The Repeated Recording Illusion: The Effects of Extrinsic and Individual Difference Factors on Musical Judgements

Manuel Anglada-Tort and Daniel Müllensiefen

Goldsmiths, University of London, London, United Kingdom

Abstract

The repeated recording illusion refers to the phenomenon in which listeners are under the impression that they hear different musical stimuli while they are in fact identical. This phenomenon has not yet been studied systematically. Thus, the present paper aims to construct an experimental paradigm to enable the systematic measurement of the repeated recording illusion, investigating individual difference factors that contribute to it as well as extrinsic factors responsible for differences in musical judgements when the acoustic input remains the same. Seventy-two participants were misled to think that they had heard three different musical performances of an original piece when in fact they were exposed to the same repeated recording. Each time, the recording was accompanied by a different text suggesting a low, medium or high prestige of the performer. Most participants (75 %) believed that they had heard different musical performances. High levels of neuroticism and openness made it significantly more likely that an individual would fall for the illusion. Musicians were not any more or any less susceptible to the illusion than nonmusicians. For participants who fell for the illusion, the explicit prestige texts influenced evaluations of the music significantly. The effect of repeated exposure was only significant in the more familiar music condition. These results suggest that musical judgements are sometimes not based on musical cues but are influenced by factors that do not depend on the music itself. The repeated recording illusion can constitute a paradigm for investigating psychological biases and individual differences in aesthetic and musical judgements because the illusion allows for the study of their effects while the music remains the same. Results are interpreted within Tversky and Kahneman's framework of judgements and decision-making.

Keywords: aesthetics, individual differences, explicit information, music performance, judgements and preferences

The Repeated Recording Illusion: The Effects of Extrinsic and Individual Difference Factors on Musical Judgements

In 1977, the German radio station WDR 3 conducted an audience participation experiment during a live programme (see the description in Behne, 1987). The radio broadcaster misled the audience to think that they would hear three different performances of the same excerpt of Bruckner Symphony No. 4, providing brief information about three different conductors (Karl Böhm, Leonard Bernstein, and Herbert von Karajan) just before each recording was played. However, the radio broadcaster played the same recording three times. The radio station received 536 calls. 81.7 % of the callers were misled and reported differences between the identical music recordings. Only the remaining 18.3 % of the listeners who called in reported that there were no differences between the three performances. Nevertheless, we note that the audience participation experiment had several shortcomings, such as a lack of control over experimental conditions and a potential sampling bias for those listeners who believed they had heard different musical performances to call the radio station. Therefore, one of the main motivations of the present paper was the replication of this phenomenon in an experimental setting.

We will refer to this phenomenon, where listeners are under the impression that they hear different musical performances while in fact they are identical, as *the repeated recording illusion*. Duerksen (1972) was amongst the first academic studies to use a similar approach. He played two tape recordings of an identical piano performance to music major and nonmusic major students. Participants were told that one performance was by an eminent professional pianist and the other one by a student. Both groups rated technical and musical characteristics of the music recording consistently lower when told the performance was by a student than when told it was by a professional.

However, Duerksen (1972) merely attributed the findings to an effect of expectations and did not investigate whether participants believed that they had heard the same or different musical performances.

There are a number of studies that used similar experimental paradigms, presenting participants with identical recordings in succession (Behne & Wöllner, 2011; Cavitt, 1997, 2002; Elliott, 1995; Griffiths, 2008; Juchniewicz, 2008; Radocy, 1976; Silvey, 2009). The main purpose of these studies was to investigate nonmusical factors that influence evaluations of musical performances, such as the effect of expectations (Cavitt, 1997, 2002; Duerksen, 1972), authority (Radocy, 1976), musicians' body movements (Behne & Wöllner, 2011; Juchniewicz, 2008), race and gender (Elliott, 1995), concert dress and physical attractiveness (Griffiths, 2008), and band labels (Silvey, 2009). None of these studies considered the implications of participants potentially falling for the repeated recording illusion. Thus, in none of these studies it is possible to determine whether the illusion occurred in the sample of participants. We considered the repeated recording illusion to be a phenomenon that merits further investigation. Exploring this phenomenon in detail could provide relevant and unique insights to the fields of aesthetics, music perception, cognition, and choice behaviour. Therefore, the present study attempts to measure systematically the repeated recording illusion, investigating individual difference factors that contribute to it as well as extrinsic factors responsible for differences in musical judgements when the acoustic input remains the same.

In relation to the individual difference factors, we suggest that the amount of music training of participants may play an important role in the repeated recording illusion. A large number of previous studies have shown that people with high levels of music training (i.e., musicians) outperform nonmusicians on many music-related tasks, indicating that music training has a positive influence on the efficiency and accuracy with which characteristics of sounds (e.g., pitch and timbre) are encoded in memory (see Pearce, 2015 for a review). For instance, musicians show greater sensitivity to fine variations and nuances in music (e.g., slurs, rests, articulation, and timbre) (Deliege, 1987) and better recognition memory for melodies than nonmusicians (Dowling & Bartlett, 1981; Dowling, 1978; Halpern, Bartlett, & Dowling, 1995; Orsmond & Miller, 1999). We therefore hypothesized that music training would have an effect on the illusion. However, the tasks involved in the above research (e.g., to recognize a melody) are very different to the task that requires an individual to realize that the same music recording is played in succession. Thus, it is difficult to predict the direction in which music training may affect the repeated recording illusion. The present study only attempts to assess whether musicians perform differently on this task compared to nonmusicians.

Arguably, the paradigm used in the repeated recording illusion relies on a judgement bias exerted by a figure of authority (i.e., participants are told by a researcher in a lab condition that they will listen to different performances). In line with Milgram's obedience to authority experiment (1963), Radocy (1976) found that the bias exerted by a figure of authority significantly influenced participants' evaluations of musical events. We therefore considered that individual differences on suggestibility could be an important factor contributing to the illusion. We hypothesized that people with higher levels of susceptibility would be more likely to fall for the repeated recording illusion.

The present research also explored music preferences and personality as possible individual difference factors related to the illusion. Individuals tend to have stronger preferences for certain genres of music, becoming more familiar with the preferred style as a result of repeated listening. Repeated exposure to a piece of music increases the liking for it and decreases its subjective complexity (see North & Hargreaves, 2008 for a review). In relation to personality, research shows that personality traits relate to specific preferences for music styles (see Greasley & Lamont, 2016 for a review). For instance, openness to experience is positively linked to preference for reflective and complex styles (e.g., classical music) (Rentfrow & Gosling, 2003). Furthermore, research on individual differences has found links between personality and suggestibility, showing for example a positive (but low) relationship between suggestibility and neuroticism (see Gudjonsson, 2003 for a review). Therefore, we hypothesized that preferences for music style and personality traits would affect participants' susceptibility to the repeated recording illusion, although we could not specify in which direction.

Extrinsic factors that may be responsible for differences in musical judgements when the acoustic input is identical include the effect of explicit information. Presenting music with explicit information has been shown to be influential in the evaluation of musical performances (Cassidy & Sims, 1991; Cavitt, 1997, 2002; Kroger & Margulis, 2016; Margulis, 2010; Margulis, Kisida, & Greene, 2015; North & Hargreaves, 2005; Silveira & Diaz, 2014; Silvey, 2009; Vuoskoski & Eerola, 2013). In an fMRI study, Kirk, Skov, Hulme, Christensen, and Zeki (2009) presented the same images of artworks with different contextual information, varying in prestige (i.e., labelled as 'gallery' or 'computer generated'). The findings revealed that when the artworks were labelled as 'gallery' they were rated higher in an aesthetic value scale than when labelled as 'computer generated'. The fMRI data showed more activity in the medial orbitofrontal cortex under the gallery context compared to the computer one, suggesting a neural system supporting contextual modulation of aesthetic ratings. In the present study, we hypothesized that participants would evaluate the same recording more positively when presented with a text suggesting high prestige of the performer than when presented with texts of lower prestige levels.

Another important extrinsic factor responsible for differences in musical judgements when the acoustic input is identical may be the effect of repeated exposure. In line with the domain-general mere exposure effect (Zajonc, 1968), liking to an initially neutral stimuli increases with repeated exposure. While the effect of mere exposure has been extensively studied using particular pieces of music as stimuli (see North & Hargreaves, 2008 for a review), only a few studies have examined this effect on evaluations of performances of individual pieces. In a recent study, Kroger and Margulis (2016) presented participants with pairs of solo piano performances and informed them that one was played by a conservatory student and the other by a world-renowned professional. After listening to each pair, participants had to select which they considered to have been performed by the professional. The results indicated that participants selected the second performance as professional more frequently than the first performance, although this effect was modulated by the actual identity of the performer. In relation to the repeated recording illusion, we hypothesized that participants' ratings of the same recording would improve with repeated exposure.

The present research had three main aims. The first was to construct an experimental paradigm to enable the systematic measurement of the repeated recording illusion. The second aim was to investigate possible individual difference factors that contribute to the illusion (i.e., music training, suggestibility, music preferences and personality). The third aim was to investigate extrinsic factors responsible for differences in musical judgements when the acoustic input remains the same (i.e., explicit information and repeated exposure). In addition, in order to capture higher-order interactions between the extrinsic and individual difference factors, an exploratory analysis of the same data aimed to identify conditions that lead to particularly positive or negative judgements.

In constructing the experimental paradigm of the repeated recording illusion, participants were misled to think that they had heard three different performances of an original music piece. However, we played the exact same recording three times in succession. Each time the recording was accompanied by a text suggesting low, medium or high prestige of the performer. We repeated this experimental procedure with two different pieces of music, a piece of classical music and a piece of popular music for which we assumed a high stylistic familiarity for most participants. In order to study the repeated recording illusion without an effect of explicit information, we examined a nonprestige group where we did not manipulate prestige of the performer.

Method

Participants

A sample of seventy-two university students took part in the experiment (36 male, 36 female), aged 19-39 (M = 24.26, SD = 3.60). Twenty-nine participants were considered as trained musicians (M = 45.74, SD = 5.73 on the Musical Training subscale of the Goldsmiths Musical Sophistication Index, Müllensiefen, Gingras, Musil, & Stewart, 2014; and had 6 to 8 years of formal music training). Forty-five participants were considered as nonmusicians (M = 22.71, SD = 7.34 on the Gold-MSI; and had 1 year of formal music training on average). Twelve participants were randomly allocated to a nonprestige condition (6 male, 6 female), aged 21-29 (M = 24.34, SD = 3.45). Participation was on a volunteer basis and unpaid.

Design

The study employed a 3x3x2 repeated measures design. Explicit information (low vs. medium vs. high prestige text), repeated exposure (first vs. second vs. third position), and genre of the original music piece (popular vs. classical music) were the within-participant factors. The three levels of the explicit information factor were fully counter-balanced with presentation order across participants. Half of the participants started with the popular music piece condition and the other half started with the piece of classical music. The dependent variables consisted of a diverse range of musical judgements provided immediately after each listening and at the end of each music condition. In order to explore the repeated recording illusion without an effect of explicit information, we examined a nonprestige group where we did not manipulate prestige of the performer. In addition, we measured individual difference factors that were expected to contribute to the illusion (i.e., music training, suggestibility, music preferences and personality).

Materials

In the popular music condition participants listened to a live recording of 'Jailhouse Rock' by Elvis Presley recorded in NBC studios in 1968. The length of the recording was 1 minute and 36 seconds. This piece was selected because we assumed a high stylistic familiarity for most participants. In the classical music condition participants listened to the final part of a live recording of 'Bruckner Symphony No. 4 Die Romantische' conducted by Günter Wand and performed by the Berliner Philarmonic Orchestra in 1998. The length of the recording was 2 minutes and 48 seconds. This piece was selected in order to replicate empirically the experiment carried out in the German radio station WDR 3(Behne, 1987). The original recordings were edited and normalised using *ableton live* computer software. In the popular live recording we edited the start and end points of the original recording in order to contain only the musical performance element of the recording. Similar to the German radio experiment (Behne, 1987), the start and end points of the classical music piece were edited to contain the final part of the performance. We then normalised the volume of the two recordings to be fixed on the same threshold. Then each recording was duplicated three times and written to the same compact disc, using *iTunes 12.2.2*. Each copy of the music recording was saved under a different name, which included performers' names as used in the texts suggesting different levels of prestige. In the nonprestige condition, the names were 'performance 1', 'performance 2', and 'performance 3'.

To manipulate the effect of explicit information we created three texts suggesting low, medium and high prestige of the performer. The texts had the same format, organisation and a length of 150 words. In the popular music condition ('Jailhouse Rock'), the three 'different' performers were presented as different Elvis impersonators. The prestige texts provided information about the three impersonators, who differed on skill and success (Appendix A). In the classical music condition ('Bruckner's Symphony No.4'), the three 'different' performers were presented as different classical conductors. The prestige texts provided information about the conductors, who differed on skill and success (Appendix B). Günter Wand, the actual conductor of the recording, was not among these conductors. In the nonprestige condition, three different texts were created with the same format, organisation and length of 150 words. While in the popular music condition the three texts provided neutral information from different parts of Elvis Presley's biography, in the classical music condition the texts provided neutral information from different parts of Anton Bruckner's biography.

In order to evaluate liking as well as more objective aspects of the performance (e.g., pitch accuracy and tempo appropriateness), we designed an evaluation form consisting of ten Likert rating scales and two open-text boxes. Nine of the rating scales consisted in sliders ranging from 0 to 100. The rating scales were provided to evaluate the following dimensions: (1) liking of the interpretation, (2) timing and rhythm, and (3) tone quality (from ‘dislike strongly’ to ‘like strongly’), (4) tempo appropriateness (from ‘very inappropriate’ to ‘very appropriate’), (5) pitch accuracy (from ‘very inaccurate’ to ‘very accurate’), (6) emotional quality and (7) overall quality of the performance (form ‘very bad’ to ‘very good’), and degree of agreement to two statements: (8) *some aspects regarding the singer’s vocal technique/ orchestral technique could be improved*, and (9) *some aspects of the overall interpretation could improve* (from ‘strongly disagree’ to ‘strongly agree’). In addition, (10) participants were asked to rate each recording using a 5-star rating scale, ranging from 1 star (strongly dislike) to 5 stars (like strongly). The Likert rating scales were designed to examine differences in musical judgements when the acoustic input is the same. After the ten Likert rating scales, two open-text boxes were provided where participants could write down anything to describe the performance and whether or not they enjoyed it. Answering the open-text boxes was optional.

At the end of each music condition, participants were requested to fill out a final evaluation form. In this final evaluation, participants were asked to rate how much they liked each recording compared to the others, on a scale from 0 (much less than the others) to 100 (much more than the others), where the midpoint of the scale (‘50’) was labelled as ‘as much as the others’. Participants also had to evaluate the familiarity to the original piece of music, on a scale from 0 (‘don’t know at all’) to 100 (‘know very well’). In all rating scales, participants were able to see the number attributed to their specific rating. We also provided an open-text box where participants could write down any optional comments regarding the experience of the experiment. The information from the open-text boxes was used to determine whether participants fell for the illusion or not. When the information from the open-text boxes was not sufficient to make a clear and objective decision, the final comparative rating scales were taken into consideration to determine whether participants fell for the illusion or not. The open-text boxes were used in conjunction with the final comparative rating scales, designed to address a clear limitation in this experiment: we could not ask participants explicitly whether the recordings were the same or different as this would have biased their subsequent evaluations and behaviour in the experiment.

In order to measure the individual difference factors, participants filled out different questionnaires corresponding to each factor. To measure participants' music training and active engagement with music we used the Goldsmiths Musical Sophistication self-report questionnaire (Gold-MSI, Müllensiefen et al., 2014). To measure participants' suggestibility, we used the Social Desirability Scale (SDS-17) (Stöber, 2001) and 8 items adopted from the Susceptibility Persuasive Strategies Scale (STPS) (Kaptein, Ruyter, Markopoulos, & Aarts, 2012), which measured bias to authority, consensus and persuadability, used in a previous study (Unal, Temizel, & Eren, 2014). To assess music preferences and stylistic familiarity, we used the Short Test of Music Preferences revised (STOMP-R, Rentfrow & Gosling, 2003). To measure personality, we used the Big Five Inventory (BFI) (John & Srivastava, 1999).

Procedure

Participants were tested individually in small cubicle rooms. They listened to the music recordings using professional headphones (KNS 8400 Studio Headphones, KRK systems) and at a comfortable listening level that could be adjusted by the individual participants prior to the actual experiment. Participants were told that the main purpose of the study was to measure people's skills in evaluating technical and musical aspects of different musical performances of the same original piece. After filling out the Gold-

MSI questionnaire, participants were instructed to listen to three different interpretations of the same piece of music and to evaluate them as accurately as possible. Before listening to each recording, participants were presented with the corresponding text suggesting different levels of prestige. Immediately after reading the text participants listened to the recording. Immediately after listening to each recording, participants completed the evaluation form, where they were presented with the ten Likert rating scales and two open-text boxes. The experiment had two parts with exactly the same procedure and experimental instructions, but using popular music ('Jailhouse Rock') and classical music ('Bruckner's Symphony No.4') respectively. Immediately after listening the three recordings of each part, participants filled the final evaluation form consisting in the final comparative rating scales and the open-text box. Between completing the two parts of the experiment participants were asked to fill out the STOMP-R questionnaire. In the nonprestige condition the procedure was the same. Participants were also instructed that they would listen to three different performances of the same piece, but the texts presented with the music did not induce any kind of prestige. The three recordings were presented as 'performer 1', 'performer 2', and 'performer 3'. Two weeks after the experiment, participants were asked via email to fill out the BFI, SDS-17, and the 8 items measuring suggestibility. The experiment and questionnaires were implemented in *Qualtrics* software (Qualtrics, Provo, UT). This research was granted ethical approval by the Ethics Committee of the Department of Psychology of Goldsmiths College, University of London.

Results

The Repeated Recording Illusion

In order to determine whether participants fell for the repeated recording illusion or not we used the following procedure: We first assessed the information provided in the open-text boxes. From a total of 14 open-text boxes (7 in the popular music condition and 7 in the classical music conditions), on average participants provided information in 12.65% of the boxes (6.33% in the popular music condition and 6.32% in the classical music condition). By using the information provided in the open-text boxes we were able to identify 48 participants out of 72 (66.67%) in the popular music condition and 50 participants out of 72 (69.45%) in the classical music condition, who provided specific information either reporting differences between performances or reporting that the recordings were the same.

There were cases wherein the information from the open-text boxes was not sufficient to make a clear and objective decision but suggested a direction: either that the participant was not aware that the recordings were identical or that the participant suspected that they were the same. In these cases, we took into consideration the scores from the final comparative rating scales where participants had to compare how much did they like each recording in comparison to the others, on a scale from 0 (much less than the others) to 100 (much more than the others), where the midpoint of the scale ('50') was labelled as 'as much as the others'. We only classified the participant when the scores from the final comparative ratings confirmed the suggested direction from the text boxes. It is important to note that we never took into consideration the scores form the final comparative ratings on its own.

When the information from the open-text boxes was not sufficient and/or too ambiguous to make a clear and objective decision, we did not include the participant's data in the subsequent analyses. Two participants provided highly ambiguous statements in the open-text boxes for both music conditions and the two participants were therefore excluded from the subsequent analyses. Furthermore, one participant provided ambiguous information in the popular music condition and a different participant in the classical music condition. Thus, we had a total of 69 participants in each music condition.

As a consequence of using the above mentioned procedure, we had a total of four possible criteria to determine whether participants fell for the repeated recording illusion or not (see Appendix C for a decision diagram depicting the decision procedure and criteria; Table S1 and S2 from the supplementary materials show the information used to make each individual decision per participant in the two music conditions):

(1) When the information provided in the open-text boxes specifically indicated any differences between performances: In the popular music condition, 37 out of 69 participants (53.62%) specifically reported information indicating differences between performances, such as “more upbeat than the two others, a happier sounding performance” or “this piece sounds more aggressive than the previous one. The tempo for me is faster”. In the classical music condition, 42 out of 69 participants (60.87%) specifically reported information indicating differences between performances, such as “the mood in this piece seemed to escalate a lot more naturally than in the other pieces” or “this interpretation sounded a bit more hesitant. Again, it was not as dramatic as the first performance, but it was clearer than the second one”.

(2) When the information in the open-text boxes specifically indicated that the participant realized that the recordings were the same: In the popular music condition, 11 out 69 participants (15.94%) specifically reported information indicating that the recordings were the same (e.g., “I reckon this is the same file repeated three time” or “this is absolutely the same as the first two”). In the classical music condition, 8 out 69 participants (11.59%) specifically reported information indicating that the recordings were the same (e.g., “This sounds exactly like the two others” or “I thought all 3 were the same”).

(3) When the information provided in the open-text boxes was not sufficient to make a clear and objective decision but suggested that the participant was not aware that the recordings were identical: In these cases, in addition to the open-text boxes, we took into consideration the scores from the final comparative rating scales. If at least one score from the final comparative ratings differed by 10% from the midpoint of the scale ('50'), or any two scores differed by 10% from each other, we considered the participant as falling for the illusion. 19 participants (27.54%) in the popular music condition and 17 participants (24.64%) in the classical music condition were classified using this third criterion.

(4) When the information provided in the open-text boxes was not sufficient to make a clear and objective decision, but suggested that the participant suspected that the performances were the same: In these cases, in addition to the open-text boxes, we took into consideration the scores from the final comparative rating scales. If the three scores from the final comparative ratings did not differ more than 10% from the midpoint of the scale ('50'), we considered the participant as not falling for the illusion. Two participants (2.90%) in the popular music condition and two different participants (2.90%) in the classical music condition were classified using this fourth criterion.

Table 1 shows the number of participants who fell for the repeated recording illusion. In the total sample of participants, 52 out of 69 participants (75.36%) believed that they had heard different musical performances in at least one of the two music conditions. By contrast, 17 participants (24.64%) recognised that the performance was the same in at least one of the two music conditions. Only 6 out of 69 participants (8.7%) realized that the recordings were identical in both music conditions. When looking at the music conditions separately, in the popular music condition 56 participants (81.16%) fell for the illusion and 13 participants (18.84%) did not. In the classical music condition, 59 participants (85.51%) fell for the illusion and 10 participants (14.49%) did not. Additionally, in the nonprestige condition (where the effect of explicit information was not manipulated), 9 out of 12 participants (75%) were susceptible to the illusion. According to a X^2 test, there was no significant association between the music conditions (popular and classical piece) and the occurrence of the repeated recording illusion, X^2 (1) = .47, p = .49. According to Fisher's Exact test, there was no significant association between the presence of prestige (i.e., prestige-suggestion and nonprestige group) and the occurrence of the illusion (p = .65).

Insert Table 1 here.

Generally, participants rated the popular music piece as more familiar (M = 72.16, SD = 21.93 on 100-point rating scale) than the classical piece (M = 13.73, SD = 21.10). This difference in familiarity was highly significant as indicated by a paired samples t-test, t (68) = 16.43, p < .001.

Individual Difference Factors

The analysis of individual difference factors was conducted using a data classification method known as the random forest (Breiman, 2001), in which the aim was to examine whether individual differences contributed to the repeated recording illusion. Random forest procedures differ in a number of ways from other classification methods in that they can handle large sets of predictor variables and do not assume a linear relationship between predictors (see Hastie, Tibshirani, Friedman, & Franklin, 2009; see Pawley & Müllensiefen, 2012 for the use of random forests in music psychology). We used the conditional random forest based on permutation tests as implemented in the R package "party" (Hothorn, Buehlmann, Dudoit, Molinaro, Van der Laan, 2006; Hothorn, Hornik, & Zeileis, 2006; Strobl, Boulesteix, Kneib, Agustin, Zeileis, 2008; Strobl, Malley, & Tutz, 2009). The random forest model was run with a size of 5000 trees. We employed a measure of variable importance for each predictor variable, which is designed to produce unbiased estimates of variable importance even in situations where significant correlations between predictor variables exist and when the dependent variable is very unequally distributed (atza, Strobl, & Boulesteix, 2013).

As predictor variables, we used 6 demographic variables as well as musical variables that were collected during the experimental session (age, gender, Gold-MSI Musical Training and Active Engagement scores, STOMP preference scores for Reflective & Complex, Intense & Rebellious, Upbeat & Conventional, and Energetic & Rhythmic). Data for 9 additional variables were collected via the follow-up questionnaire measuring the big five personality traits (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness) as well as suggestibility (Authority score, Consensus score, Persuadability score, and Social Desirability score). Using these 17 predictor variables we computed two different models with two different binary dependent variables: (a) a strict criterion model in which only those participants who fell for the illusion in both music conditions were considered as not falling for the illusion, and (b) a less strict criterion model where we considered as not falling for the illusion those participants who fell for the illusion in at least one of the two music conditions. A variable importance score was obtained for each predictor variable, describing how predictive each variable was compared to the others. We applied a “confidence interval” criterion in order to select the top performing variables. Only the variables whose variable importance scores were positive and greater than the absolute value of the lowest negative variable importance score were selected (Strobl et al., 2008; Strobl et al., 2009).

The two models (strict and less strict criterion) delivered very similar results, indicating that there were two variable importance scores that met the above criterion (*neuroticism* and *openness*). In both models, neuroticism was the most important variable contributing to the repeated recording illusion, followed by openness (see Appendix D for graphs with the 17 variable important scores in the two models). In the strict criterion model, neuroticism was approximately 3.5 times more important than openness. In this model, those participants falling for the illusion in the two music conditions scored higher in neuroticism (*M*= 23. 41, *SD* = 5.17) and openness (*M*= 40.12, *SD* = 5.14) than those participants who did not fall for the illusion (*M* = 17.43, *SD* = 6.85 on the neuroticism factor; *M* = 35.28, *SD* = 7.02 on the openness factor). In the less strict criterion model, neuroticism was approximately 3 times more important than openness. In this model, those participants who fell for the illusion in at least one of the two music conditions scored higher in neuroticism (*M*= 23.14, *SD* = 5.55) and openness (*M*= 40.12, *SD* = 5.42) than those participants who did not fall for the illusion (*M* = 17.43, *SD* = 6.85 on the neuroticism factor; *M* = 35.28, *SD* = 7.02 on the openness factor).

Extrinsic Factors: The Effects of Explicit Information and Repeated Exposure

The subsequent analyses included the sixty participants of the main experimental group (i.e., where we manipulated the effect of explicit information). In the popular music condition, three participants were excluded from the analyses and ten fell for the illusion. Therefore, in the popular music condition we had a total of 47 participants. In the classical music condition, three participants were excluded from the analyses and nine fell for the illusion. Therefore, in the classical music condition we had a total of 48 participants.

Participants' ratings on the ten Likert rating scales were aggregated into a single scale. First, the ratings of each participant on each rating scale were transformed into z-scores across the ratings of all six recordings (three in the popular music condition and three in the classical). Then, a principal component analysis (PCA) was conducted on the z-transformed data of the ten rating scales. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis, KMO = .93 ('marvellous' according to Hutcheson & Sofroniou, 1999). In addition, all KMO values for individual rating scales were greater than .86, which is well above commonly accepted limit of .5 (Field, 2013). The scree plot of the different factor solution was very clear and indicated a solution with just one factor. Moreover, there was only one PCA component with an eigenvalue >1 which explained 64.56% of the variance. Thus, this 1-factor PCA solution was accepted and component scores for all participants' ratings were computed using the regression method.

Because the two music recordings used in the popular and classical music conditions differed substantially in several aspects (i.e., music genre, familiarity, presence of words/ vocalizations, duration of the excerpt and quality of the recording), we ran two separate models, one with the ratings obtained in the popular music condition and one with the ratings obtained in the classical music condition (see Appendix E for a summary table of both models). In addition, the ratings were standardised separately for each music condition.

To test the hypothesis regarding the effects of explicit information and repeated exposure we used the R packages lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016) to perform a linear mixed effects analysis with the z-scores of the participants' ratings as the dependent variable. In the two models, explicit information (low, medium, and high prestige of the text) and repeated exposure (first, second, and third position) were the fixed effect independent factors, whereas participants were the random effect factor.

The linear mixed-effect model of the popular music condition revealed that there were significant main effects of explicit information ($p < .001$) and repeated exposure ($p < .001$). Because the interaction between explicit information and repeated exposure was not significant we ran the model again only with the two main factors. The effects of explicit information and repeated exposure become visible in Figure 1. The effect of explicit information shows that when the recording was presented with a high prestige text the ratings were significantly higher than when presented with low and medium texts. The effect of repeated exposure of the recording shows that when the recording was heard in the second and third positions the ratings were significantly higher than when heard in the first position.

Insert Figure 1 here.

The linear mixed-effect model of the classical music condition revealed that there was a significant main effect of explicit information ($p < .001$). However, the effect of repeated exposure and the interaction between explicit information and repeated exposure were not significant. Because the interaction between explicit information and repeated exposure was not significant we ran the model again only with the two main factors. The effect of explicit information shows that when the recording was presented with a high prestige text the ratings were significantly higher than when presented with low and medium texts (Figure 2).

Insert Figure 2 here.

The R^2 for the classical music model was 0.16 and therefore lower than the R^2 of 0.28 of the popular music model, indicating that the extrinsic factors explained more of the variance in the more familiar popular music condition.

Exploratory Analysis (Regression Model Tree)

In order to capture higher order interactions between extrinsic and individual difference factors and identify conditions that lead to particularly low and high ratings, we computed a regression tree model based on permutation tests as implemented in the R package “party” (Hothorn et al., 2006; Hothorn et al., 2006; Strobl et al., 2008; Strobl, et al., 2009). Statistical tree models differ in a number of ways from linear regression models (see Hastie et al., 2009) in that they use a built-in variable selection mechanism and therefore can handle large sets of predictor variables. In addition, tree models do not assume a linear relationship between predictors and the dependent variable and they are very useful for modelling higher-order interaction effects between predictor variables automatically. For this study we used a particular family of tree models called conditional inference trees that combine the rigorous theory of permutation statistics (Hothorn et al., 2006) with the principle of recursive partitioning (Zeileis, Hothorn, & Hornik, 2008**).**

For the regression tree model, the z-transformed participants’ ratings served as dependent variable. In addition to the two extrinsic factors (explicit information and repeated exposure), we added the factor music genre (popular and classical music) and six individual difference variables (1. music training, 2. self-rated familiarity with the music piece, 3. preference for the STOMP meta-genre reflective & complex, 4. preference for the STOMP meta-genre Intense & Rebellious, 5. neuroticism, and 6. Openness), resulting in a total of nine independent variables. Figure 3 shows the structure of the regression tree. The model makes use of only 3 of the nine independent variables and has an R^2 value of 0.23. For each node of the tree, the p-values indicating the significance of the split based on the permutation statistics are presented as well as a description of the two subgroups of the split on the independent variable. For the terminal nodes at the bottom of the graph, the distribution of the ratings on the standardised rating scale are depicted as box- and whiskers plots.

The tree model can be interpreted by starting at the top and following each branch down, to arrive at a terminal node. A path to a terminal node describes the interaction of experimental conditions that lead to a particular subset of ratings. To arrive at the subset with the highest (i.e. most positive) average ratings, follow the first "Explicit Information" node down the "High Prestige" branch (left-hand side) and then descend to the left at the "Repeated Exposure" node down the "2nd and 3rd Positions" branch. This branch can be interpreted as follows: when participants listened to the music recording presented with a high prestige text in the second and third positions, the average ratings were around 1 and, therefore, the highest compared to the other terminal branches of the model. In contrast, the lowest ratings, which were around -1, were given when the recording was presented with low and medium prestige texts, in the popular music condition, and when the recording was heard for the first time. Overall, the regression tree model confirms the effects of explicit information and repeated exposure, but it also shows higher-level interactions between the extrinsic factors and the two pieces of music. None of the individual difference factors were significant in the tree model. This indicates that after participants had fallen for the illusion, individual difference factors did not play an important role and musical judgements were mainly influenced by the extrinsic factors.

Insert Figure 3 here.

Discussion

The primary aim of the present study was to construct an experimental paradigm to enable the systematic measurement of the repeated recording illusion. Participants were misled to think that they had heard three different performances of an original piece when in fact they were exposed to the same repeated recording. Each time, the recording was accompanied by a different text suggesting a low, medium or high prestige of the performer. Most participants (75.36%) believed that they had heard different musical performances when in fact they were identical. In contrast, seventeen participants (24.64%) recognised that the performance was the same in at least one of the two music conditions. Only six participants (8.7%) realized that the recordings were identical in both music conditions. Nearly three-quarters of the participants provided verbal comments indicating specific differences between the performances (e.g., "this piece sounds more aggressive than the previous one. The tempo for me is faster") or that they were the same (e.g., "I reckon this is the same file repeated three times"). Thus, it can be concluded that the majority of the participants fell for the repeated recording illusion. This finding suggests that musical judgements are sometimes not based on perceptual features and musical cues but are influenced by factors that do not depend on the music itself. This is at least true when a mild deception is applied and participants believe that they had heard different performances.

It could be argued that the repeated recording illusion occurs in part because participants are not familiar with the original piece of music. Therefore, we examined the illusion using two different pieces that were significantly different on familiarity, a highly familiar piece of popular music ('Jailhouse Rock' by Elvis Presley) and a highly unfamiliar piece of classical music (‘Bruckner’s Symphony No. 4’). The repeated recording illusion occurred similarly in the two music conditions. However, these two recordings differed substantially in several other aspects, including music genre, complexity, length of the excerpt, presence of vocals and quality of the recording. Thus, these variables are confounded in this experimental setup. Any interpretation of differences between the two musical stimuli will have to take this into account. Further studies should explore the repeated recording illusion with a larger range of different performances and recordings.

It is important to note that there is a main methodological restriction to be considered in the experimental design used here: an implicit bias of authority figure. In other words, the fact that participants were told they would listen to ‘three different performances’ by an investigator in a lab situation may account, at least partly, for the occurrence of the illusion. It would be interesting for future research to investigate the repeated recording illusion using an experimental paradigm without any implicit bias of authority. This paradigm could consist in presenting participants with pairs of different and identical musical performances. Participants would be instructed to rate how different are the two performances using several rating scales. In the cases where the performances were identical, participants’ ratings would indicate to what extent people hear differences when listening to the same repeated recording without relying on a judgements bias excreted by a figure of authority.

The second aim of the study was to investigate possible individual difference factors that contribute to the repeated recording illusion. The most important individual difference factor related to the illusion was the personality trait of neuroticism, which is in line with previous research showing a positive (but low) link between vulnerability to suggestion and neuroticism (see Gudjonsson, 2003). This finding suggests that people who tend to be anxious, pessimistic, shy, fearful, vulnerable and emotionally unstable are more likely to fall for the repeated recording illusion. Although less important, openness to experience also was a significant factor related to the occurrence of the illusion, suggesting that people who tend to be curious, imaginative, artistic, excitable and unconventional are more likely to fall for the illusion. Importantly, none of the other individual difference factors that were expected to contribute to the illusion were significant, including music training, suggestibility and preferences for music style. We consider particularly interesting that different levels of suggestibility (including bias to authority, consensus, persuadabiliy and social desirability) were not related with the occurrence of the illusion. Moreover, in our sample of participants, highly trained musicians were not any more or any less susceptible to the repeated recording illusion than participants with low levels of music training. Thus, it remains still open the question of which are the main individual differences contributing to the repeated recording illusion. For instance, what would occur when using participants with a greater range of music training and expertise (e.g., top-level professional musicians and music critics)? Would other individual differences (e.g., intelligence, memory, perceptual abilities) be able to explain why some people fall for the illusion while others seem no be unaffected by it?

The third aim of the present research was to investigate extrinsic factors responsible for differences in musical judgements when the acoustic input remains the same. As predicted, we found that the effect of explicit information contributed significantly to differences in musical judgements. This effect was clear in the two music conditions, where participants rated the same music recording significantly better when presented with a high prestige text than when presented with low and medium prestige texts. This finding is consistent with previous research on the effects of explicit information upon aesthetic reactions to music (e.g., Kroger & Margulis, 2016; Margulis, 2010; Margulis, Kisida, & Greene, 2015; North & Hargreaves, 2005). Using a similar paradigm, where identical artworks were presented with different contextual explicit information varying in prestige, Kirk et al. (2009) found that prefrontal and orbitofrontal cortices recruited by aesthetic judgements were significantly influenced by the explicit information presented with the same stimuli. We suggest that this neural system could also be responsible for the modulation of aesthetic reactions to music by explicit contextual information.

The effect of repeated exposure was only significant in the more familiar popular music condition, but not in the more unfamiliar classical music condition. This finding supports partly previous research on the effects of repeated exposure to music (North & Hargreaves, 2008 for a review). In one of the few studies using musical performances as stimuli, Kroger and Margulis (2016) found that evaluations of performances were driven by a combination of repeated exposure and the actual identity of the performer. Interestingly, in a second experiment, Kroger and Margulis (2016) found that the effect of explicit information was mitigated by the influence of the actual performer and repeated exposure, showing interplay between intrinsic and extrinsic factors. In the present study, the two original pieces of music differed in a number of important aspects. For instance, the classical piece was a minute longer than the popular piece, did not contain vocals and was highly unfamiliar to most of the participants. Furthermore, while the popular music piece was a live recording from 1968 that had a notably worse recording quality than ordinary studio recordings, the quality of the classical music piece (recorded live in 1998) was superior. Therefore, it may be possible that the effect of repeated exposure did not affect participants in the classical music condition because of the nature of the music recording. Moreover, the explicit information presented with the recordings might have had a different impact on participants in the two music conditions. Future studies will need to explore the strength of the effect of repeated exposure across a larger range of different performances and recordings.

In an attempt to explore higher-order interactions between the extrinsic and individual difference factors, we used a regression tree model in which we identified conditions that lead to particularly low and high ratings. The highest ratings were given when the music recording was presented with a high prestige text and heard in the second and third positions. In contrast, the lowest ratings were found when participants listened to the popular music piece in the first position and presented with low and medium prestige texts. Overall, the regression tree model confirmed the effects of explicit information and repeated exposure, but it also showed higher-level interactions between the extrinsic factors and the two pieces of music. None of the individual difference factors used in the model (music training, familiarity with the original piece, music preferences, neuroticism, and openness) were significant in the regression tree model. This finding suggests that after participants had fallen for the illusion, individual difference factors did not play an important role and musical judgements were mainly influenced by the extrinsic factors.

The present study focussed on extrinsic factors in order to examine differences in musical judgements when the acoustic input remains the same. Nevertheless, one could argue that the factors of explicit information and repeated exposure might also be responsible, in part, for the occurrence of the illusion. The results from a nonprestige group, where the effect of explicit information was not manipulated, indicated that 75 % participants were susceptible to the illusion. This finding suggests that the effect of explicit information is not essential for the occurrence of the illusion. By contrast, we consider it likely that the effect of repeated exposure contributes to the illusion. In an extensive investigation of repetition in musical experience, Margulis (2014) provides relevant insights to this matter. She stated that, "[a]t a minimum, a repeated element will sound different from its initial presentation by virtue of coming later and having been heard before" (Margulis, 2014, p. 35). Although in this quote Margulis refers to repetition within individual pieces of music, we find it plausible that the same principle should apply to the repeated recording illusion: while the musical input remains the same, repeated exposure modifies the listening experience, giving rise to the feeling that the performances are different.

Two relevant questions arise from the results of this study. Why are some individuals more susceptible to the illusion than others? One way to approach this question is the study of further individual difference factors (e.g. intelligence, memory, perceptual abilities) that may be associated with the repeated recording illusion. The second question refers to a more fundamental issue: did participants in this study actually perceive differences between the repetitions of the same recording? Or, alternatively, did they believe they heard differences because they were misled to think so? We encourage the use of neuroimaging techniques as one possible approach to investigate whether the illusion is a perceptual phenomenon or rather a bias in a secondary and later stage of cognitive processing and decision-making.

Taking a wider perspective, the research framework developed by Tversky and Kahneman (Kahneman & Tversky, 1984; Tversky & Kahneman, 1974; see Kahneman, 2011 for a review) could provide a theoretical framework by which the results of the current study could be interpreted. Although it does not involve music and is mainly concerned with economic decision processes, Tversky and Kahneman's framework offers insight into how to investigate traditional psychological biases in musical judgements by using recent research on human judgements and decision-making. However, this framework has not yet been applied explicitly to the study of evaluative judgement processes involving music.

The effect of explicit information may fall within a broad heuristic principle, namely, the affect heuristic (Kahneman & Frederick, 2002; Slovic, Finucane, Peters, & MacGregor, 2002), which refers to the reliance on good or bad feelings experienced in relation to a stimulus. Thus, if the emotions associated with a stimulus are positive, people will be more likely to judge characteristics of the pertinent stimulus more positively, as found in the present study when the music recording was presented with a high prestige text. Similarly, the effect of repeated exposure is one of several mechanisms within the bias of perceptual fluency (Kahneman, 2011), which has been widely shown to influence human judgements and decision-making in many areas (see Reber, Schwarz, & Winkielman, 2004 for a review). Such findings suggest that perceptual fluency gives rise to feelings of familiarity and a positive affective response that results in an increase in preference judgements. In the present study, this is evident only when participants listened to the more familiar popular music recording.

Our results suggest that at least in certain situations, evaluations of music rely on judgement biases and heuristics that do not depend on the stimuli themselves, which is in line with models of decision-making and the research framework developed by Tversky and Kahneman. However, when applying Tversky and Kahneman's framework to the study of evaluative and judgment processes involving music, one should consider the implications and difficulties of using music as stimuli (e.g., familiarity, complexity, presence of vocals, individual preferences to music, personality). This approach wherein biases in musical judgements are linked to comparable research in behavioural economics could be used to investigate and better understand musical judgements, preferences and choice behaviour. This general approach, that could be termed *the behavioural economics of music,* would attempt to create a solid understating of the role that behavioural economics can play in the study of musical judgements and preferences, two fields that have been surprisingly unconnected in the literature so far.

In summary, the findings of the present study show that most participants believed that they had heard different musical performances when in fact they were identical. This illusion occurred regardless of participants' levels of suggestibility, music training, and preferences for music style. However, high levels on the personality traits of neuroticism and openness made it significantly more likely that an individual would fall for the illusion. While the explicit information presented with the music influenced participants' evaluations of music significantly, the effect of repeated exposure affected participants' ratings only in the more familiar popular music recording. These findings support previous research showing that musical judgements are sometimes not based on musical cues and features but are influenced by factors that do not depend on the music itself. Beyond the findings and limitations of the present research, the repeated recording illusion can constitute a useful paradigm for investigating psychological biases and individual differences in aesthetic and musical judgements because the illusion allows for the study of their effects while the music remains the same.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. doi: 10.1023/A:1010933404324

Behne, K. -E. (1987). Urteile und Vorurteile: Die Alltagsmusiktheorien jugendlicher Hörer. In H. Motte-Haber (Ed.), *Psychologische Grundlagen des Musiklernens* (pp. 221-272). Kassel, DE: Bärenreiter.

Behne, K. -E., & Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, *15*(3), 324-342. doi: 10.1177/1029864911410955

Cassidy, J. W., & Sims, W. L. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education*, *39*(1), 23. doi: 10.2307/3344606

Cavitt, M. E. (1997). Effects of expectations on evaluators' judgments of music Performance. *Texas Music Education Research,* p. 7-10. Retrieved from https://www.tmea.org/assets/pdf/research/Cav1997.pdf

Cavitt, M. E. (2002). Differential expectation effects as factors in evaluations and feedback of musical performance. *Texas Music Education Research,* p. 2-6. Retrieved from https://www.tmea.org/assets/pdf/research/Cav2002.pdf

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4*(4), 325-60.

Dowling, W. J. (1978). Scale and contour: Two componenets of a theory of memory for melodies. *Psychological Review, 85*(4), 341-354. doi: 10.1037/0033-295X.85.4.341

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, *1*, 30-49.

Duerksen, G. L. (1972). Some effects of expectation on evaluation of recorded musical performance. *Journal of Research in Music Education*, *20*(2), 268-272. doi: 10.2307/3344093

Elliott, C. A. (1995). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education, 127,* 50-56.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage.

Greasley, A., & Lamont, A. (2016). Musical Preferences. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology (second edition)* (pp. 263-281). Oxford, UK: Oxford University Press.

Griffiths, N. K. (2008). The effects of concert dress and physical appearance on perceptions of female solo performers. *Musicae Scientiae*, *12*(2), 273-290. doi: 10.1177/102986490801200205

Gudjonsson G.H. (2003). *The psychology of interrogations and confessions: A handbook*. West Sussex, UK: John Wiley & Sons.

Halpern. A. R., Bartlett, J., & Dowling, W. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging*, *10*(3), 325–342. doi: 10.1037/0882-7974.10.3.325

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Hierarchical Clustering. In T. Hastie, E. Tibshiran, & J. Friedman (Eds.), *The elements of statistical learning: Data*

Mining, inference and prediction (2nd ed.) (pp. 520-528). New York, NY: Springer.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A, & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics, 7*(3), 355-373. doi: 10.1093/biostatistics/kxj011

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, *15*(3), 651-674. doi: 10.1198/106186006X133933

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist. Introductory statistics using generalized linear models*. doi: 10.4135/9780857028075

Janitza, S., Strobl, C., & Boulesteix, A. –L. (2013). An AUC-based permutation variable importance measure for random forests. *BMC Bioinformatics, 14*(1), 119. doi: 10.1186/1471-2105-14-119

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(510), 102-138.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music*, *36*, 417-427

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Friffin, D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive thought* (pp. 49-81). New York, USA: Cambridge University Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American psychologist*, *39*(4), 341. doi: 10.1037/0003-066X.39.4.341

Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems*, *2*(2), 1-25. doi: 10.1145/2209310.2209313

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. doi: 10.1016/j.neuroimage.2008.10.009

Kroger, C., & Margulis, E. H. (2016). "But they told me it was professional": Extrinsic factors in the evaluation of musical performance. *Psychology of Music*, *45*(1), 49-64. doi: 10.1177/0305735616642543

Kuznetsova, A., Brockhoff, P. B., & Christensen R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models. *R Package Version*. doi: http://CRAN.R-project.org/package=lmerTest

Margulis, E. H. (2010). When program notes don't help: Music descriptions and enjoyment. *Psychology of Music, 38*, 285-302.

Margulis, E. H. (2014). *On repeat: How music plays the mind*. New York, NY: Ocford University Press.

Margulis, E. H., Kisida, B., & Greene, J. P. (2015). A knowing ear: The effect of explicit information on children's experience of a musical performance. *Psychology of Music*, *43*(4), 596-605. doi: 10.1177/0305735613510343

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal Psychology*, *67*(4), 371–378. doi: 10.1037/h0040525

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), e89642. doi: 10.1371/journal.pone.0089642

North, A. C., & Hargreaves, D. J. (2005). Brief report: Labelling effects on the perceived deleterious consequences of pop music listening. *Journal of Adolescence*, *28*(3), 433-440. doi: 10.1016/j.adolescence.2004.09.003

North, A., & Hargreaves, D. (2008). *The social and applied psychology of music*. New York, NY: Oxford University Press.

Orsmond, G. I., & Miller, L. K. (1999). Cognitive, musical and environmental correlates of early music instruction. *Psychology of Music*, *27*, 18-37. doi: 10.1177/0305735699271003

Pawley, A., & Müllensiefen, D. (2012). The science of singing along: A quantitative field study on sing-along behavior in the north of England. *Music Perception, 30*(2), 129–146. doi: 10.1525/mp.2012.30.2.129

Pearce, M. T. (2015). Effects on processes involved in musical appreciation. In J. P. Huston, M. Nadal, F. Mora, L. Agnati, F. Mora, & C. J. Cela-Conde (Eds.), *Art, aesthetics and the brain* (pp. 319-338). Oxford, UK: Oxford University Press.

Radocy, R. E. (1976). Effects of authority figure biases on changing judgments of musical events. *Journal of Research in Music Education*, *24*(3), 119-128. doi: 10.2307/3345155

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364-382. doi: 10.1207/s15327957pspr0804_3

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236-1256. doi: 10.1037/0022-3514.84.6.1236

Silveira, J. M., & Diaz, F. M. (2014). The effect of subtitles on listeners' perceptions of expressivity. *Psychology of Music*, *42*(2), 233-250.

Silvey, B. A. (2009). The effects of band labels on evaluators' judgments of musical performance. *Update: Applications of Research in Music Education*, *28*(1), 47-52. doi: 10.1177/8755123309344111

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *Journal of Socio-Economics*, *31*(4), 329-342. doi: 10.1016/S1053-5357(02)00174-9

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, *17*(3), 222-232. doi: 10.1027//1015-5759.17.3.222

Strobl, C., Boulesteix, A. -L., Kneib, T., Augustin, T. & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(23), 307. doi: 10.1186/1471-2105-9-307

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. doi: 10.1037/a0016973

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.)*, *185*(4157), 1124-31. doi: 10.1126/science.185.4157.1124

Unal, P., Temizel, T. T., & Eren, P. E. (2014, May). *An exploratory study on the outcomes of influence stra-tegies in mobile application recommendations*. Paper presented at Proceedings of the Second International Workshop on Behavior Change Support Systems (BCSS2014). Padova, IT.

Vuoskoski, J. K., & Eerola, T. (2013). Extramusical information contributes to emotions induced by music. *Psychology of Music*, *43*(2), 262–274. doi: 10.1177/0305735613502373

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9* (2p2), 1-27. doi: 10.1037/h0025848

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492-514. doi: 10.1198/106186008X319331

Appendix A

Prestige Texts (Low, Medium, and High) used in the Popular Music Condition

('Jailhouse Rock' by Elvis Presley)

Popular Music Condition: Low Prestige Text – Larry Leigh

Larry Leigh was a humble truck driver who developed an obsessive love for Elvis Presley from an early age. His lack of musical training did not hold him back from impersonating his beloved star on stage.

Although Leigh participated in *the Georgia Elvis Festival* and *the ETA Preliminary Competition*, he never succeed in his career as an Elvis imitator and his critics labelled him as an amateur singer.

Leigh had some similarities to Elvis' voice, however, due to the lack of success he gave up his dream. A close friend expressed that his unsuccessful career as an Elvis impersonator made him spiral into a deep depression.

Information source: <u>Impersonators in America</u> (2001) by Esther Newton.

Popular Music Condition: Medium Prestige Text – Drew Polsun

Drew Polsun was not just an owner of a music record store, but also a lifelong lover and impersonator of Elvis. With some musical training from an early age, he began to perform as Elvis in his local area.

Putting his business on hold he began to compete more seriously in Elvis impersonator contests. At *the Ultimate Elvis Tribute Artist Contest* he placed 17th out of roughly 40 impersonators.

Never able to reach the top, he turned back to his business at the music store and focused on his family. Still to this day, he continues to perform in his hometown sharing his love for the King.

Information source: <u>Impersonating Elvis</u> (2009) by Leslie Rubinkowski.

Popular Music Condition: High Prestige Text – Shawn Klush

Starting at the age of 6, Shawn Klush would sing and dance like Elvis. He now works as a professional actor and entertainer, and has a strong musical background, from classical to jazz music, in singing and guitar.

Klush became very successful as one of the top Elvis impersonators. He released 3 CDs, became the grand champion at the $150,000 *World Elvis Tribute Artist Competition*, and received the Heart of the Kind, *Worldwide Ambassador of Elvis Award*.

In 2007, he was named the *World's Greatest Elvis* by 6.5 million international viewers on BBC1 Television in the United Kingdom. Since then, Klush is considered one of the world's most professional Elvis Tribute Artists.

Information source: www.shawnklush.com

Appendix B

Prestige Texts (Low, Medium, and High) used in the Classical Music Condition

('Bruckner Symphony No.4')

Classical Music Condition: Low Prestige Text – Kurt Schlichter

With an undergraduate degree from Royal Holloway, University of London, Kurt Schlichter, continued his education as a Masters student in conducting at the Royal College of Music.

As a young student, Kurt found it quite challenging to conduct a broad range of classical works in real-life rehearsal and performance situations. But these experiences allowed him to grow and become one of the top students in his class.

Talking to Kurt about his experiences he says, "The opportunities at the Royal College have been highly varied and rewarding. For my final project I was able to conduct one of my favourite Bruckner pieces which allowed me to engage fully with the complex compositional style of this genius composer".

Information source: www.rcm.ac.uk/conducting/story/kurt

Classical Music Condition: Medium Prestige Text – Pablo Giménez

As the principal conductor of the Royal Seville Symphony Orchestra, Pablo Giménez, put Spain on the map as a place for beautiful interpretations of classical music. Based in Seville, his orchestra has played some very well know interpretations of classical favourites.

Gimenez had become an up-and-coming conductor, performing *Carmen* at the *Liceu's theatre* of Barcelona and international concert halls in China and Australia. His album of *Zarzuela* has been released throughout Europe, but did not receive any awards.

Even though his interpretations of works by Brahms and Handel have been regarded as "beautiful orchestral performances", his interpretations of Beethoven and Wagner were labelled as "lifeless and inexpressive" by some reviewers.

Information source: Guide to Spanish Conductors (2010) by Jose Luis-Garcia.

Classical Music Condition: High Prestige Text – Claudio Abbado

Claudio Abbado was widely considered one of the greatest conductors of the 20th century. He served as principal conductor of the Berlin Philharmonic and London Symphony Orchestra, one of the best ensembles in the world.

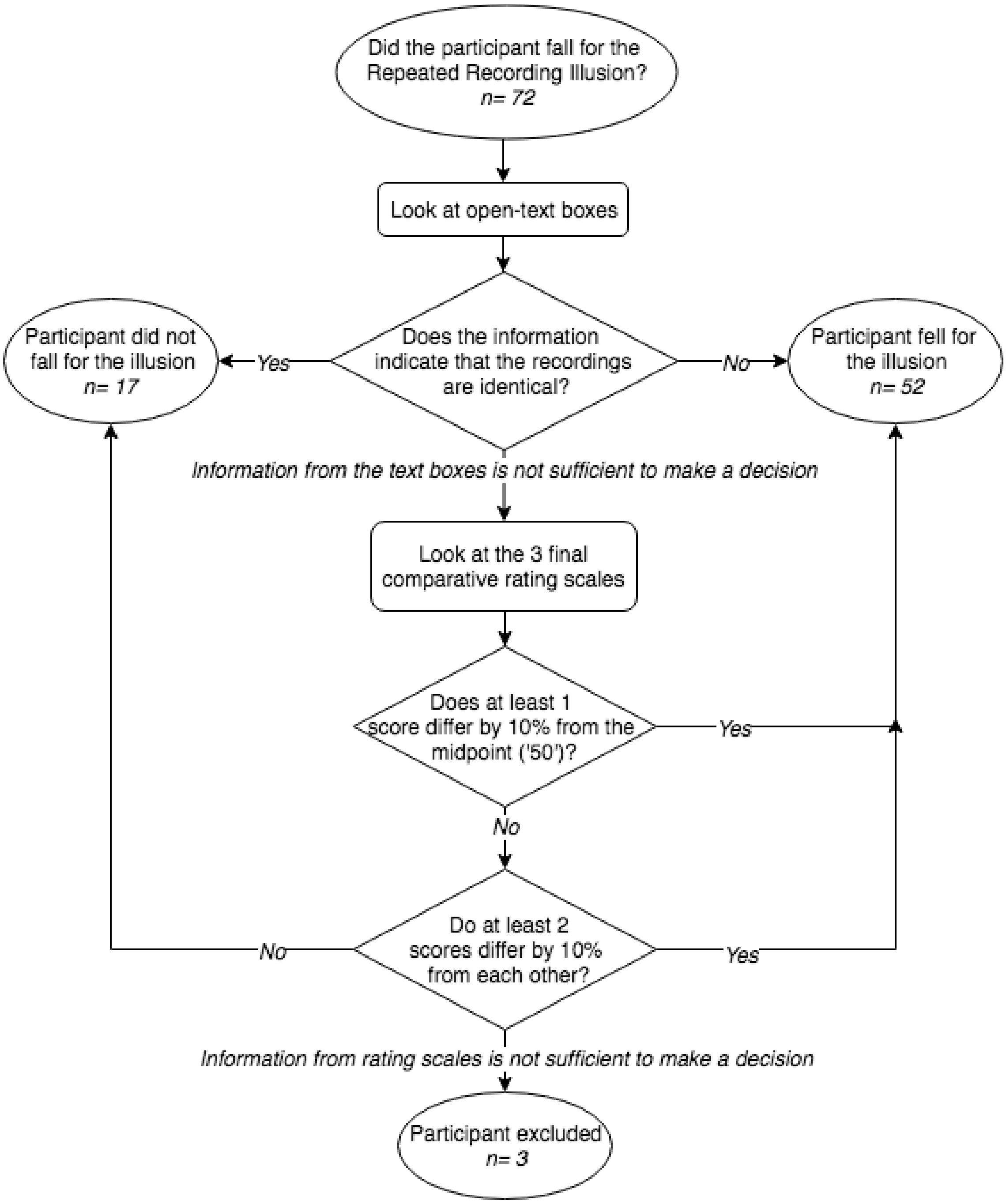
Claudio's career couldn't be more successful. He won the International Competition for Composers and the Grammy Award in the Best Small Ensemble Performance. Additionally, the Orchestra Academy of the Berlin Philharmonic established the Claudio Abbado Composition Prize in his honour.

In 2004, Abbado conducted the Berlin Philharmonic to performed Bruckner's Symphony No. 4 in a series of recorded live concerts. The resulting CD won Best Orchestral Recording of the Year in *Gramophone awards.*

Information source: Stories of the Great Contemporary Conductors (2012) by Maurice Hinson.

Appendix C

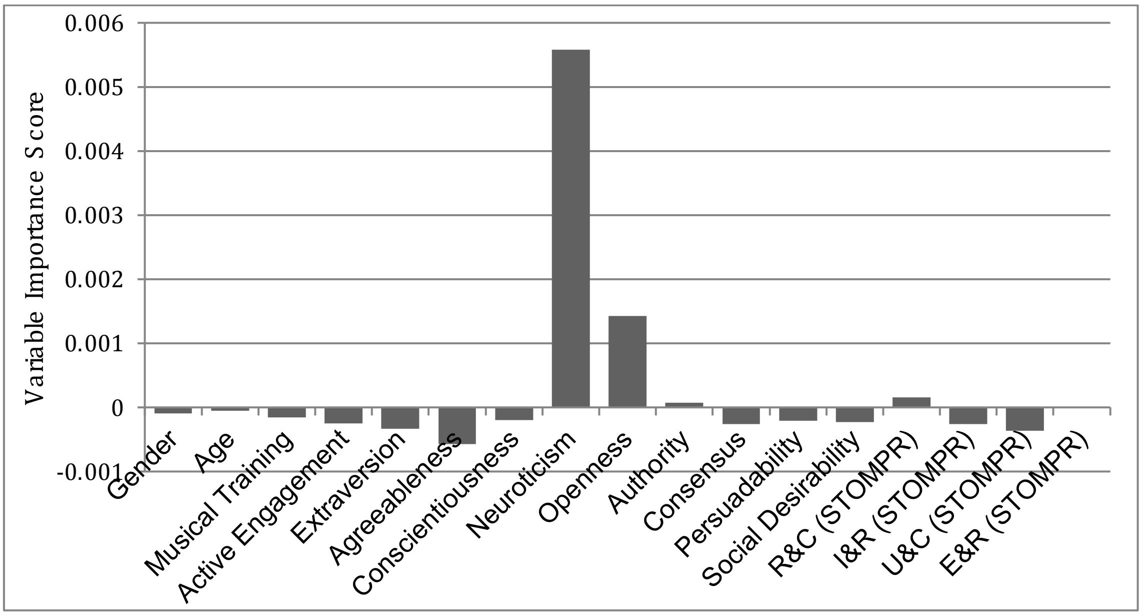
Decision Diagram of the Procedure used to Determine Whether Participants Fell for the Repeated Recording Illusion



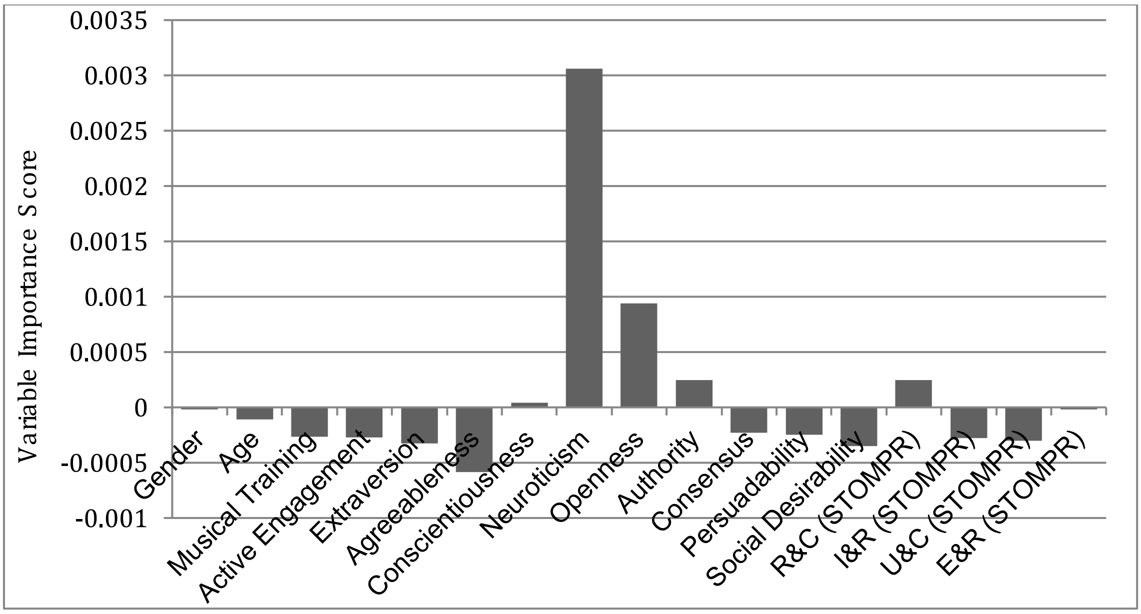
Appendix D

Variable Importance Scores for the 17 Variables

Variable Importance Scores for Predictor Variables in Random Forest Model using Strict Criterion (i.e., Participants Falling for the Illusion in Both Music Conditions)



Variable Importance Scores for Predictor Variables in Random Forest Model using Less Strict Criterion (i.e., Participants Falling for the Illusion in One Music Condition)



Appendix E

Summary Tables of the Two Linear Mixed-Effects Models (Popular Music and Classical Music)

Popular Music Condition

	Sum of Squares	DF	F	p-*value*
Explicit Information (EI)	7.89	2	7.79	< .001***
Repeated Exposure (RE)	17.42	2	17.20	< .001***
CI * RE	1.34	4	.66	.62

Classical Music Condition

	Sum of Squares	DF	F	p-*value*
Explicit Information (EI)	12.61	2	10.66	< .001***
Repeated Exposure (RE)	.23	2	.19	.82
CI * RE	3.96	4	1.67	.16

Author Note

Manuel Anglada-Tort, Department of Psychology, Goldsmiths, University of London; Daniel Müllensiefen, Department of Psychology, Goldsmiths, University of London, London, United Kingdom.

Manuel Anglada-Tort is now at Department of Audio Communication, Technical University of Berlin, Berlin, Germany.

Correspondence concerning this paper should be addressed to Manuel Anglada-Tort, Department of Audio Communication, Technical University of Berlin, Germany. E-mail: mangl001@gold.ac.uk

Table 1

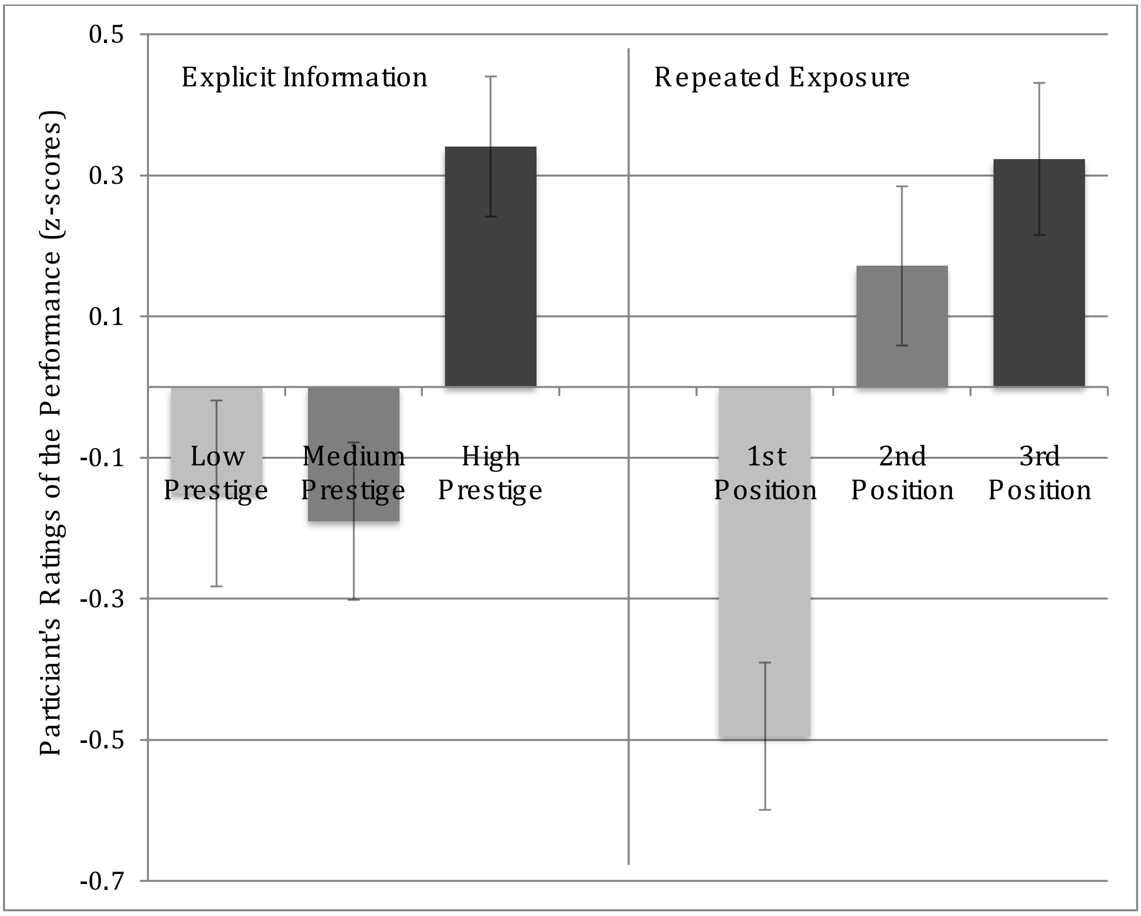
Numbers of Participants Falling for the Repeated Recording Illusion

Did participants fall for the repeated recording illusion?	Yes	%	No	%
Total	52	*75.36*	17	*24.64*
Popular music	56	*81.16*	13	*18.84*
Classical music	59	*85.51*	10	*14.49*
Prestige-suggestion group	43	*80.70*	14	*24.56*
Nonprestige group	9	*75*	3	*25*

Note. Participants were classified as NO if they identified the three recordings as identical in at least one of the two music conditions.

Figure 1

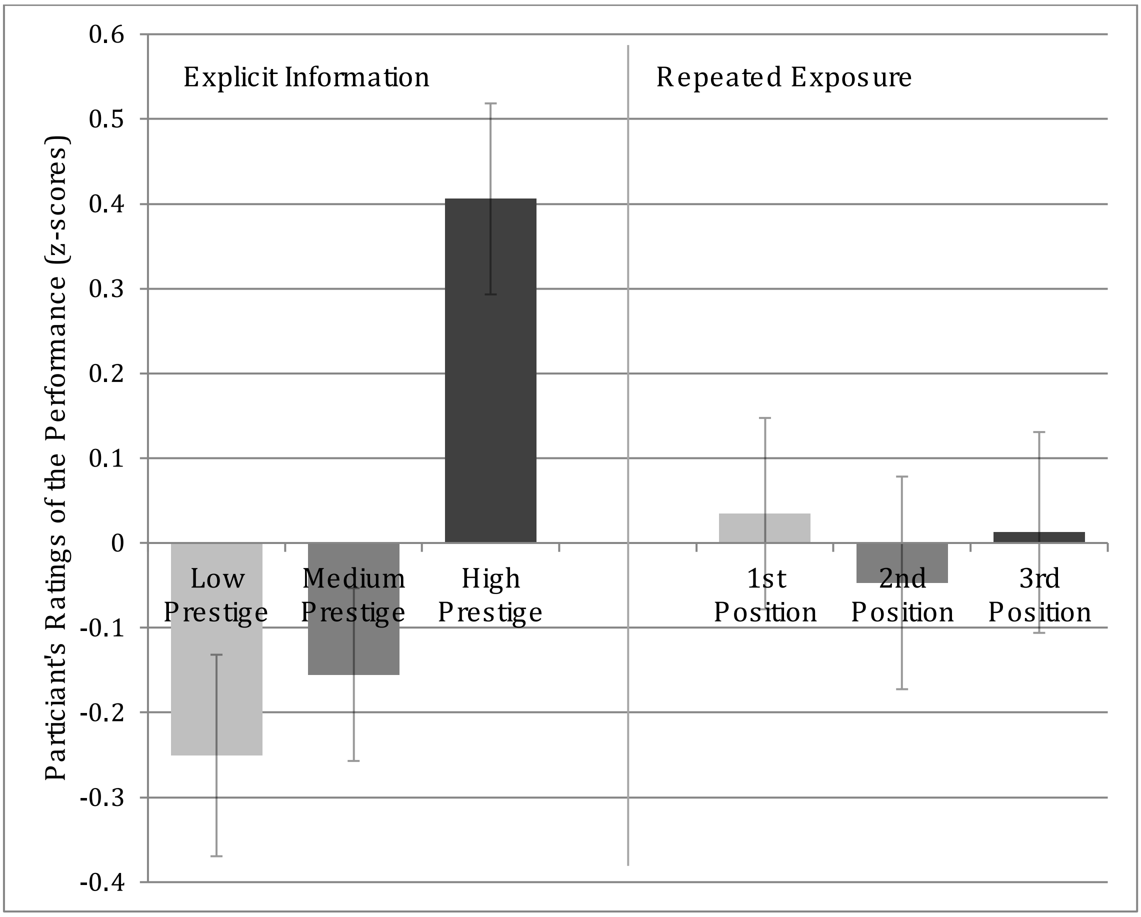
Effects of Explicit Information and Repeated Exposure in the Popular Music Condition



Note. Error bars represent the standard error.

Figure 2

Effects of Explicit Information and Repeated Exposure in the Classical Music Condition



Note. Error bars represent the standard error.

Figure 3

Regression Tree Model