Psychology of Aesthetics, Creativity, and the Arts

Some Like It Sharp: Song Familiarity Influences Musical Preference for Absolute Tuning

Stephen C. Van Hedger, Noah R. Bongiovanni, and Huda Khudhair

Online First Publication, December 14, 2023. https://dx.doi.org/10.1037/aca0000636

CITATION

Van Hedger, S. C., Bongiovanni, N. R., & Khudhair, H. (2023, December 14). Some Like It Sharp: Song Familiarity Influences Musical Preference for Absolute Tuning. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://dx.doi.org/10.1037/aca0000636

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

© 2023 American Psychological Association
ISSN: 1931-3896
https://doi.org/10.1037/aca0000636

Some Like It Sharp: Song Familiarity Influences Musical Preference for Absolute Tuning

Stephen C. Van Hedger[1, 2], Noah R. Bongiovanni[3], and Huda Khudhair[1]
[1] Department of Psychology, Huron University College at Western
[2] Department of Psychology and Western Institute for Neuroscience, University of Western Ontario
[3] Department of Music, University of Notre Dame

Listening to music is an enjoyable activity for most individuals, yet the factors that relate to aesthetic preferences are not completely understood. In the present article, we investigate whether the absolute tuning of music influences listener evaluations of music. Across three experiments, participants rated musical excerpts, tuned conventionally (A4 = 440 Hz) versus unconventionally (±50 cents from conventional tuning), in terms of aesthetic preference. In Experiment 1, participants rated single musical instrument digital interface piano excerpts on each trial in terms of liking, interest, and unusualness. In Experiments 2 and 3, participants heard two versions of the same excerpt on each trial, only differing in terms of tuning, and made a forced-choice judgment as to which version they preferred. Experiment 2 used the same piano excerpts as Experiment 1, whereas Experiment 3 introduced both highly familiar and unknown song excerpts by professional recording artists. Overall, the results suggest that absolute tuning influences aesthetic preferences under limited circumstances. Although there was no strong evidence for tuning influencing judgments in either Experiments 1 or 2, we found a robust effect in Experiment 3 depending on the familiarity of the recording. Whereas participants clearly preferred the conventionally tuned version for highly familiar recordings, they tended to prefer the version that was highest in absolute pitch if the recording was unfamiliar. Overall, these results suggest that absolute tuning can influence musical preferences, although the specific nature of the effect depends on familiarity.

Keywords: aesthetics, music cognition, tuning, absolute pitch, relative pitch

Supplemental materials: https://doi.org/10.1037/aca0000636.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Most individuals find listening to music to be a pleasurable experience, despite the fact that music provides no obvious biological advantages (Mas-Herrero et al., 2014). The ability of music to both represent and evoke emotions in listeners has been discussed for millennia; for example, Plato and Aristotle both discussed how music could regulate emotion and was an integral part of perfecting human nature (Schoen-Nazzaro, 1978). More recent neuroscientific approaches have found that music-induced pleasure is associated with neural signatures that overlap with both primary (e.g., food, sex) and secondary (e.g., money) rewards (Blood & Zatorre, 2001; Blood et al., 1999; Montag et al., 2011), highlighting the importance of music as a pleasure-inducing signal.

How is music able to generate pleasurable, aesthetic reactions in listeners? Although this question at first glance might appear to be ill-formed, as musical preferences might be too variable and idiosyncratic to be characterized more broadly, there have been several complementary contributions to understanding more general aesthetic reactions to music. Rentfrow et al. (2011) discuss a five-factor mellow, urban, sophistication, intensity, and campestral model that is theoretically dissociable from musical genre and can explain individual preferences to music. Other approaches to answering this question have focused on more general features of music, such as predictability and uncertainty (Gold et al., 2019) and perceptions of musical consonance (Bowling & Purves, 2015; McDermott et al., 2010). The focus on more general musical features in eliciting aesthetic responses from listeners is particularly compelling, as it suggests that there are basic auditory features that may contribute to musical pleasure responses regardless of individual differences in musical taste.[1]

In the present article, we investigate whether musical tuning—defined here as the specific mapping of musical notes to auditory frequencies (e.g., A4 = 440 Hz) may serve as an additional feature that influences the aesthetic evaluation of music. Before exploring why

Stephen C. Van Hedger https://orcid.org/00000-0002-2448-9088
Noah R. Bongiovanni https://orcid.org/0000-0002-1615-5352

This research was supported in part through a Huron University College Faculty of Arts and Social Sciences Research Grant awarded to Stephen C. Van Hedger. The authors have no known conflicts of interests to disclose. All materials and data associated with this article can be accessed via Open Science Framework (https://osf.io/zjcvd/ and https://doi.org/10.17605/OSF.IO/ZJCVD).

Correspondence concerning this article should be addressed to Stephen C. Van Hedger, Department of Psychology, Huron University College at Western, 1349 Western Road, London, ON N6G 1H3, Canada. Email: svanhedg@uwo.ca

[1] This is not meant to suggest that accounting for these general features alone will sufficiently explain all patterns of musical preference; indeed, it is clear that a variety of individual differences have the potential to influence aesthetic preferences as well (e.g., North, 2010).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

musical tuning might influence aesthetic responses to music, it is useful to describe current approaches to musical tuning. Most contemporary Western music adheres to equal temperament, in which the octave (a 2:1 frequency ratio) is divided into 12 equal steps (comprising the 12 notes of Western music). Given this approach, the relative change in frequency between notes is fixed and determined by this tuning system (i.e., two adjacent notes are always separated by one-twelfth of an octave, conveniently referenced in logarithmic terms as 100 cents). Thus, if one changes the tuning of a subset of musical notes, this will change the absolute tuning of the changed notes, as well as the relative tuning between the changed and unchanged notes (as the division of the octave into twelve equal steps necessarily would be disrupted). However, if one changes the tuning of all musical notes by a uniform amount, this will only affect absolute tuning, as the relative pitch changes between notes would be preserved. To illustrate this difference between relative and absolute tuning, consider a mischievous piano tuner. If the piano tuner decided to alter the tuning of a single note, then any melody or chord containing this altered note would exhibit altered relative tuning. In contrast, if the piano tuner decided to alter every note on the piano by a fixed amount (e.g., 50 cents), then the relative tuning would remain preserved, as all adjacent notes would still be separated by 100 cents.

Musical tuning has already been shown to influence listener evaluations of music; however, this prior work has focused on relative (rather than absolute) tuning. For example, regardless of formal musical training, listeners appear to be adept at identifying mistuned notes in melodic sequences that are sung (Larrouy-Maestri, 2018; Larrouy-Maestri et al., 2013, 2019) as well as played on an instrument (Hutchins et al., 2012; Lynch & Eilers, 1992; Lynch et al., 1990). Listeners are able to detect mistuned notes in otherwise well-tuned melodies presumably because they have developed rich implicit representations of the typical relative pitch changes in musical sequences (cf. Tillmann & Bigand, 2000). However, the role of absolute tuning (i.e., shifting all notes by a uniform amount) in listener perceptions and evaluations of music remains unexplored.

Intriguingly, there is intentional variability in absolute tuning standards across different musical contexts, suggesting that absolute tuning is at least thought to influence listener experiences of music, even if this question has received little empirical attention. Although most modern Western music recordings adheres to the absolute tuning standard established by the International Standards Organization, in which the "A" above "middle C" (i.e., A4) is tuned to 440 Hz (hereafter referred to as A440 tuning), this value is arbitrary and there are several examples of music that deviates from this standard. Historical music (e.g., Baroque pieces) often adopt more period-appropriate tuning standards, such as A415 tuning, which is an entire semitone (100 cents) lower than A440 and can thus pose initial challenges for listeners with absolute pitch (AP; Lebeau et al., 2020). Additionally, several modern symphonic orchestras tune to different standards for ostensibly aesthetic reasons (e.g., to achieve a more "brilliant" timbre; see Farnsworth, 1968). Some notable examples of this higher, sharpened tuning within the past several decades include the New York Philharmonic (A442/+8 cents higher than A440), the Berlin Philharmonic (A448/+31 cents higher than A440), and the Moscow Symphony Orchestra (A450/+39 cents higher than A440; Abdella, 1989).

This tendency to tune to increasingly higher frequencies reflects a general trend of pitch inflation over the last couple of centuries in Western orchestral music (Lawson & Stowell, 1999); however, it is unclear whether most listeners would be able to appreciate these differences in absolute tuning. Although prior research has demonstrated that listeners might have a preference for sharper tuning in orchestral recordings (Geringer, 1976), the participants in this study were all musicians and some reported possessing AP. Additionally, the tuning of the recordings in this study was manipulated via audiotape, meaning pitch changes also resulted in changes to playback speed (tempo). Thus, it is unclear whether listeners—regardless of formal musical training or possession of AP—would demonstrate tuning-related musical preferences when factors such as tempo are held constant.

It is reasonable to be skeptical that musically untrained listeners would be sensitive to changes in absolute tuning. This is because, as alluded to previously, relative pitches can be maintained regardless of the tuning standard (e.g., a perfect fifth is 700 cents regardless of the absolute frequencies of the notes). However, more recent work has suggested that an implicit or latent memory for AP is a widespread ability regardless of formal musical training. Research in this area has consistently demonstrated that most listeners, regardless of musical training, have an implicit understanding of the musical key of well-known music recordings (Frieler et al., 2013; Jakubowski & Müllensiefen, 2013; Levitin, 1994; Schellenberg & Trehub, 2003; Van Hedger et al., 2023; Van Hedger, Heald, & Nusbaum, 2018). These results are thought to be driven by familiarity with specific recordings; however, more recent research has demonstrated that implicit AP memory is influenced by a broader sense of familiarity with musical notes, learned through statistical regularities in the environment (e.g., Ben-Haim et al., 2014). Specifically, Ben-Haim et al. (2014) had listeners rate isolated musical notes in terms of pleasantness, and found that less common pitch classes (e.g., C#) were rated as significantly more pleasant than more common pitch classes (e.g., C). These findings could not be entirely explained by the pitch inflation hypothesis, as pleasantness ratings were nominally but not significantly correlated with the pitch height of the tested notes. The result from Ben-Haim et al. (2014) thus suggests that musical preferences are influenced in part by how frequently one hears particular musical notes, providing an important theoretical link that implicit statistical learning mechanisms (cf. Saffran et al., 1999) can lead to musical preferences based on AP.

Although the results from Ben-Haim et al. (2014) suggest that AP can influence aesthetic evaluations of musical sounds, the findings appear to be at odds with the mere exposure effect (e.g., Zajonc, 1968), in which more frequently experienced items are more positively evaluated. One potential explanation, provided by the authors, is that listeners might have favored the least familiar notes because the sounds were overall quite simple (isolated notes), and thus the most familiar notes might have not been sufficiently arousing (cf. the "inverted U" model of Berlyne, 1970). Another possibility, supported by subsequent work (e.g., Van Hedger et al., 2017), has to do with how the task is framed. Using a perceptual categorization task which had an objectively correct answer, Van Hedger et al. (2017) found that non-AP listeners could differentiate conventionally tuned (A440) from unconventionally tuned (A453/+50 cents higher than A440) isolated musical notes. This effect was similarly explained by the authors in terms of statistical regularities, as listeners have considerably greater experience listening to conventionally tuned Western music and thus should have developed a representation of category typicality (e.g., Rosch, 1973). However, in Van Hedger et al. (2017), the findings were more consistent with the mere exposure effect, as the categorization task framed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

conventionally tuned notes as "good" and unconventionally tuned notes as "bad." Taken together, the results of Ben-Haim et al. (2014) and Van Hedger et al. (2017) suggest that listeners' implicit familiarity with APs can influence aesthetic evaluations of music, but the specific direction of this effect might depend on how the task is framed. Perhaps most importantly for the present article, both of these studies tested isolated musical notes. As such, it is unclear how these reported findings would extend to musical sounds with greater ecological validity (e.g., excerpts from actual pieces of music), which was hypothesized by Ben-Haim et al. (2014) to potentially influence the relationship between familiarity and affective responses.

The present experiments were therefore designed to assess how infrequently heard, nonstandard tuning might influence listener perceptions and evaluations of music in more ecologically valid contexts. In Experiment 1, participants rated both conventional (A440) and unconventional (+50 cents) musical excerpts in terms of how much they liked the excerpt, how interesting they found the excerpt, and how unusual they found the excerpt. In Experiment 2, participants provided more direct forced-choice judgments of whether they preferred unfamiliar piano excerpts tuned conventionally (A440) versus unconventionally (−50 or +50 cents from conventional tuning). Experiment 3 adopted the same forced-choice paradigm as Experiment 2, but instead used unfamiliar and highly familiar excerpts from popular recording artists. Across all experiments, tuning was not explicitly mentioned to participants, given the association between "mistuned" and negative affective associations (cf. Van Hedger et al., 2017). Moreover, by not explicitly mentioning tuning to participants, these experiments are able to assess whether absolute tuning can implicitly influence aesthetic judgments.

Grounding our hypothesis in the prior literature on implicit AP (e.g., Ben-Haim et al., 2014; Van Hedger et al., 2017), we predicted that listeners across all experiments would differentially rate musical excerpts based on tuning. The specific direction of this effect was not clear, given that prior work has shown both higher (Ben-Haim et al., 2014) and lower (Van Hedger et al., 2017) preferences for less commonly heard musical notes. Additionally, the design of Experiments 2 and 3 allowed us to disentangle absolute tuning preferences (i.e., preferring either conventional or unconventional tuning) from pitch inflation preferences (i.e., preferring the version that was higher in pitch). Overall, these experiments were designed to test whether listeners demonstrated any evidence of absolute tuning-based influences to aesthetic and perceptual evaluations of music using more ecologically valid musical sounds, as this provides a clearer assessment of whether absolute tuning may influence aesthetic experiences outside of constrained experimental contexts.

Experiment 1

Method

Participants

A total of 100 participants were recruited from Amazon Mechanical Turk and 92 were retained for the final analyses ($M = 41.87$ years, $SD = 12.26$ years, range of 20–71 years old). The Data Exclusion section provides details on participant exclusion. The sample size for all experimental conditions in the article was set at $n = 100$ based on two considerations: availability of funds, and a desire to match the recruited samples of Van Hedger et al. (2017), who similarly investigated perceptions of absolute tuning using sounds shifted by the same amount (50 cents) with samples ranging from $n = 94$ to $n = 105$. We used Cloud Research (Litman et al., 2017) to recruit a subset of participants from the larger Mechanical Turk participant pool. Participants were required to have at least a 90% approval rating from a minimum of 50 prior Mechanical Turk assignments, and had to have passed internal attention checks administered by Cloud Research. All participants received monetary compensation upon completion of the experiment. The protocol (#04-202101) was approved by the Research Ethics Board of Huron University College.

Materials

All materials associated with the experiment are provided on Open Science Framework. The experiment was programmed in jsPsych 6 (de Leeuw, 2015). The 30 musical excerpts were selected from four multimovement piano compositions, accessed from an online musical instrument digital interface (MIDI) repository for classical piano music (http://piano-midi.de). Nine excerpts were selected from *Opus 109* (18 Etudes) by Friedrich Burgmüller (1858), seven excerpts were selected from *Petite Suite* by Alexander Borodin (1885), six excerpts were selected from *Opus 165* (España) by Isaac Albéniz (1890), and eight excerpts were selected from *Suite Española* by Isaac Albéniz (1886). These particular pieces were selected because each contained several short piano movements and were hypothesized to be generally unfamiliar to participants. The excerpts were imported into Reason 4 (Propellerhead: Stockholm) as MIDI files and exported using a grand piano timbre. The MIDI files additionally contained expressive cues (including the use of a sustain pedal, dynamic changes in volume, and dynamic changes in expressive timing including tempo, depending on the piece). In-tune excerpts were exported as audio files with Reason's master tuning of Reason set at +0 cents (i.e., A440 tuning), whereas out-of-tune excerpts were exported as audio files with the master tuning set at +50 cents (i.e., adhering to ∼A453 Hz tuning). All musical excerpts had a 44.1 kHz sampling rate and 16-bit depth. Excerpts were then trimmed to 15 s in duration with a 1,000 ms linear fade out and then root-mean-square normalized to −20 dB relative to full scale in Matlab (MathWorks: Natick, Massachusetts).

The auditory calibration stimuli consisted of a 30-s brown noise, generated in Adobe Audition (Adobe: San Jose, California), as well as sine tones meant to assess whether participants were wearing headphones (see Woods et al., 2017). All sine tones were presented in stereo. The "standard" sine tones in phase across stereo channels, whereas the "quiet" sine tone was 180° out-of-phase across the stereo channels. Differentiating the standard and quiet sine tones is easy when the right and left channels are clearly separated (e.g., when wearing headphones) because of phase cancelation but virtually impossible when listening over standard computer speakers.

Procedure

After reading the letter of information and providing informed consent by clicking on a checkbox on the computer, participants completed the auditory calibration. Participants first heard a calibration noise and were instructed to adjust their computer's volume such that the noise was being presented at a comfortable volume. Following the volume adjustment, participants completed the headphone assessment (Woods et al., 2017). The headphone assessment

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

consisted of six trials. On each trial, participants heard three sine tones and had to judge which tone was quietest. Similar to Woods et al. (2017), correctly responding to at least five of six trials was taken as evidence of headphone use.

Next, participants completed the main tuning judgment task. Participants were instructed that they would hear thirty 15-s excerpts of piano music and would be asked to rate each excerpt on a few terms. The tuning of the excerpts was not explicitly mentioned to participants. For each participant, half of the 30 excerpts were randomly assigned to be conventionally tuned (+0 cents) and the other half were selected to be unconventionally tuned (+50 cents). The excerpts were presented in a randomized order. Following each excerpt, participants were asked to rate, on a 100-point slider scale, (a) how much they liked the music, (b) how interesting they found the music, and (c) how unusual they found the music. These questions were selected specifically for the purposes of this study (i.e., they were not taken verbatim from prior studies), and they were designed to form a broader evaluative profile beyond preference. Specifically, the questions assessing how interesting and unusual participants found the music were selected as measures of arousal, which may differ as a function of tuning familiarity and may also influence preference ratings (cf. Ben-Haim et al., 2014). After each excerpt, participants were asked whether they recognized the piece of music and, if so, to provide as many details as they could about the piece title and composer. Two auditory attention checks were added at the end of the tuning judgment task, in which a recording prompted participants to click a specific button on the computer.

Following the tuning judgment task, participants completed a short questionnaire. The questionnaire assessed participants' age, gender, level of education, primary language, hearing aid use, self-reported musical skill, self-reported intonation perception, self-reported AP ability, and formal musical training. Following the questionnaire, participants were given a unique completion code, which they entered into Mechanical Turk to receive compensation for participating.

Data Exclusion

Of the 100 recruited participants, one participant's data were not successfully transferred to our server. Of the remaining 99 participants, we adopted three exclusion criteria: (a) failing at least one of the two auditory attention checks, (b) reporting the use of a hearing aid or otherwise indicating that they had a health concern that might affect the results of the study, or (c) self-identifying as an AP possessor. Five participants were excluded because of failing at least one auditory attention check, no participants were excluded because of the reported use of a hearing aid or disclosing a health-related concern, and two participants were excluded for self-identifying as AP possessors. Thus, 92 participants were included in the main analyses.

Data Analysis

To assess whether the tuning of the musical excerpts influenced participants' ratings, we constructed linear mixed-effect models using the "lme4" package in R (Bates et al., 2020). We generated three models—one using the liking rating as the dependent variable, one using the interesting rating as the dependent variable, and one using the unusual rating as the dependent variable. Each model contained random intercepts for both participant and musical excerpt. Each model also contained a dummy term for tuning (conventional vs. unconventional tuning). To assess the importance of the tuning term in the model, we used both a null hypothesis significance testing (NHST) and a Bayesian approach. For the NHST approach, we calculated p values associated with the tuning terms using package "lmerTest" in R (Kuznetsova et al., 2017). For the Bayesian approach, we compared each model containing the tuning term to a null (intercept-only) model and calculated Bayes Factors (BFs). The reported BF_{10} represent how likely the tuning model is, relative to the null model, given the data. For example, a BF_{10} of 5 would mean that the model containing the tuning term is five times likelier than the null, intercept-only model given the data, whereas a BF_{10} of 0.20 would mean that the model containing the tuning term is one-fifth as likely as the null, intercept-only model given the data. Exploratory correlations between questionnaire measures and performance, which are reported in the online supplemental materials, were assessed via Pearson Product–Moment Correlations and a Bayesian equivalent in JASP.

Results

The Influence of Intonation on Excerpt Ratings

Tuning was not a significant predictor of how much participants reported liking an excerpt, $B = -0.94$, $SE = 0.74$, $p = .204$. On average, participants' mean liking rating was 58.19 ($SD = 12.12$) for conventionally tuned excerpts and 57.42 ($SD = 12.58$) for unconventionally tuned excerpts. The BF_{10} in comparing the tuning model to the intercept-only model was 0.095, meaning the null model was 10.53 times more likely than the tuning model given the data. Tuning was also not a significant predictor of how interesting participants perceived the excerpts, $B = -0.89$, $SE = 0.77$, $p = .248$. On average, participants' mean interesting rating was 57.73 ($SD = 12.07$) for conventionally tuned excerpts and 57.14 ($SD = 12.24$) for unconventionally tuned excerpts. The BF_{10} in comparing the tuning model to the intercept-only model was 0.081, meaning the null model was 12.35 times more likely than the tuning model given the data. Finally, tuning was not a significant predictor of how unusual participants perceived the excerpts, $B = -0.85$, $SE = 0.75$, $p = .257$. On average, participants' mean unusual rating was 39.67 ($SD = 15.95$) for conventionally tuned excerpts and 39.11 ($SD = 15.04$) for unconventionally tuned excerpts. The BF_{10} in comparing the tuning model to the intercept-only model was 0.080, meaning the null model was 12.50 times more likely than the tuning model given the data. Mean ratings are plotted in Figure 1.

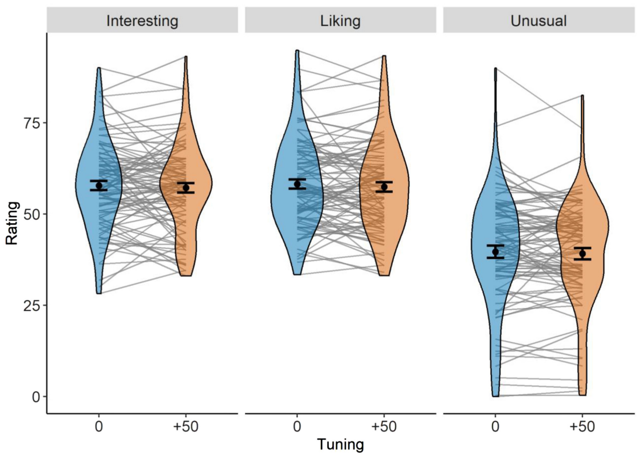
Excerpt Familiarity

Participants reported essentially no prior familiarity with the musical excerpts. Out of the 30 total excerpts, the mean percentage that participants reported to recognize was just 1.23% (i.e., just 0.37 excerpts per participant). The modal number of recognized pieces across participants was zero, and the maximum number of recognized pieces was five. Even when an excerpt was reported to be familiar, no participant was able to provide additional details related to the composer or the name of the piece. As such, it can be concluded that the piano pieces were, as hypothesized, highly unfamiliar to participants.

Discussion

The results of Experiment 1 do not support the hypothesis that absolute tuning influences listener evaluations of musical excerpts.

Figure 1
Violin Plots of Aesthetic Ratings Split by Musical Excerpt Tuning in Experiment 1



Note. Tuning values of 0 represents conventional tuning, whereas tuning values of +50 represent unconventional tuning (50 cents sharper than conventional tuning). Lines connecting the violin plots represent individual participant data. Error bars represent plus or minus one standard error of the mean. See the online article for the color version of this figure.

Participants were asked to rate each excerpt of music in terms of how much they liked the music, how interesting they found the music, and how unusual they found the music. Notably, the tuning of the musical excerpts did not influence these ratings, with the null model (i.e., that did not include the intonation term) being 10.53 to 12.50 times more likely than the tuning model given the data. Musical self-report measures also did not relate to tuning differences in ratings. Finally, perhaps most strikingly, the three ratings were all within a single point of each other for conventionally and unconventionally tuned excerpts on a 100-point scale.

These findings, taken together, suggest that tuning had no measurable influence on listener evaluations of musical excerpts in the present experiment. Although null results cannot determine that tuning has no effect on aesthetic ratings in all circumstances, it is worth discussing why null results were observed in the present experiment, particularly given prior reports of implicit absolute tuning ability (Van Hedger et al., 2017). One likely possibility has to do with the design of the present experiment, which may have encouraged participants to respond to cues other than tuning. Although the ratings themselves were explicit, tuning was never mentioned to participants at any point during the task, and participants only heard a single version of each musical excerpt. As such, participants might have relied on other musical features that were more salient (e.g., tempo, pitch range, tonality) in making their ratings. Put another way, the variability in musical features across the compositions in the stimulus set might have overwhelmed any tuning-driven effects on aesthetic ratings. By adopting a modified paradigm—in which participants hear two versions of the same excerpt (one conventionally tuned, one unconventionally tuned) and must make a forced-choice judgment as to which version they like better—would solve for this issue, as each judgment would control for these musical features unrelated to tuning.

There is also a limitation in the design of Experiment 1, particularly with respect to determining whether listeners might prefer music that has been shifted up in pitch (i.e., consistent with the pitch inflation hypothesis). Specifically, Experiment 1 is unable to disentangle absolute tuning and pitch inflation explanations of tuning preference, as all unconventionally tuned pieces were shifted by +50 cents. Experiment 2 addresses these concerns by modifying the design of Experiment 1 in two critical ways. First, the experimental design was changed such that participants heard two versions of the same excerpt on each trial, with absolute tuning being the only difference between the two. Second, the unconventionally tuned excerpt could either be flat (−50 cents) or sharp (+50 cents) relative to conventional tuning.

Experiment 2

Method

Participants

A total of 100 participants were recruited from Amazon Mechanical Turk and 90 were retained for the final analyses ($M_{\mathrm{AGE}} = 37.72$, $SD_{\mathrm{AGE}} = 10.15$, range of 23–68 years). The Data Exclusion section provides details on participant exclusion. Similar to Experiment 1, we used Cloud Research (Litman et al., 2017) to recruit a subset of participants from the larger Mechanical Turk participant pool. Participants were required to have at least a 90% approval rating from a minimum of 50 prior Mechanical Turk assignments, could not have participated in Experiment 1, and had to have passed internal attention checks administered by Cloud Research. All participants received monetary compensation upon completion of the experiment. The protocol was approved by the Research Ethics Board of Huron University College.

Materials

The experiment was programmed in jsPsych 6 (de Leeuw, 2015). The musical excerpts came from the same files as Experiment 1 and were all created in Reason 4 (Propellerhead: Stockholm) using a piano timbre. Given that the musical sounds were initially generated in MIDI format, the tuning of each sound was altered prior to exporting each as an audio file. Specifically, the in-tune sounds were exported by setting the master tuning option to the default (+0 cents; A440 tuning), whereas the out-of-tune sounds were exported by setting the master tuning option to either +50 cents (∼A453 tuning) or −50 cents (∼A428 tuning). All sounds had a sampling rate of 44.1 kHz and 16-bit depth. Each excerpt was trimmed to 5,000 ms, with a 500 ms linear fade in and fade out. Additionally, to ensure even balancing of trial types (see Procedure), we reduced the number of tested excerpts to 28. All musical stimuli were root-mean-square normalized to −20 dB full spectrum.

Procedure

After reading a letter of information outlining the details of the study and providing informed consent, participants completed the same auditory calibration as reported in Experiment 1. Following the auditory calibration, participants were introduced to the main task. Participants were instructed that they would hear two versions of a music excerpt and would be asked to judge which version they preferred. Tuning was not explicitly mentioned in the instructions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Participants then completed the main task, in which they judged the 28 musical excerpts in a randomized order. On each trial, participants heard two versions of the same musical excerpt, with one of the excerpts always being conventionally tuned and the other excerpt being unconventionally tuned. There were four trial types (seven trials each) to ensure that participants could not rely on general pitch height or excerpt position to identify the conventionally tuned excerpt: (a) conventional first/unconventional (flat) second, (b) conventional first/unconventional (sharp) second, (c) unconventional (flat) first/conventional second, and (d) unconventional (sharp) first/conventional second. Thus, a participant who always selected the first (or second) excerpt across all trials, or a participant who always selected the excerpt highest in pitch, would be choosing the conventionally tuned excerpt exactly 50% of the time. Participants repeated this procedure until all 28 pairs of excerpts had been judged. At the end of the main task, participants were presented with two auditory attention checks, in which they were instructed to click on a specifically marked button on the computer via spoken instructions. Following the main rating task, participants completed a short questionnaire that was identical to Experiment 1. Participants were then provided a unique completion code and compensated.

Data Exclusion

The data exclusion criteria were identical to Experiment 1. Four participants failed at least one of the two auditory attention checks, two of the remaining participants reported the use of a hearing aid, and four of the remaining participants self-identified as possessing AP. This left 90 participants in the primary analysis.

Data Analysis

To assess whether participants systematically chose either (a) the conventionally tuned excerpts or (b) the excerpts highest in pitch (pitch inflation hypothesis), we calculated two mean scores for each participant—the proportion of responses that could be characterized under the AP model (i.e., selecting the conventionally tuned excerpt) and the proportion of responses that could be characterized under the pitch inflation model (i.e., selecting the excerpt highest in pitch). For each trial, responses were coded in a binary fashion (1 or 0) based on adherence to both explanations. For example, if a participant heard a conventional excerpt (± 0 cents) following by an unconventional excerpt (+50 cents) and reported preferring the unconventional excerpt, they would receive a "0" for the *conventional tuning* measure and a "1" for the *pitch inflation* measure for that given trial (as they selected the version highest in AP). Although some trials resulted in identical values (e.g., receiving a "1" or "0" for both the *conventional tuning* and the *pitch inflation* measures depending on the response), across all trials these response strategies were independent. For example, two participants could both achieve a score of 50% for the conventional tuning measure, with one participant achieving 100% for the pitch inflation measure (i.e., always selecting the highest excerpt) and another participant achieving 0% for the pitch inflation measure (i.e., always selecting the lowest excerpt). The independence of these response strategies was assessed in the present experiment through Pearson Product–Moment Correlations and a Bayesian equivalent in JASP. The two measures (proportion of conventional tuning responses and proportion of pitch inflation responses) were analyzed via one-sample t tests against a known mean of 50% and a Bayesian equivalent, performed in JASP. Thus, all analyses report both p values and BFs. Exploratory correlations between questionnaire measures and performance, which are reported in the online supplemental materials, were assessed via Pearson Product–Moment Correlations and a Bayesian equivalent in JASP. Normality of the response distributions were assessed using the Kolmogorov–Smirnov test. Both the AP ($p = .275$) and the pitch inflation ($p = .107$) response distributions were normal.

Results

Analysis of Absolute Tuning Versus Pitch Inflation Responses

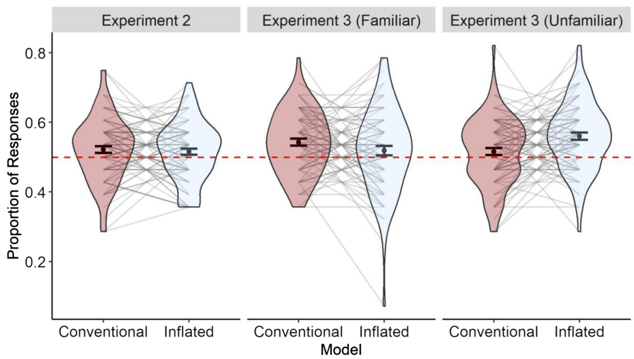
Participants selected the conventionally tuned excerpt 52.2% ($SD = 9.1\%$) of the time, $t(89) = 2.27$, $p = .026$, $d = 0.24$, and selected the excerpt highest in pitch 51.5% ($SD = 8.5\%$) of the time, $t(89) = 1.73$, $p = .087$, $d = 0.18$. Although both values were nominally above the chance estimate of 50%, with the absolute tuning explanation reaching statistical significance, these results should be interpreted cautiously for several reasons. First, the effect sizes were small (Cohen's d of 0.24 and 0.18, respectively). Second, neither explanation was strongly supported in a Bayesian framework. The BF_{10} for the absolute tuning explanation was 1.32, meaning the alternative hypothesis was only 1.32 times more likely than the null hypothesis. The BF_{10} for the pitch inflation explanation was 0.49, meaning the alternative hypothesis was 0.49 times more likely than the null hypothesis (i.e., the null hypothesis was approximately twice as likely as the alternative hypothesis). These results are plotted in Figure 2. Confirming the independence of response strategies, responses adhering to the absolute tuning model were not correlated with responses adhering to the pitch inflation model, $r(88) = .14$, $p = .191$, $BF_{10} = 0.306$.

Discussion

Similar to Experiment 1, the results of Experiment 2 suggest that listeners do not show strong preferences for pieces of unfamiliar music based on absolute tuning. Despite the use of a more direct paradigm, in which participants selected which of two musical excerpts they preferred (with the two excerpts only differing in terms of tuning), participants only showed weak evidence for systematically selecting versions that were conventionally tuned. Moreover, the effect size of this finding was small ($d = 0.26$) and the BF was equivocal (i.e., it did not suggest strong evidence for either the null or the alternative hypothesis).

One reason why the present experiment might have found equivocal results is because of the nature of the musical stimuli. Specifically, we used MIDI renderings of piano excerpts, selected to be unfamiliar to participants. As such, it is unclear whether the present findings would generalize to (a) familiar recording excerpts or (b) non-MIDI recordings. It is possible that using highly familiar musical recordings, similar to other investigations of implicit AP memory (Jakubowski & Müllensiefen, 2013; Schellenberg & Trehub, 2003; Van Hedger, Heald, & Nusbaum, 2018), would have afforded participants the ability to judge the music based on how the current tuning adheres to (or deviates from) their prior experiences with that recording. Beyond recording familiarity, it is also

Figure 2
Violin Plots of Responses Adhering to Different Pitch Models in Experiments 2 and 3



Note. Correct refers to responses in which the conventionally tuned excerpt was selected. Inflated refers to responses in which the excerpt highest in pitch was selected. Lines connecting the violin plots represent individual participant data. Error bars represent plus or minus one standard error of the mean. See the online article for the color version of this figure.

possible that tuning-based preference judgments would show greater sensitivity for non-MIDI recordings. MIDI offers experimental control at the expense of ecological validity (e.g., Dieleman et al., 2018), and thus, despite selecting MIDI files that contained expressive performance parameters (e.g., sustain pedal, expressive timing), it might have been the case that participants found both versions of the excerpt to be unfavorable, which could have encouraged random selection that was not based on tuning.

Experiment 3 was designed to address these two primary limitations. Rather than using MIDI excerpts, participants listened to audio recording excerpts from well-known recording artists. Some participants listened to highly familiar song excerpts (familiar condition), whereas other participants listened to recordings that were specifically selected to be unfamiliar (unfamiliar condition), despite being chosen from the same recording artists as the familiar condition. As such, Experiment 3 provides a more direct assessment of the influence of absolute tuning on aesthetic preference in more ecologically valid listening situations, and additionally examines how tuning might interact with piece familiarity.

Experiment 3

Method

Participants

A total of 200 participants were recruited from Amazon Mechanical Turk and 175 were retained for the final analyses (familiar condition: $n = 83$; unfamiliar condition: $n = 92$; $M_{\text{AGE}} = 39.95$, $SD_{\text{AGE}} = 11.05$, range of 21–70 years). Similar to Experiments 1 and 2, we used Cloud Research (Litman et al., 2017) to recruit a subset of participants from the larger Mechanical Turk participant pool. Participants were required to have at least a 90% approval rating from a minimum of 50 prior Mechanical Turk assignments, could not have participated in Experiments 1 or 2, and had to have passed internal attention checks administered by Cloud Research. All participants received monetary compensation upon completion of the experiment. The protocol (#21-202112) was approved by the Research Ethics Board of Huron University College.

Materials

The experiment was programmed in jsPsych 6 (de Leeuw, 2015). The musical excerpts, which came from well-known recording artists, were selected by two of the authors (Huda Khudhair and Stephen C. Van Hedger) based on the (un)familiarity of the song. Both sets of excerpts (familiar, unfamiliar) were carefully matched in terms of artist and genre. For example, an excerpt from "Firework" by Katy Perry was included in the familiar excerpt set, and this was balanced with a relatively unfamiliar recording by Katy Perry ("Into Me You See") in the unfamiliar excerpt set. This approach was continued for all stimuli in both sets, meaning that both the unfamiliar and familiar excerpts consisted of the same recording artists. Tuning was manipulated using the "Change Pitch" effect in Audacity, with the "high quality stretching" option selected (exactly preserving the length of the excerpts). Additionally, all excerpts were subjected to the pitch-shifting algorithm the same number of times. The 50-cent sharp excerpts were shifted up in pitch by 25 cents twice, the 50-cent flat excerpts were shifted down in pitch by 25 cents twice, and the conventionally tuned excerpts were first shifted up in pitch by 25 cents and then shifted down in pitch by 25 cents. All sounds had a sampling rate of 44.1 kHz and 16-bit depth. Each excerpt was trimmed to 5,000 ms, with a 500 ms linear fade in and fade out. There were 28 familiar excerpts and 28 unfamiliar excerpts. All musical stimuli were root-mean-square normalized to −20 dB full spectrum.

Procedure

The procedure was nearly identical to Experiment 2, with a couple of exceptions. First, participants were judging familiar or unfamiliar song excerpts (as a between-participant condition), rather than MIDI versions of unknown piano excerpts. Second, the postpreference familiarity rating was slightly modified from Experiment 2. In the present experiment, immediately following the preference judgment, participants rated their familiarity with the recording on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extremely*). Given that Experiment 3 was designed to assess how song familiarity interacted with absolute tuning, we discarded trials where participants reported no familiarity (familiar condition) and trials where participants reported some prior familiarity (unfamiliar condition). Participants were removed if too many trials were removed (see Data Exclusion section). Of the included participants, the recognition rate was 85.5% in the familiar condition and 12.9% in the unfamiliar condition, suggesting that the recordings selected by the authors were representative of familiar and unfamiliar categories, respectively.

Data Exclusion

The data exclusion criteria were identical to Experiments 1 and 2 with one additional consideration: participants had to report familiarity with at least 50% of the excerpts (familiar condition) or no familiarity with at least 50% of the excerpts (unfamiliar condition). Twelve participants were excluded from analysis in the familiar condition for being below the 50% threshold, and four participants were excluded from analysis in the unfamiliar condition for reporting familiarity above the 50% threshold. After applying the additional

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

exclusion criteria from Experiments 1 and 2, 83 and 92 participants were considered for analysis in the familiar and unfamiliar conditions, respectively.

Data Analysis

For each participant, we calculated two mean scores—the proportion of responses that could be characterized under the AP model (i.e., selecting the conventionally tuned excerpt) and the proportion of responses that could be characterized under the pitch inflation model (i.e., selecting the excerpt highest in pitch). The data analysis plan was nearly identical to Experiment 2, with two additional analyses. The first addition was an independent-samples *t* test and Bayesian equivalent to assess, as a manipulation check, whether the familiar and unfamiliar excerpts differed in terms of self-reported familiarity. The second additional analysis was a 2 × 2 analysis of variance (ANOVA) and Bayesian equivalent, which formally assessed how participants' response types (AP, pitch inflation) interacted with condition (familiar excerpts, unfamiliar excerpts). Similar to Experiment 2, we assessed normality of the response distributions using the Kolmogorov–Smirnov test. For both conditions, the AP and pitch inflation response types were normal (all $ps > .545$).

Results

Familiar Song Excerpts

Participants selected the conventionally tuned excerpt 54.8% ($SD = 9.2\%$) of the time, which was significantly above the chance estimate, $t(82) = 4.77$, $p < .001$, $d = 0.52$. In contrast, participants selected the excerpt highest in pitch 51.6% ($SD = 12.4\%$) of the time, which was not above the chance estimate, $t(82) = 1.17$, $p = .246$, $d = 0.13$. The BF_{10} for the absolute tuning explanation of the data were 2,088, meaning the alternative hypothesis (i.e., that participant preferred the correctly tuned version of the excerpt) was 2,088 times more likely than the null hypothesis given the data. In contrast, the BF_{10} for the pitch inflation hypothesis was 0.23, meaning the null hypothesis was approximately four times as likely as the alternative hypothesis. Responses adhering to the absolute tuning model were not correlated with responses adhering to the pitch inflation model, $r(81) = .05$, $p = .656$, $BF_{10} = 0.151$.

Unfamiliar Song Excerpts

Participants selected the conventionally tuned excerpt 51.7% ($SD = 9.7\%$) of the time, which did not differ from the chance estimate, $t(91) = 1.71$, $p = .090$, $d = 0.18$. In contrast, participants selected the excerpt highest in pitch 56.0% ($SD = 10.1\%$) of the time, which was significantly above the chance estimate, $t(91) = 5.70$, $p < .001$, $d = 0.59$. The BF_{10} for the absolute tuning explanation of the data were 0.47, meaning the null hypothesis was about twice as likely than the alternative hypothesis given the data. In contrast, the BF_{10} for the pitch inflation hypothesis was 86,929, meaning the alternative hypothesis (i.e., that participants systematically preferred excerpts highest in pitch) was 86,929 times more likely than the null hypothesis given the data. Responses adhering to the absolute tuning model were not correlated with responses adhering to the pitch inflation model, $r(90) = .02$, $p = .839$, $BF_{10} = 0.133$.

Combined Analysis of Familiar and Unfamiliar Excerpt Conditions

The results reported in the prior sections suggest that participants form preferences based on different cues depending on song familiarity. To formally test this, a 2 (cue: absolute, pitch inflation) × 2 (condition: familiar, unfamiliar) ANOVA and Bayesian equivalent were constructed. In these analyses, the BF reflects the extent to which the data support the inclusion of the term in the model. In this model, there was no main effect of cue, $F(1, 173) = 0.22$, $p = .639$, $\eta_p^2 = .001$, $BF = 0.148$, nor was there a main effect of condition, $F(1, 173) = 0.33$, $p = .566$, $\eta_p^2 = .002$, $BF = 0.153$. There was, however, a significant interaction of cue and condition, $F(1, 173) = 11.78$, $p < .001$, $\eta_p^2 = .064$, $BF = 70.66$, with participants in the familiar condition preferring the excerpt version that was conventionally tuned relative to participants in the unfamiliar condition, and participants in the unfamiliar condition preferring the excerpt version that was highest in pitch relative to participants in the familiar condition. These results are plotted in Figure 2.

Discussion

The results of Experiment 3 suggest that absolute tuning influences preference for more ecologically rich musical excerpts, with the specific influence depending on excerpt familiarity. When listening to familiar songs, participants preferred the conventionally tuned version (i.e., the version they would encounter outside of the testing context). In contrast, when listening to unfamiliar songs, participants systematically preferred the version that was highest in pitch, supporting the pitch inflation hypothesis. The results from Experiment 3 thus clearly establish that absolute tuning influences participant preferences depending on familiarity, which is particularly notable given the controlled construction of the stimuli sets. Each stimulus was pitch shifted, trimmed, and normalized in the same way to control for unintended acoustic cues across stimuli. Additionally, the familiar-unfamiliar pairings were deliberately taken from the same artist, which provides strong control with respect to factors such as vocal timbre and genre.

General Discussion

The present set of experiments were designed to assess whether absolute tuning (i.e., the mapping of musical notes to specific, absolute frequencies while preserving relative pitch structure) influences preferences for music. Although we found evidence that absolute tuning influences listener evaluations of music, these effects were (a) limited to ecologically rich (non-MIDI) musical recordings and (b) depended on familiarity with the musical recording. The contextually limited nature of these tuning influences on preference are highlighted by the results from Experiments 1 and 2, which were either null or equivocal. In Experiment 1, we found no tuning-related influence on participants' evaluations of unknown MIDI piano excerpts in terms of liking, interest, or unusualness, with null models being approximately 10–12 times likelier than models including absolute tuning. Despite using a more direct paradigm in Experiment 2, in which participants heard two excerpts differing only in absolute tuning and made a forced-choice preference judgment, we found no strong evidence that participants had systematic preferences based on tuning. These null and equivocal findings make the results from

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Experiment 3, which used an identical paradigm as Experiment 2 but tested more ecologically valid musical recordings, all the more striking. In Experiment 3, we found strong evidence that participants systematically preferred conventionally tuned versions of familiar recordings from popular artists. In contrast, when presented with unfamiliar recordings from the same recording artists, participants systematically preferred the higher-tuned version of the song, not the version adhering to conventional tuning. These findings thus suggest that listeners form contextual and dynamic preferences for musical recordings based on absolute tuning, with the specific nature of these preferences depending on factors such as recording type (audio recordings of contemporary pop music vs. MIDI versions of piano music) and familiarity with the specific recording.

The finding from Experiment 3 that participants systematically prefer the conventionally tuned versions of familiar popular recordings conceptually aligns with previous studies on latent AP memory, in which participants can determine which version of a familiar piece of music is presented at the correct AP (e.g., Jakubowski et al., 2017; Schellenberg & Trehub, 2003; Van Hedger, Heald, Uddin, & Nusbaum 2018). However, the present findings are notable when compared to this prior work for at least two reasons. First, the "incorrect" version (i.e., the unconventionally tuned version) in the present study was only 50 cents removed from the correct version, highlighting the remarkable precision of listeners' representations of familiar recordings. To put this in context, the tuning of the "incorrect" version in prior work (e.g., Schellenberg & Trehub, 2003) was either 100 or 200 cents—that is, two to four times the amount in the present experiment. Second, and perhaps most notably, the findings demonstrate that the present paradigm can serve as an accurate implicit measure of pitch memory. Participants in the present study were simply asked to select which version of the song they liked more; as such, there were no "correct" or "incorrect" answers to the questions in the task. It is thus notable that participants preferred the conventionally-tuned versions of these familiar songs. A robust body of literature supports the notion that familiarity with a particular stimulus correlates with an increased affinity for it (Bornstein, 1989; Fang et al., 2007; Zajonc, 1968); this general phenomenon, called the "mere exposure effect," has been confirmed specifically in the musical contexts as well (Schubert, 2007). As a result, a plausible explanation for the present findings is that participants, having attended to multiple playings of the songs with which they are familiar, also become implicitly habituated to the conventional tuning standard in which they experience these recordings, and thus show preferences for this familiar, conventional tuning.

In contrast, participants displayed a clearly dissociable pattern of results when the recording was unfamiliar. In these cases, participants ostensibly cannot reference prior experience with a particular song in the process of making a preference-based choice and instead seem to be drawn to the highest option, supporting the pitch inflation hypothesis. The affinity that people show for sharper-pitched music is conceptually supported from prior findings. For example, Geringer (1976) found that when given the choice to tune excerpts of classical music to their preferred pitch, people show a significant propensity for pitch levels an average of 1.5 semitones sharper than the original recording. One potential mechanism for why this might be the case comes from Eitan and Timmers (2010), who found that listeners associated increased pitch height with increased ratings of intensity and brightness. Even beyond the scientific literature, this association between pitch sharpness and brightness is well-documented historically. Effects of this belief manifested in the relative increasing of the tuning note "A" over the last several centuries, attributed to the continual desire by composers and conductors to achieve a more "brilliant" sound than their peers both past and present (Anthon, 1941; Farnsworth, 1968; Rosenberg, 2021). The present study, along with previous works, affirms that this preference for higher, "brighter" pitched music in novel settings is robust.

Although the findings from Experiment 3 clearly establish dissociable patterns of tuning-based preference depending on familiarity, it is important to comment on why these effects might have only been observed in Experiment 3. Perhaps the most likely explanation for the nonsignificant or equivocal results of the first two experiments is that they were not particularly well-suited for preference judgments. First, despite the use of expressive performance parameters (e.g., sustain pedal and expressive timing) in the present study, the use of MIDI recordings reduces ecological validity and may limit aesthetic expression (Dieleman et al., 2018). Second, the genre of the musical excerpts was essentially homogeneous—each excerpt was sourced from a selection of 19th-century classical piano works, which is not likely to be a particularly revered genre among a sample of online participants not specifically selected for their musical backgrounds. Consequently, even in Experiment 2, which used the same design as Experiment 3, the artificial MIDI nature of the sounds, and the homogeneity and niche status of the musical genre may have interjected noise into participants' preference (as participants may have felt apathetic toward the excerpts in general and might have thus been likelier to randomly respond).

Another possibility, which is not based on the (un)suitability of the stimuli from Experiments 1 and 2 for preference judgments per se, is that the excerpts only used a single musical instrument (piano). In contrast, both the familiar and unfamiliar excerpts from Experiment 3 were multiinstrumental. The use of multiple instruments significantly increased the perceptual centroid of the excerpts (see the online supplemental materials), which may have also changed the way in which pitch shifting influenced perceived brightness. Put another way, the excerpts from Experiments 1 and 2 might have shown more evidence for the pitch inflation hypothesis if we had used multiinstrumental (e.g., orchestral) recordings with similar perceptual centroids as the recordings from Experiment 3. At present, the significantly lower perceptual centroids might have meant that subtle pitch shifts of 50 cents were not sufficient to influence the perceived brightness of the sound.

As an avenue for future work, these potentially confounding variables differentiating the piano recordings from the pop song recordings—MIDI rendering, genre appeal, and (multi)instrumentation—could be dissociated to analyze their effects on absolute tuning judgment accuracy. For example, one follow-up approach could use MIDI renderings of the stimulus set from Experiment 3 to assess whether the findings replicate; another follow-up could examine whether absolute tuning effects manifest for (non-MIDI) audio recordings of classical piano music, or other genres (ideally containing multiple instruments) with which participants are unlikely to have significant listening experience accrued.

Beyond the myriad differences between MIDI piano excerpts and recordings from popular artists, which limit the extent to which we can draw strong conclusions across all three experiments, there are a couple of other limitations to acknowledge. First, the choice of adopting a forced-choice paradigm in Experiments 2 and 3 might have encouraged participants to more explicitly consider tuning in their preference responses. Although Experiments 2 and 3 emphasized in the instructions that there were no correct answers, and that participants should simply select the version they preferred, future work might consider more nuanced means of assessing preference or changes to the experimental design to assess the extent to which these tuning-based preferences are implicit.

Second, the approach taken in Experiment 3 to maximize (un) familiarity of the recordings leaves open the question of how familiarity—treated as a more graded variable—might influence participant responses. Although every song in the familiar category was of course not equally familiar to each person, all recordings had been selected based on their presumed high levels of familiarity and presence in popular culture. In future studies, it would be worthwhile to specifically analyze how graded or relative differences in familiarity levels (e.g., "less familiar" vs. "more familiar") influence perception and preference, which could provide more detailed information about the role that familiarity plays in preference for conventionally-tuned recordings. If tuning preferences operate similarly to implicit AP memory for familiar recordings (cf. Schellenberg et al., 2019), then it might take only a few experiences with a previously unfamiliar recording before individuals begin to prefer the version that is presented in the "correct" tuning.

Third, the online nature of the experiments afforded no direct interaction with or monitoring of the participants. However, to mitigate the effects of unsupervised delivery, we incorporated attention checks to ensure that our participants were engaged, with participants incorrectly answering attention checks being discarded from analysis. Furthermore, the recruitment parameters used in CloudResearch allowed for additional layers of data quality assurances, including only recruiting participants who had previously passed internal attention checks administered by CloudResearch. Additionally, prior studies conducted by the authors (e.g., Van Hedger et al., 2017) have shown internal replications of effects in both online and in-lab samples, using similar recruitment parameters as those used in the present study. Finally, the literature asserts the benefits of online studies more broadly given that they are carefully controlled (e.g., Eerola et al., 2021; Zhao et al., 2022). As such, we would expect these general findings to replicate in an in-person, laboratory setting.

Conclusion

Most listeners find music to be aesthetically pleasing; however, the specific features that relate to aesthetic evaluations of music are a subject of active investigation. The present set of experiments assessed whether absolute tuning could influence musical preference among listeners not specifically recruited for their musical expertise or AP ability. Whereas Experiments 1 and 2 suggested null or equivocal effects of absolute tuning on musical preference, these two experiments used MIDI renditions of unfamiliar piano excerpts, lowering the ecological validity of the findings. In contrast, when using excerpts of audio recordings from a pop music genre, we found that participants systematically preferred the conventionally tuned versions of familiar songs and systematically preferred the higher-tuned versions of unfamiliar songs (Experiment 3). These findings support the argument that non-AP possessors have some level of implicit AP memory for familiar music, and that this memory has the capacity to influence aspects of musical preference. The data also support the pitch inflation hypothesis, particularly for listening contexts in which there is no prior familiarity with the piece of music. Taken together, then, we find that preferences for music systematically relate to absolute tuning in complex ways. Notably, this relationship is strongest in cases where both ecological validity and familiarity with the stimulus are demonstrated.

References

Abdella, F. T. (1989). As pitch in Opera rises, so does debate. *The New York Times*.

Anthon, C. (1941). *Absolute pitch*. The Atlantic. https://www.theatlantic.com/magazine/archive/1947/04/absolute-pitch/655473/

Bates, D., Maechler, M., Bolker, B., Walker, S., Haubo, R., Christensen, B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., & Fox, J. (2020). *Package "lme4"* (Lme4.r-Forge.r-Project.Org).

Ben-Haim, M. S., Eitan, Z., & Chajut, E. (2014). Pitch memory and exposure effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 24–32. https://doi.org/10.1037/a0033583

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5-A), 279–286. https://doi.org/10.3758/BF03212593

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 of the United States of America*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 382–387. https://doi.org/10.1038/7299

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289. https://doi.org/10.1037/0033-2909.106.2.265

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), 11155–11160. https://doi.org/10.1073/pnas.1505768112

de Leeuw, J. R. (2015). Jspsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavioral Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dieleman, S., van den Oord, A., & Simonyan, K. (2018). The challenge of realistic music generation: Modelling raw audio at scale. *ArXiv*. https://doi.org/10.48550/arxiv.1806.10474

Eerola, T., Armitage, J., Lavan, N., & Knight, S. (2021). Online data collection in auditory perception and cognition research: Recruitment, testing, data quality and ethical considerations. *Auditory Perception & Cognition*, *4*(3–4), 251–280. https://doi.org/10.1080/25742442.2021.2007718

Eitan, Z., & Timmers, R. (2010). Beethoven's Last Piano Sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–22. https://doi.org/10.1016/j.cognition.2009.10.013

Fang, X., Singh, S., & Ahluwalia, R. (2007). An examination of different explanations for the mere exposure effect. *Journal of Consumer Research*, *34*(1), 97–103. https://doi.org/10.1086/513050

Farnsworth, P. R. (1968). *The social psychology of music*. The Iowa State University Press.

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*(3), 334–349. https://doi.org/10.1177/1029864913493802

Geringer, J. M. (1976). Tuning preferences in recorded orchestral music. *Journal of Research in Music Education*, *24*(4), 169–176. https://doi.org/10.2307/3345127

Gold, B. P., Pearce, M. T., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

music: A reward for learning? *Journal of Neuroscience*, *39*(47), 9397–9409. https://doi.org/10.1523/JNEUROSCI.0428-19.2019

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*(2), 147–159. https://doi.org/10.1525/mp.2012.30.2.147

Jakubowski, K., & Müllensiefen, D. (2013). The influence of music-elicited emotions and relative pitch on absolute pitch memory for familiar melodies. *Quarterly Journal of Experimental Psychology*, *66*(7), 1259–1267. https://doi.org/10.1080/17470218.2013.803136

Jakubowski, K., Müllensiefen, D., & Stewart, L. (2017). A developmental study of latent absolute pitch memory. *Quarterly Journal of Experimental Psychology*, *70*(3), 434–443. https://doi.org/10.1080/17470218.2015.1131726

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Larrouy-Maestri, P. (2018). "I know it when I hear it": On listeners' perception of mistuning. *Music & Science*, *1*), https://doi.org/10.1177/2059204318784582

Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, *51*(2), 663–675. https://doi.org/10.3758/s13428-019-01225-1

Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2013). The evaluation of singing voice accuracy: A comparison between subjective and objective methods. *Journal of Voice*, *27*(2), 259.e1–259.e5. https://doi.org/10.1016/j.jvoice.2012.11.003

Lawson, C., & Stowell, R. (1999). *The historical performance of music: An introduction*. Cambridge University Press.

Lebeau, C., Tremblay, M. N., & Richer, F. (2020). Adaptation to a new tuning standard in a musician with tone-color synesthesia and absolute pitch. *Auditory Perception & Cognition*, *3*(3), 113–123. https://doi.org/10.1080/25742442.2021.1886846

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/BF03206733

Litman, L., Robinson, J., & Abberbock, T. (2017). Turkprime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics*, *52*(6), 599–608. https://doi.org/10.3758/BF03211696

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*(4), 272–276. https://doi.org/10.1111/j.1467-9280.1990.tb00213.x

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

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*(11), 1035–1041. https://doi.org/10.1016/j.cub.2010.04.019

Montag, C., Reuter, M., & Axmacher, N. (2011). How one's favorite song activates the reward circuitry of the brain: Personality matters!. *Behavioural Brain Research*, *225*(2), 511–514. https://doi.org/10.1016/j.bbr.2011.08.012

North, A. C. (2010). Individual differences in musical taste. *The American Journal of Psychology*, *123*(2), 199–208. https://doi.org/10.5406/amerjpsyc.123.2.0199

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. https://doi.org/10.1037/a0022406

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosenberg, R. (2021). Perfect pitch. *Journal of Popular Music Studies*, *33*(1), 137–154. https://doi.org/10.1525/jpms.2021.33.1.137

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*(3), 262–266. https://doi.org/10.1111/1467-9280.03432

Schellenberg, E. G., Weiss, M. W., Peng, C., & Alam, S. (2019). Fine-grained implicit memory for key and tempo. *Music & Science*, *2*, Article 205920431985719. https://doi.org/10.1177/2059204319857198

Schoen-Nazzaro, M. B. (1978). Plato and Aristotle on the ends of music. *Laval Théologique et Philosophique*, *34*(3), 261–273. https://doi.org/10.7202/705684ar

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music*, *35*(3), 499–515. https://doi.org/10.1177/0305735607072657

Tillmann, B., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*(4), 885–913. https://doi.org/10.1037/0033-295X.107.4.885

Van Hedger, S. C., Bongiovanni, N., & Khudhair, H. (2023). *Some like it sharp: Song familiarity influences musical preference for absolute tuning* [Open Science Framework Repository]. https://doi.org/10.17605/OSF.IO/ZJCVD

Van Hedger, S. C., Heald, S. L. M., Huang, A., Rutstein, B., & Nusbaum, H. C. (2017). Telling in-tune from out-of-tune: Widespread evidence for implicit absolute intonation. *Psychonomic Bulletin and Review*, *24*(2), 481–488. https://doi.org/10.3758/s13423-016-1099-1

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2018). Long-term pitch memory for music recordings is related to auditory working memory precision. *The Quarterly Journal of Experimental Psychology*, *71*(4), 879–891. https://doi.org/10.1080/17470218.2017.1307427

Van Hedger, S. C., Heald, S. L. M., Uddin, S., & Nusbaum, H. C. (2018). A note by any other name: Intonation context rapidly changes absolute note judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(8), 1268–1282. https://doi.org/10.1037/xhp0000536

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

Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt. 2), 1–27. https://doi.org/10.1037/h0025848

Zhao, S., Brown, C. A., Holt, L. L., & Dick, F. (2022). Robust and efficient online auditory psychophysics. *Trends in Hearing*, *26*, Article 23312165221118792. https://doi.org/10.1177/23312165221118792

Received June 1, 2023
Revision received July 27, 2023
Accepted July 29, 2023 ■