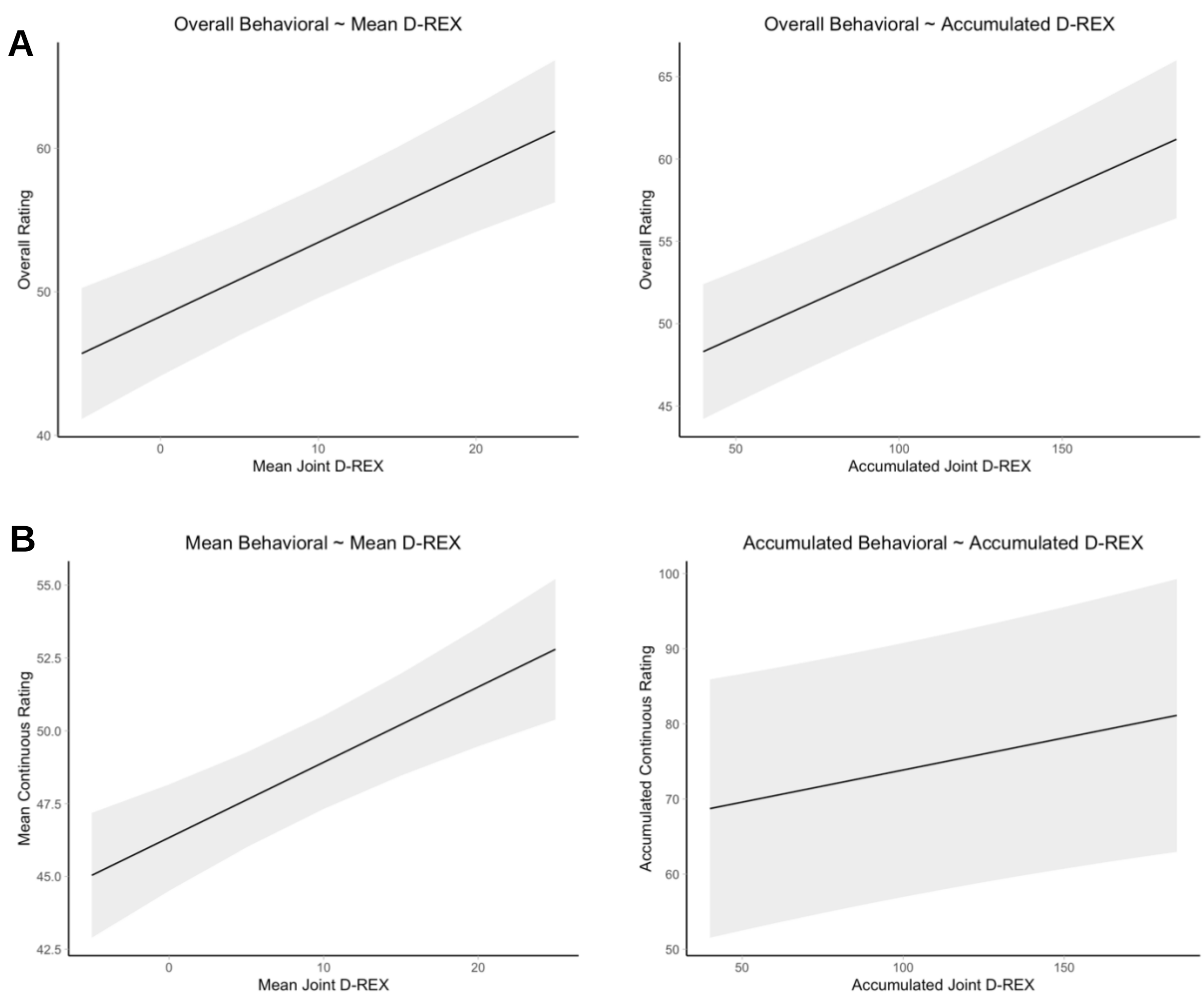


Figure 11. Results for models showing that overall surprisal is predicted by both D-REX summary measures **(A)** and that mean continuous surprisal and accumulated continuous surprisal are predicted by their respective D-REX summary measures **(B)**.

Second, we also generated models to predict the Behavioral Continuous Surprise measures provided by the participants. In particular, we tried to predict behavioral mean continuous surprisal and accumulated continuous surprisal, from *D-REX Mean Surprisal* and *D-REX Accumulated Surprisal,* respectively. each D-REX measure had a significant main effect in its respective model ($p < 0.0001$; $p = 0.01$; Table 24; Figure 11B).

	Behavioral Mean Surprisal			Behavioral Accumulated Surprisal		
Coefficient	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
Intercept	46.33	44.51 – 48.16	**<0.001**	65.30	47.40 – 83.21	**<0.001**
DREXMeanSurprisal	0.26	0.15 – 0.37	**<0.001**			
DREXAccumSurprisal				0.09	0.02 – 0.15	**0.014**

Table 24. Summary of linear mixed regression models for behavioral mean and accumulated continuous surprisal using D-REX mean and accumulated continuous surprisal as fixed factors, respectively.

Discussion for Experiment 3

The results of Experiment 3 show that participants report elevator music as being less surprising than classical music, and that this is only affected by preference for accumulated continuous surprisal (Figure 7). We found that participants' mean continuous surprisal better explained the variance in overall surprisal ratings compared to participants' accumulated surprisal ratings (Figure 9), which suggests that the mean is more related to the overall feeling of surprisal, while accumulated surprisal might be capturing more "online," "in the moment," or "dynamic" feelings of surprisal.Turning to the computational model results, we use the same measures to summarize D-REX's surprisal output over time and compare using two-sample t-tests, showing that both mean and accumulated surprisal is higher for classical music than elevator music (Figure 10).

Regarding the result showing that higher musical training predicted lower surprisal ratings (for overall and accumulated surprisal), those with musical training may have had a richer and more technically formed set of expectations, thus having lower overall prediction error (Figure 8A and 8B). On the other hand, much of the experience of pleasure in music is tied to the balance between expectation, violation, and resolution. Thus, those who are more highly sensitive to musical pleasure may be experiencing and reporting higher surprisal, as evidenced by BMRQ's role in mean continuous surprisal ratings (Figure 8D). This is in contrast with the result for accumulated surprisal, which shows that higher BMRQ predicts lower surprisal ratings.

The best model for accumulated surprisal shows a main effect of Gold-MSI but not BMRQ. Interestingly, there is a main effect BMRQ in mean surprisal, but not Gold-MSI. This suggests that the baseline/mean surprisal value is more affected by sensitivity to musical reward, while the increases and variability in ratings is more tied to musical training or expertise. Lastly, we show that there is a consistent relationship between behavioral ratings of surprisal and a new computational model of surprise (Figure 11).

General Discussion

The present project characterizes a genre which was developed with the specific intention of being emotionally neutral to listeners with a Western enculturation: elevator music. We used behavioral data collected from a large cohort of participants over three different experimental designs, as well as a new computational model of musical surprise. Model predictions from Experiments 1 and 2 show that elevator music elicited pleasure and familiarity responses within the neutral part of the scale (between -1 to 1 on a -4 to 4 scale; see Figures 1, 3, and 5). Experiment 2 also shows that people reported elevator music as being less pleasurable than other well-known genres, such as rock/pop and classical music. Finally, both computational and behavioral measures of musical surprise show that elevator music is less surprising and more predictable than other well-known genres (Figure 7 and 10).

These results confirm our hypothesis that elevator music triggers neutral reward-related responses (i.e., pleasure). Importantly, the second part of Experiment 2 allowed us to assess the role that music preference plays in modulating the experience that participants have when listening to elevator music. Participants reported higher pleasure ratings for classical music regardless of preference – the pattern of model predictions of pleasure for classical vs. elevator music was maintained, albeit attenuated, even for participants who did not indicate a preference for classical music. In other words, participants who did not like classical music still liked elevator music less. This suggests that *above and beyond individual music preference*, elevator music is rated as more emotionally neutral than other genres. Regarding familiarity, separating the question of explicit recognition from stylistic familiarity, affirmed our conjecture that though listeners would likely not have heard the exact excerpts before, they would be relatively familiar with the structural, harmonic, and melodic aspects of elevator music. Participants also consistently reported elevator music to have higher positive valence that rock/pop or classical. This implies that there is less variation in elevator music across songs that in other genres, where valence can change drastically from one musical piece to another (e.g., a rock ballad is still part of the rock genre but it usually negative valanced). Finally, model selection in both experiments confirmed that sensitivity to musical reward (as measured by the BMRQ) as well as musical training modulated ratings, but we did not find evidence to suggest that either measure affected the difference between pleasure ratings across conditions. Valence was rated on the higher end of the scale, between 1.5 and 2.

Considering why Muzak was originally created, our results provide concrete behavioral evidence supporting the functional intention behind the genre. Within the music cognition field, computational models of higher-level musical features (such as tension and surprise) are of high interest (Pearce, 2018; Farbood, 2012). The gold standard thus far has been Information Dynamics of Music (IDyOM; Pearce & Wiggins, 2012), which uses statistical learning (via a training set) to generate a continuous measure of entropy and surprise. While being an incredibly useful tool for the computational modelling of musical surprise, IDyOM is constrained to symbolic, monophonic data. To overcome these constraints, we turned to a newly validated algorithm that can be leveraged to compute musical surprise: D-REX. This algorithm uses Bayesian inference to extract statistical regularities in auditory sequences to generate predictions, resulting in continuous surprisal over time. Importantly, unlike IDyOM, D-REX can can be applied to polyphonic music and is not restricted to symbolic data. The results of Experiment 3 further validate this model as predictive of subjective ratings of surprisal, showing that D-REX summary measures of joint surprisal predict behavioral surprisal ratings (both overall and continuous) in elevator music.

Turning to cognitive neuroscience methods, the properties elucidated here depict neutrality in terms of pleasure and surprise, both of which are elements of music which typically characterize preference and activate the dopaminergic system. While there is extensive evidence showing that dopamine neurons encode reward prediction errors (involving motivational *value*), they have also been shown to respond to motivational *salience* in general, which includes aversive and novel stimuli as well as pleasurable stimuli (Matsumoto and Hikosaka, 2009; Lisman et al., 2011). Additionally, recent work shows that midbrain dopaminergic neurons respond to abstract reward as well as information *about* that abstract reward (Bromberg-Martin and Hikosaka, 2009). Given this body of research showing the complexity of the reward system's response to abstract stimuli above and beyond pleasure, elevator music seems well suited to serve as a control musical stimulus to explicitly pleasurable (and self-selected) music. This broadly speaks to the issue of whether, given individual differences in music preference, there is a universal neutral musical stimulus within the same musical enculturation. Our results suggest the possibility that even given vast differences across individuals in pleasure, preference, and the experience of reward, elevator music contains musical elements which are experienced as neutral across Western music listeners. The specific acoustic and musical features which govern this neutrality have yet to be isolated.

Limitations and future directions

Using an online platform limited our ability to control the listening experience (i.e. volume or headphones) for the participant, or confirm that they were not listening to anything else in the background. A replication in a lab-controlled environment collecting physiological responses (e.g., electrodermal activity) would help to further characterize the effect of elevator music.

Additionally, a more formal music theoretical analysis of specific elevator music excerpts would provide a missing piece of this endeavor. This could show that the functional intention is not only confirmed by listener ratings, but also in the structural and compositional aspects of the music itself.

References

Assaneo, M.F., Ripollés, P., Orpella, J., Lin, W.M., de Diego-Balaguer, R., Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat Neurosci. 22*(4), 627-632. doi: 10.1038/s41593-019-0353-z.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences, 98(20), 11818-11823. doi:10.1073/pnas.191355898

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron, 63(1), 119-126. doi:10.1016/j.neuron.2009.06.009

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Cheung, V.K.M., Harrison, P.M.C., Meyer, L., Pearce, M.T., Haynes, J. & S. Koelsch. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology, 29*, 1–9. https://doi.org/10.1016/j.cub.2019.09.067.

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. Proceedings of the National Academy of Sciences, 117(4), 1924-1934. doi:10.1073/pnas.1910704117

Cowen, A.S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, 117(4), 1924-1934. DOI: 10.1073/pnas.1910704117

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PloS one, 8(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Farbood, M. M. (2012). A Parametric, Temporal Model of Musical Tension. Music Perception, 29(4), 387-428. doi:10.1525/mp.2012.29.4.387

Ferreri, L., Mas-Herrero, E., Cardona, G., Zatorre, R.J., Antonijoan, R.M., Valle, M., Riba, J., Ripollés, P. and Rodriguez-Fornells, A. (2021), Dopamine modulations of reward-driven music memory consolidation. Ann. N.Y. Acad. Sci.. https://doi.org/10.1111/nyas.14656

Ferreri, L., Mas-Herrero, E., Zatorre, R.J., Ripolles, P., Gomez-Andres, A., Alicart, H., Olive, G., Marco- Pallares, J., Antonijoan, R.M., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *PNAS, 116*(9), 3793-3798.

Ferreri, L., Singer, N., Mcphee, M., Ripollés, P., Zatorre, R. J., & Mas-Herrero, E. (2021). Engagement in Music-Related Activities During the COVID-19 Pandemic as a Mirror of Individual Differences in Musical Reward and Coping Strategies. Frontiers in Psychology, 12. doi:10.3389/fpsyg.2021.673772

Goethem, A. van & Sloboda, J. The functions of music for affect regulation. Music. Sci. (2011) doi:10.1177/1029864911401174.

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The Song Is You. Social Psychological and Personality Science, 7(6), 597-605. doi:10.1177/1948550616641473

Hennessy, S., Sachs, M., Kaplan, J., & Habibi, A. (2021). Music and mood regulation during the early stages of the COVID-19 pandemic. Plos One, 16(10). doi:10.1371/journal.pone.0258027

Juslin, P. N. & Laukka, P. Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. J. New Music Res. 33, 217–238 (2004).

Juslin, P. N. & Vastfjall, D. Emotional responses to music: The need to consider underlying mechanisms. Behav. Brain Sci. 31, 559 (2008).

Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nat. Rev. Neurosci., 15, 170-180.

Koelsch, S. (2020). A coordinate-based meta-analysis of music-evoked emotions. NeuroImage, 223, 117350. doi:10.1016/j.neuroimage.2020.117350

Lanza, J. (2004). Elevator Music: A Surreal History of Muzak, Easy-Listening, and Other Moodsong; Revised and Expanded Edition. University of Michigan Press. ISBN 0472089420.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. PNAS, 114(10), E2016-E2025. doi:https://doi.org/10.1073/pnas.1619316114

Lisman, J., Grace, A. A., & Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. Trends in Neurosciences, 34(10), 536-547. doi:10.1016/j.tins.2011.07.006

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2016). Neural correlates of specific musical anhedonia. Proceedings of the National Academy of Sciences, 113(46), E7337. https://doi.org/10.1073/pnas.1611211113

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2019). White Matter Microstructure Reflects Individual Differences in Music Reward Sensitivity. The Journal of Neuroscience, 39(25), 5018-5027. doi:10.1523/jneurosci.2020-18.2019

Mas-Herrero E., Zatorre R.J., Rodriguez-Fornells A., Marco-Pallarés J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Curr Biol, 24*(6), 699- 704. doi: 10.1016/j.cub.2014.01.068.

Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2017). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. Nature Human Behaviour, 2(1), 27-32. doi:10.1038/s41562-017-0241-z

Mas-Herrero, E., Maini, L., Sescousse, G., & Zatorre, R. J. (2020). Common and distinct neural correlates of music and food-induced pleasure: A coordinate-based meta-analysis of neuroimaging studies. doi:10.1101/2020.08.14.250894

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Barcelona Music Reward Questionnaire (BMRQ) [Database record]. APA PsycTests.

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

Mas-Herrero, E., Singer, N., Ferreri, L., Mcphee, M., Zatorre, R., & Ripolles, P. (2020). Rock ’n’ Roll but not Sex or Drugs: Music is negatively correlated to depressive symptoms during the COVID-19 pandemic via reward-related mechanisms. doi:10.31234/osf.io/x5upn

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature, 459(7248), 837-841. doi:10.1038/nature08028

Müllensiefen, D., Harrison, P., Caprini, F., & Fancourt, A. (2015). Investigating the importance of self-theories of intelligence and musicality for students' academic and musical achievement. Frontiers in Psychology., 6:1702. doi: 10.3389/fpsyg.2015.01702

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. doi:10.1111/nyas.13654

Pearce, M. T., & Wiggins, G. A. (2012). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. Topics in Cognitive Science, 4(4), 625-652. doi:10.1111/j.1756-8765.2012.01214.x

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. Journal of Personality and Social Psychology, 100(6), 1139-1157. doi:10.1037/a0022406

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nature Neuroscience, 14(2), 257-262. doi:10.1038/nn.2726

Skerritt-Davis, B., Elhilali, M. (2018). Detecting change in stochastic sound sequences. PLoS Comput Biol., 14(5), e1006162. doi: 10.1371/journal.pcbi.1006162.

Sloboda, J. A. & O'Neill, S. A. Emotions in everyday listening to music. in Series in affective science. Music and emotion: Theory and research 415–429 (2001).

Stewart, J., Garrido, S., Hense, C., & Mcferran, K. (2019). Music Use for Mood Regulation: Self-Awareness and Conscious Listening Choices in Young People With Tendencies to Depression. Frontiers in Psychology, 10. doi:10.3389/fpsyg.2019.01199

Symonds, M. R. E. & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol,* 65:13–21 DOI: 10.1007/s00265-010-1037-6

Weinberg, M. K. & Joseph, D. If you're happy and you know it: Music engagement and subjective wellbeing. Psychol. Music 45, 257–267 (2017).

Supplementary Tables

Rock/Pop Music		*Elevator Music*	
Artist	**Title**	**Artist**	**Title**
Adele	Rolling in the Deep	Muzak Orchestra	40 Years Young
Aerosmith	I Don’t Want to Miss a Thing	Muzak Orchestra	At Seventeen
Aretha Franklin	Respect	Muzak Orchestra	Blue Skies
Bruno Mars	Uptown Funk	Muzak Orchestra	C’mon Smile
Dolly Parton	Jolene	Muzak Orchestra	How Little We Know
Journey	Don’t Stop Believin’	Muzak Orchestra	It’s One of Those Nights
Lorde	Royals	Muzak Orchestra	Lady Blue
MGMT	Kids	Muzak Orchestra	Teach Me Tonight
Oasis	Don’t Look Back in Anger	Muzak Orchestra	The One You Say Good Morning To
The Strokes	You Only Live Once	Muzak Orchestra	Java (Experiment 1) / 3 Days of the Condor (Experiment 2)

Supplementary Table 1. *Rock/Pop* and *Elevator* music selected for Experiments 1 and the Rock/Pop version of Experiment 2. “Java” was replaced with “3 Days of the Condor” in Experiment 2 due to being rated as too highly pleasurable.

Classical Music		*Elevator Music*	
Composer	**Title**	**Artist**	**Title**
Beethoven	Für Elise	Muzak Orchestra	3 Days of the Condor
Beethoven	Symphony No.9, Op.125, Mov.2	Muzak Orchestra	40 Years Young
Beethoven	Moonlight Sonata	Muzak Orchestra	At Seventeen
Dvorak	New World Symphony No.9, Mov.4	Muzak Orchestra	Blue Skies
Holst	Jupiter, the Bringer of Jollity	Muzak Orchestra	C’mon Smile
Pachelbel	Canon In D	Muzak Orchestra	How Little We Know
Tchaikovsky	Dance Of the Sugar Plum Fairy	Muzak Orchestra	It’s One of Those Nights
Tchaikovsky	Swan lake Op.20 Scene finale	Muzak Orchestra	Lady Blue
Vivaldi	The four seasons "Spring" Mov.1	Muzak Orchestra	Teach Me Tonight
Vivaldi	The four seasons "Winter" Mov.1	Muzak Orchestra	The One You Say Good Morning To

Supplementary Table 2. *Classical* and *Elevator* music selected for the Classical version of Experiment 2.

Composer	**Title**
Muzak Orchestra	3 Days of the Condor
Muzak Orchestra	40 Years Young
Muzak Orchestra	50 Million Frenchman
Nick Perito Orchestra	Am I On Time
Aisha Duo	Amanda
Muzak Orchestra	At Seventeen
Muzak Orchestra	Blue Skies
Nick Perito Orchestra	C’mon Smile
Nick Perito Orchestra	Canida

Muzak Orchestra	Dance With Me
David O'Brien	Elevator
Bohoman	Elevator Music
Muzak Orchestra	Environs
Muzak Orchestra	Flashback
Laurie Johnson	Happy-Go-Lively
Nick Perito Orchestra	He's My Guy
Nick Perito Orchestra	How Little We Know
Nick Perito Orchestra	It Never Rains in California
Mel Davis	It's Impossible
Muzak Orchestra	It's One of Those Nights
Muzak Orchestra	Jubilation
Muzak Orchestra	Kate McShane
Muzak Orchestra	Lady Blue
Nick Perito Orchestra	Last Tango in Paris
Frank Hunter	Law and Disorder Theme
Muzak Orchestra	Leave Me Alone
The Noveltones	Left Bank Two
Frank Hunter	Lolita
Muzak Orchestra	Loving You
Muzak Orchestra	Nancy
Muzak Orchestra	Paradise Program
Al Calola	Rose Garden
Muzak Orchestra	Star Eyes
Muzak Orchestra	Teach Me Tonight
Benjamin Tissot	The Elevator Bossa Nova
Nick Perito Orchestra	The First Time I Ever Saw Your Face
Muzak Orchestra	The One You Say Good Morning To
Muzak Orchestra	When Things were Rotten
Muzak Orchestra	Whole Lotta Sunlight
Muzak Orchestra	You and Me Against the World

Supplementary Table 3. *Elevator* music selected for Experiment 3.

Composer	**Title**
Beethoven	Symphony No.5, Op.67, Mov.2
Bach	Choral Der Gott
Bach	Cantata Bwv 208
Beethoven	Symphony No.7 in A major, Op.92, Mov.2
Beethoven	Symphony No.4 in B Flat major, Op.60, Mov.2
Beethoven	Für Elise
Beethoven	Piano Sonata No.8 in C Minor, Mov.3
Beethoven	Symphony No.2 in D major Op.36, Mov.1

Beethoven	Violin Sonata No. 5 “Spring" Mov.1
Brahms	String quartet No.1, Mov.2
Chopin	Mazurka in A minor, Op.17, No.4
Chopin	Nocturne in G minor, Op.37, No.1
Chopin	Prelude Op.28, No. 4 in E Minor
Debussy	Clair de lune
Desprez	Ile fantazies de Joskin
Dvorak	New World Symphony No.9, Mov.4
Dvorak	New World Symphony No.9, Mov.2
Dvorak	Symphony No. 8, Mov. 4
Elgar	Cello Concerto in E minor, Op.85, Mov.1
Fauré	Violin Sonata in A Major Mov. 1
Gibbons	Fantasies a6
Haendel	Organ Concerto Op.4, No.2 In B Flat major
Haendel	II Concerto Grosso Op.6, No.4 In A minor
Haydn	Symphony No.38 in C major Mov.3
Haydn	Symphony No.101 in D major
Holst	The Planets -Jupiter, the Bringer of Jollity
Holst	First Suite in E Flat major Op.28, No.1
Holst	The Planets - Venus, The Bringer of Peace
Liszt	Danse Macabre
Mahler	Symphony No.2 "Résurrection", Mov.1
Mahler	Symphony No.1 "Titan", Mov.4
Mozart	Requiem Lacrimosa
Mozart	Symphony No.25 in G minor Mov.2
Mozart	Symphony No.29 in A major Mov.2
Pachelbel	Canon In D
Pärt	Tabula rasa I (Clip3)
Pärt	Tabula rasa IV (Clip1)
Penderecki	Threnody For the Victims of Hiroshima
Rameau	Suite La triomphante Mov.2
Rameau	Pieces De Clavecin Suite In D Minor
Ravel	String quartet in F major Mov.2
Rimski-Korsakov	Sherezade "The Kalender Prince"
Schönberg	String Quartet No.1 in D minor Op.7, Mov.3
Stravinsky	Firebird Suite, Finale
Tchaikovsky	Dance Of the Sugar Plum Fairy
Tchaikovsky	Swan lake Op.20 Scene finale
Vivaldi	The four seasons "Spring" Mov.1
Vivaldi	The four seasons "Winter" Mov.1
Webern	Symphony Op.21, Mov.1

Supplementary Table 4. *Classical* music selected for Experiment 3.