Abstract

Human imagination is generative and creative yet deeply rooted in culture and familiarity. Recent studies have quantified the effects of culture on stories that are imagined during music listening, but the music used in previous work was always drawn from a tradition familiar to participants from at least one of the cultures. Here we report the first study of imagined stories to music written in a musical system that is novel to participants from each culture, thus allowing for a direct comparison of narratives prompted by the same set of excerpts that is comparably unfamiliar to both groups. Music composed in the Bohlen-Pierce scale was presented to participants from two geographically-defined cultures: Boston, USA and Beijing, China. We also examined how individual differences, such as in musicality and sensitivity to musical reward, might affect narrative engagement and semantic content of the imagined stories as measured by tools from natural language processing. Results showed that semantic spaces of music-evoked imaginings differed between Boston and Beijing cohorts. While both cultures were similarly engaged by the story response task, differences emerged in the semantic content of the imagined stories. . Boston participants who reported being more absorbed by music wrote more unconventional stories, whereas Beijing participants who reported more emotional responses to music wrote more conventional stories. These results reveal the roles of culture and individual differences in modes of narrative engagement and imagination during music listening.

Keywords

Imagination; creativity; scale; music; culture; individual differences

Introduction

Imagination is a core cognitive capacity that underlies human artistic, scientific, and technical creation (Finn & Wylie, 2021; Vygotsky, 2004). Far from being a passive idling of the mind, imagination is a cognitive act that involves generative thought, and relies on neurocognitive mechanisms, such as perception, attention, and memory, that have developed throughout a lifespan of culturally organized experiences (Pelaprat & Cole, 2011). It can involve thinking about things as they were or might be but are not in the present environment (Liao and Gendler, 2019), and can call on mental imagery in any modality, whether conscious or unconscious, voluntary or involuntary (Nanay, 2023).

Human imagination is also deeply rooted in culture (Zittoun & Gillespie, 2015). Culture has been deemed "the collective programming of the mind" that distinguishes the members of one group or category of people from others (Hofstede, 2001). The contents of imagination, in particular, often draw from resources such as stories, ideas, and images derived from cultural artifacts (Zittoun & Cerchia, 2013). In that sense, culture serves as a mediator of imagination, and cultural artifacts contribute to the structure and content of imagined stories (Vygotsky, 2004; Pelaprat & Cole, 2011).

Music plays an important role in both reflecting and shaping culture. Composers and musical artists have long drawn from the structural and acoustic features associated with particular musical traditions to convey a sense of cultural identity. For example, the nineteenth-century composer Claude Débussy used piano techniques such as pedaling to simulate the timbre of the Javanese gamelan in *Pagodes* (Parker, n.d.); K-pop artists moved from the more traditional pentatonic scale to diatonic scales to signal Western urbanization (Lie, 2012); and choices in instrumentation, melodic, rhythmic, and scale structures are all used in rap and hip hop music to

project the cultural identity of marginalized minorities in Cuba, France, Italy, New Zealand and South Africa (Bennett, 2017). Empirical studies have shown that rhythm patterns from music within individual cultures tend to mirror patterns found in that culture's language (Patel & Daniele, 2003), suggesting that part of the way music signals cultural identity may come from adopting linguistic patterns within the culture.

One form of imagined response during music listening is the generation of music-evoked mental imagery (Taruffi & Küssner, 2019; Jakubowski, 2020). This imagery can at times constitute a form of stimulus-independent mind wandering, at other times it is closely shaped by the music (Herff, Cecchetti, Taruffi & Déguernel, 2021; Herff, McConnell, Ji & Prince, 2022). This imagery tends to be reflective of the music's conveyed emotions (Koelsch et al., 2019), and shaped by individual differences (Floridou, Peerdeman & Schaefer, 2022).

Within the general category of music-evoked imagery, one domain in which structural and acoustic features of music from different traditions are known to serve as cues is the generation of fictional imagined stories while listening (Margulis, Williams, et al., 2022). People from multiple cultures have been shown to imagine stories during music listening (Margulis et al., 2019; Margulis, Wong, et al., 2022; Margulis & McAuley, 2022). Culture affects the stories that are imagined during music listening: participants from Arkansas and Michigan in the US, when listening to Chinese and Western classical musical excerpts, imagined broadly similar stories to individual excerpts, but participants from Dimen in China imagined stories that were quite different for those same excerpts (Margulis et al, 2022). However, the music used in these studies was always drawn from a tradition familiar to participants from at least one of the cultures. When one culture is more familiar with the music than the other culture, comparisons between generated narratives across cultures is difficult, as familiarity itself may affect imagined

content (Lancashire, 2017; Margulis 2017). To disentangle familiarity from imagined content, one solution is to present participants with material that is dissimilar to both cultural traditions. One salient structural feature of many musical traditions that can signal cultural identity is the musical scale, a set of notes that are arranged in ascending or descending order of pitch. Most musical scales are organized around the octave, which features a 2:1 ratio in frequency. For example, a musical A that is one octave above the A that is 220 Hz is 2 x 220 Hz = 440 Hz. Within this 2:1 frequency ratio, different cultures use different collections of step sizes. For example, the equal-tempered chromatic scale divides this 2:1 ratio into 12 logarithmically-even steps, and the common-practice Western major diatonic scale has eight unequal-sized steps traversing this 2:1 ratio. Other cultures divide the octave in different ways. For example, the traditional Javanese *slendro* scale has five pitches spaced approximately equally over the octave, whereas the Chinese pentatonic scale uses five uneven step sizes over the octave, resulting in pitches that approximate whole-integer ratios of 1:1, 8:9, 4:5, 2:3, and 3:5 in frequency. These scales can function as signals of cultural identity in music – for example, Chen Gang's Butterfly Lovers Violin Concerto (1959) uses a prominent pentatonic scale structure to signal a Chinese cultural identity within the context of a Western orchestral instrumentation.

A much less common musical scale is the Bohlen-Pierce scale, an artificial musical scale that was discovered when Heinz Bohlen and John Pierce independently questioned the necessity of the octave by exploring the ramifications of a 3:1 instead of 2:1 ratio. The Bohlen-Pierce scale features 13 logarithmically even steps of a 3:1 frequency ratio, and some of its intervals approximate low-integer ratios such as 3:5 and 5:7 (Krumhansl, 1987; Mathews et al., 1988). While the BP scale is mostly unknown across cultures, a small subculture of microtonal composers such as John Chowning, Max Mathews, and Georg Hajdu have adopted the BP scale

and developed notation systems for experimental compositions (Hajdu, 2015). Despite being relatively little known and regarded as non-standard musical practice (Hajdu, 2015), listeners are reliably and rapidly able to learn the structure of the BP scale from a single session of listening (Loui et al., 2010). This learning includes developing sensitivity to its spectral and temporal content (Loui, 2022), and showing brain signatures of learning that bear similar patterns as within-culture music after just 20 minutes of exposure (Loui et al., 2009).

For the present study we asked the following questions: Do individuals from different cultures readily imagine stories for music from an unfamiliar tuning system? If so, how consistent within and between cultures are these imaginings, and what individual and cultural differences predict the consistency or conventionality of these imaginings? While the effects of culture may accumulate over the lifespan, it is also nonlinear with age, as research on immigrant populations suggests that there may be sensitive periods for acculturation over the lifespan (Cheung et al, 2011; Bierwiaczonek et al, 2021). Individual differences in music reward sensitivity and music training are prevalent in many cultures (Mullensiefen et al, 2014; Mas-Herrero et al, 2013; Saliba et al, 2013; Sadakata et al, 2022) and also vary with age (Mullensiefen et al, 2014; Belfi et al, 2020). This is subserved by developmental plasticity in the reward system especially in adolescence and young adulthood (Davidow et al, 2016; Fasano et al, 2022). While the same reward system may be involved in autobiographical memory for music (Barrett & Janata, 2016), and may be linked to the reminiscence bump effect of stronger autobiographical associations for music from adolescence and young adulthood (Belfi & Jakubowski, 2021), the roles of development and individual differences in musical engagement on imagined narratives are relatively unknown. We hypothesize that developmental maturation and individual differences in

musical engagement would affect the vividness of imagined content during music listening, and that would vary between cultures.

Here we report the first study of imagined stories to music written in a tradition that is novel to participants from each culture tested. This allows for a direct comparison of narratives prompted by the same set of excerpts, each of which uses a style that is unfamiliar to both groups. Music composed in the Bohlen-Pierce scale was presented to participants from two geographically defined cultures: Boston, USA and Beijing, China. After hearing each excerpt, participants wrote down the story they imagined, and completed the Narrative Engagement Scale (NES) (Margulis, 2017) in which they rated the ease, clarity, and vividness of the imagined story. To test the hypothesis that age and musical engagement would affect imagined content for each culture, we applied cross-linguistically validated individual difference measures for musical background and training, perceptual skills, emotional engagement, and reward sensitivity in addition to collecting demographic variables on both groups of participants (Mullensiefen et al, 2014; Lin et al, 2021; Wang et al., 2021; Cardona et al, 2022). By combining these individual differences measures with tools from natural language processing to analyze the content of the stories, the present work addresses hypothesized differences in music-evoked imaginings both within and between cultures.

Methods

Participants

Participants were recruited in two cohorts: a US American, specifically Boston-based cohort from Northeastern University (hereafter “the Boston cohort”), and a Chinese, specifically Beijing-based cohort from Beijing Normal University (hereafter “the Beijing cohort”).

Both groups were recruited and tested online. The Boston cohort (n = 241) was recruited and tested in English via Psylink, a website that allows students to participate in online studies in return for course credit. Participants indicated their identified gender, race and ethnicity categories as specified by guidelines published by the US National Institutes of Health (https://www.nih.gov/nih-style-guide). 141/241 participants identified as female, 91 participants as male, 3 as non-binary, and 6 did not provide an answer. 89/241 participants identified as White/Caucasian, 62 as Asian, 35 as Black, 35 as Hispanic or Latino, 2 as American Indian or Alaska Native, 1 as Native Hawaiian or Pacific Islander, 11 as belonging to two or more races. In terms of first language, 179/241 participants identified their first language as English, 14 as Spanish, 11 as Mandarin, 7 as Chinese, 4 as French, 4 as Russian, 4 as Vietnamese, 3 as Haitian Creole, 3 as Hindi, 2 as Portuguese, and 1 as each of Amharic, Cantonese, Harari, Japanese, Kannada, Korean, Polish, and Romanian.

The Beijing cohort (n = 213) was recruited via WeChat, a Chinese instant messaging app, and tested in Chinese in return for payment. A poster containing a QR code was sent in several group messages of Beijing college students, who subsequently shared this code via word of mouth and personal WeChat messages. 137/213 participants identified as female, 68 as male, and 8 did not answer. All participants identified as being from China. 135/213 participants identified their first language as Han Chinese, 71 as Chinese, 3 as Mandarin, and 1 as Szechuanese. Both cohorts were tested online using written instructions on Qualtrics. The Boston cohort was tested in English whereas the Beijing cohort was tested in simplified Chinese (written Hànzi characters) using instructions translated from English by one of the authors (Y.O.) who is a native Mandarin speaker, and checked by all authors.

Stimuli

Musical stimuli for this study consisted of eight recorded excerpts ranging from 20 to 90 seconds long, that were composed in the Bohlen-Pierce scale. These included four solo-clarinet compositions that were composed by our lab for another study (Kathios et al., 2022) and excerpts from four recorded music performances that were composed for a music festival on the Bohlen-Pierce scale in 2010 (Johnson, 2010). The recorded music performances were: *Reminiscences* by Steven Yi, *Bohlen-Pierce Pan Flute World Premiere Folk Tune* by Arturo Raffaele Grolimund, *Hoquetus* by Johannes Kretz, and *Beyond the Horizon* by Georg Hajdu. Table 1 lists the excerpts along with their assigned numerical labels with which we will refer to them in all subsequent analyses. The full recordings are available online: http://bohlen-pierce-conference.org/concert-1 and the selected stimuli for this study, along with analysis code and data, are available on https://osf.io/qjy8h/. Notably, the stimuli were generally unfamiliar, and did not differ in familiarity between US and Chinese groups: In a norming study for another study (Kathios et al, in review), we obtained familiarity ratings from a Chinese sample (n = 156) and a US sample (n = 169), where participants rated familiarity using a Likert-scale ranging from 1 ('not familiar at all') to 6 ('very familiar') for each melody. Both groups rated the melodies as low on familiarity (Chinese sample mean = 2.89, 95% confidence interval = 2.68 – 3.09; US sample mean = 2.86 , 95% confidence interval = 2.64 – 3.09). Familiarity ratings showed no difference between groups, as assessed using a mixed effects model treating participants as a random intercept (ß = 0.02, $t(322) = 0.21$, $p = 0.83$).

Excerpt	Title	Composer	Duration
1	BP_Clarinet_11	Euan Zhang	0:20
2	BP_Clarinet_16	Euan Zhang	0:20
3	BP_Clarinet_18	Euan Zhang	0:20
4	BP_Clarinet_21	Euan Zhang	0:20
5	Bohlen-Pierce Pan Flute World Premiere Folk Tune	Arturo Raffaele Grolimund	1:30
6	Hoquetus	Johannes Kretz	0:45
7	Reminiscences	Steven Yi	1:34
8	Beyond the Horizon	Georg Hajdu	0:25

Table 1. Musical excerpts used in the study.

Procedure

Both cohorts were tested online using a survey on Qualtrics in English for the Boston cohort, and in Chinese for the Beijing cohort. Participants first provided informed consent in accordance with Northeastern University's approved IRB protocol. After informed consent, participants answered basic demographic questions (age, gender identity, race, ethnicity). They then began the listening task. The listening task consisted of eight trials, presented in randomized order without replacement. For each trial, participants were first instructed to listen to a musical excerpt, then they were given the following Story Response Question (SRQ, Margulis et al, 2022):

"Sometimes when people listen to music, they imagine an accompanying story, or elements of a story. Did you imagine a story or elements of a story while listening to this music? Elements of a story include characters, events, and settings." [Yes or no]

[If yes] Please describe the story you imagined in as much detail as possible, but do not spend more than about a minute on the response.

[If no] Why do you think this music didn't trigger an imagined story for you? What were you thinking about/experiencing while listening? Answer these questions in as much detail as possible, but do not spend more than about a minute on the response.

After the Story Response Question, participants completed the Narrative Engagement Scale (Busselle & Bilandzic, 2009), adapted with permission for use with music (Margulis, 2017). The NES includes four statements about each excerpt, which participants rated on a 7-point Likert-like scale ranging from "strongly disagree" to "strongly agree": 1) It was easy to imagine a story while listening to the music. 2) I imagined a vivid story. 3) I imagined a story with a clear setting, characters, and events. 4) I imagined a story while the music was playing, not afterwards.

After the SRQ and the NES, participants completed a battery of surveys to assess individual differences in music reward sensitivity, general physical anhedonia, and musical engagement. The battery consisted of the Extended Barcelona Music Reward Questionnaire (eBMRQ) (Cardona et al., 2022), the Physical Anhedonia Scale (Chapman et al., 1976), and the Goldsmiths Musical Sophistication Index (Gold-MSI) (Müllensiefen et al., 2014).

The eBMRQ grew from the BMRQ (Mas-Herrero et al., 2013), a 20-item scale that consisted of six subscales used to measure individual differences to multiple aspects of reward sensitivity to music: sensorimotor, music seeking, emotion evocation, mood regulation, and social reward. The absorption to music subscale was later added as four additional items to capture the tendency to

experience music-driven states of cognitive immersion (Cardona et al., 2022). The BMRQ has been translated into Chinese (Wang et al., 2021); here we used a new Chinese translation and kept all questions of the BMRQ (instead of removing items that did not fit the factor structure as in Wang et al). We additionally translated the four items of the absorption subscale and added it to form the Chinese eBMRQ. The resulting Chinese version of the eBMRQ is available on https://osf.io/qjy8h/.

The Gold-MSI refers to music-related behaviors, skills, and achievements (Müllensiefen et al., 2014). Its self-report measure is a 38-item scale that captures self-reported musicality along five subscales: musical engagement (9 items), perceptual ability (9 items), musical training (7 items), emotion (6 items), and singing ability (7 items). It has been translated to Chinese (Lin et al., 2021); we used the English version for the Boston cohort and the Chinese version for the Beijing cohort.

Data analysis

Since we were interested in cultural level effects and their relationship to individual differences in narrative engagement with music, we compared the semantic similarity between stories provided by individual participants and the typical story based on all responses to an excerpt by all participants within each location. Stories provided in response to the same music excerpt at the same geographic location were combined into a single nardoc, as in Margulis et al (2022). Nardocs were created for each of the eight excerpts heard at each of the two geographic locations (Boston or Beijing) in order to capture the narrative themes typically experienced when a person from a particular location hears a particular excerpt. Stories from Beijing participants were translated into English by a professional translator blind to the hypotheses of the study.

Preprocessing of the story texts followed the methods used in Margulis et al. (2022) and was done in Python version 3.6.2 (Python Software Foundation, https://www.python. org/). First, we identified and manually corrected misspelled words. Narrative texts were then stripped of all punctuation and capitalization, and English stop-words from Natural Language Toolkit version 3.5 (Bird et al., 2009) were removed. We also removed words that referenced acoustic features of music excerpts (instrument names from a catalog of the world's musical instruments and music traditions, e.g., "Chinese" and "Western") and words that stemmed from task instructions (e.g., "imagine," "music," and "story"); see Table S1 in Supplementary materials for the full list. Finally, we used lemmatize() from the Gensim package version 3.8.3 (R Rehurek & P Sojka, 2010) to convert words to their dictionary forms so that inflected forms of words were treated as the same word. Unlike in Margulis et al. (2022), we included words that were only mentioned a single time across all nardocs because the embedding space approach in the present paper was not limited by term frequency in the same manner as the term frequency inverse document frequency measure used in the previous study.

We measured the semantic relationships between individual stories and nardocs by projecting text into a word embedding space based on a word2vec model pretrained on the Google News dataset (https://code.google.com/archive/p/word2vec/). The model provided a means to obtain feature vectors for words that reflect semantic similarities based on a large corpus of ~3 million words and phrases rather than relying on the smaller collection of stories provided in response to the excerpts (~26,000 words). The closer two words are in embedding space, the greater the semantic similarity. Each word within a story was projected into the 300-dimensional semantic space defined by the pretrained model. Proper nouns referring to specific people and places

missing from the Google News dataset were excluded from further analysis (see Table S2 from Supplementary materials for complete list of excluded nouns). We calculated the average location in semantic space for each story by averaging across embeddings for each word in a story. We calculated the average location in semantic space for each nardoc, representing the average story or the consensus narrative, by averaging across the average embeddings for each participant's story, resulting in each person's story receiving equal influence within a nardoc. Thus, each story (at the level of individual participants) and nardoc (at the level of music excerpts) were transformed into a 300-dimensional feature vector that represented the semantic content of the unique composition of words in every story and nardoc.

Margulis et al. (2022) used the same approach to visualize the semantic relationships between nardocs. In the present paper, we extended this approach by comparing the similarity between embedding space vectors for individual stories and nardocs, rather than using the term frequency inverse document frequency approach used in Margulis et al. (2022) to examine nardoc similarity. Using word embeddings allowed us to avoid the sparse feature vectors that arise from relying on term frequency to characterize individual stories that are much shorter in length than nardocs. Research has shown that taking the average of embedding space vectors across words is an effective way to capture semantic similarities between sentences and paragraphs, though the method ignores the order of words within a document (Arora et al., 2017).

We used cosine similarity to measure the semantic similarity between story and nardoc feature vectors, as in Margulis et al. (2022). Cosine similarity measures the cosine of the angle between two feature vectors, and varies between 0 and 1, where 1 indicates maximum similarity (an angle of 0 degrees between the vectors), and 0 indicates minimum similarity (the vectors are at 90 degrees). We used the cosine similarity measure to identify how semantically similar any

individual's story was to the within-culture nardoc (i.e., the typical story prompted by that excerpt for that cohort) as well as the similarity between nardocs.

Following the visualization method used in Margulis et al. (2022), we used principal components analysis (PCA()) as implemented in the Scikit-learn package version 0.21.3 (Pedregosa et al., 2011) to simultaneously reduce the dimensionality of all story and nardoc feature vectors within-culture to 50 before further reducing to two-dimensional space using the Scikit-learn implementation of t-distributed stochastic neighboring entities (T-SNE()), and, finally, plotting the two-dimensional nardoc vectors for each location. We used the default TSNE() parameters with perplexity set to 30 and random state set to 5. While the relationships between feature vectors in this space are interpretable (e.g., the closer two points are in embedding space, the greater the semantic similarity), the reduced dimensions for the space are not interpretable. While stories from both locations were projected into the same embedding space, each location was visualized separately, thus, the axes and semantic spaces represented cannot be directly compared across figures.

We further measured individual factors contributing to semantic conventionality and narrative engagement using multiple linear regression models. We defined linear regression models to test for individual differences on two outcome measures: narrative engagement as operationalized by average score on the NES, and semantic conventionality as operationalized by cosine similarity from the within-culture consensus narrative, averaged across excerpt trials for each participant. For each outcome measure, our model included the predictor variables of demographics (age, gender), musical skills and expertise (Gold-MSI, 5 subscales: active engagement, perceptual abilities, musical training, singing abilities, emotions), individual differences in reward

sensitivity (extended BMRQ, 6 subscales: absorption to music, music seeking, social reward, sensorimotor, mood regulation, emotion evocation). It was important from a theoretical perspective for narrative engagement and semantic conventionality to each be considered both as an outcome measure and also as a predictor of the other variable. As such, the model for narrative engagement also included semantic conventionality as a predictor, and the model for semantic conventionality also included narrative engagement as a predictor. Data from the Boston and Beijing cohorts were entered into separate models to allow for different patterns of individual differences across cultures. Variance inflation factors were below 4 for all four models (2 cohorts: Boston and Beijing, x 2 outcome variables: semantic conventionality and narrative engagement).

Results

Cohort	**Boston**		**Beijing**	
	Mean	**SD**	**Mean**	**SD**
N	241		213	
age	19.11	1.82	23.19	5.78
eBMRQ Absorption	3.14	0.76	3.70	0.82
eBMRQ_Emotion Evocation	4.01	0.88	4.23	0.60
eBMRQ_Music Seeking	3.85	0.78	3.95	0.76
eBMRQ_Mood Regulation	4.54	0.60	4.45	0.60
eBMRQ_Sensorimotor	4.13	0.83	3.76	0.77
eBMRQ_Social Reward	3.80	0.83	4.01	0.68
eBMRQ_all	3.91	0.57	4.02	0.52

PAS	12.88	6.86	13.03	7.16
Gold-MSI_Active Engagement	4.12	1.09	3.56	1.00
Gold-MSI_Perceptual Ability	5.02	1.03	4.69	1.00
Gold-MSI_Music Training	3.09	1.52	2.75	1.03
Gold-MSI_Emotion	5.43	1.03	5.27	0.98
Gold-MSI_Singing Ability	4.34	1.23	3.71	1.15
Gold-MSI_General Sophistication	3.90	1.09	3.65	1.02
NES avg.	3.80	1.08	3.77	1.06
Avg. # Words /Subject *	52.14	33.12	63.34	39.82

Table 2 sample size and descriptive summary statistics on word count, narrative engagement, and individual difference measures for the two tested geographic locations. Further descriptive summary statistics on nardoc composition and length are given in Table S3 in Supplementary materials. * Note: Excerpts missing a story were removed before averaging # of words per excerpt.

Table 2 shows descriptive summary statistics on all tested variables. It was exceedingly common to imagine stories even for unfamiliar music, as 98.2% of participants (237/239 from Boston and 211/213 from Beijing) reported stories for at least one excerpt. Visualizing these stories as word clouds (Fig. 1) revealed some similarities and some differences in frequently used words across cultures: in response to Excerpt 7 (Reminiscences by Steven Yi), the Boston cohort reported stories involving a mysterious someone moving in space or something in a dark space, such as a room or water or the ocean. The Beijing cohort reported stories involving a person, often a woman, in a dark forest moving slowly at night.

<insert Figure 1 here>

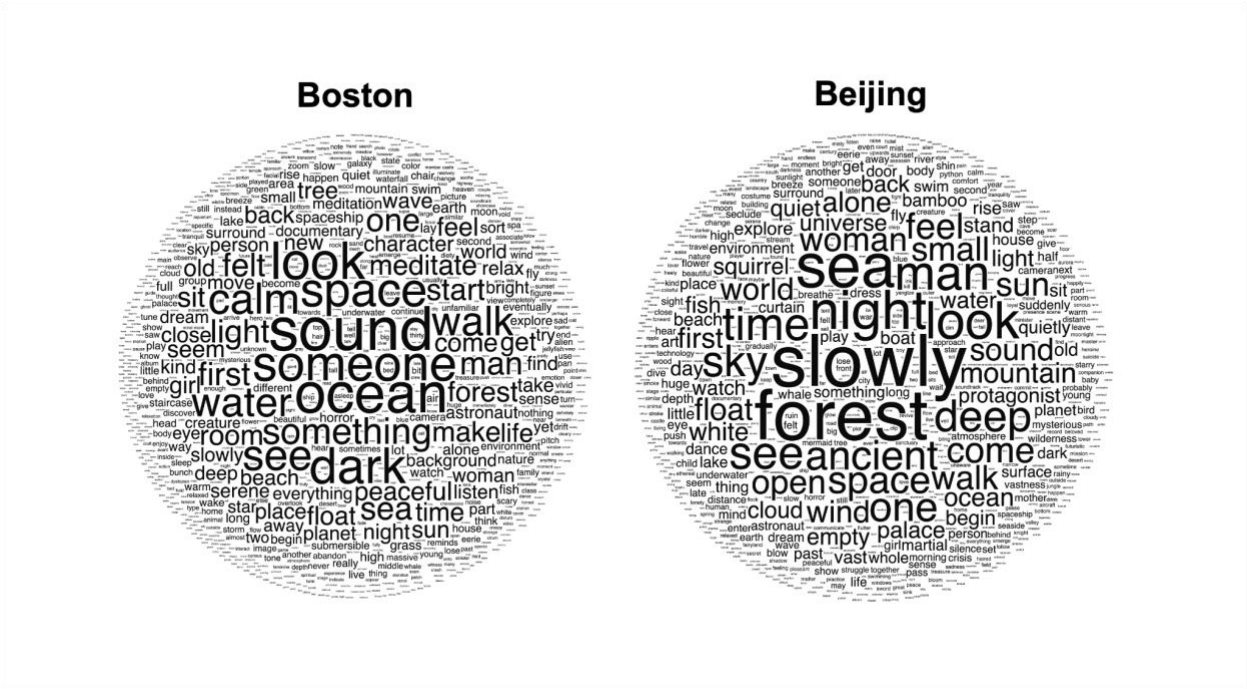

Figure 1. Word clouds in response to Excerpt 7 (Reminiscences by Steven Yi).

A 2D visualization of the semantic space for the Boston cohort's stories showed that the first four excerpts clustered together, and the last four excerpts clustered together (Fig. 2a). This was confirmed by the heat map of semantic similarity as indexed by cosine distances, which showed high semantic similarity across the first four excerpts (lower right quadrant of Fig. 3). Since the first four excerpts were all shorter solo clarinet clips whereas the others were longer and used fuller orchestration, timbre and duration appear to play important roles in how these clips shaped the Boston cohort's imagined narratives.

<insert Figure 2 here>

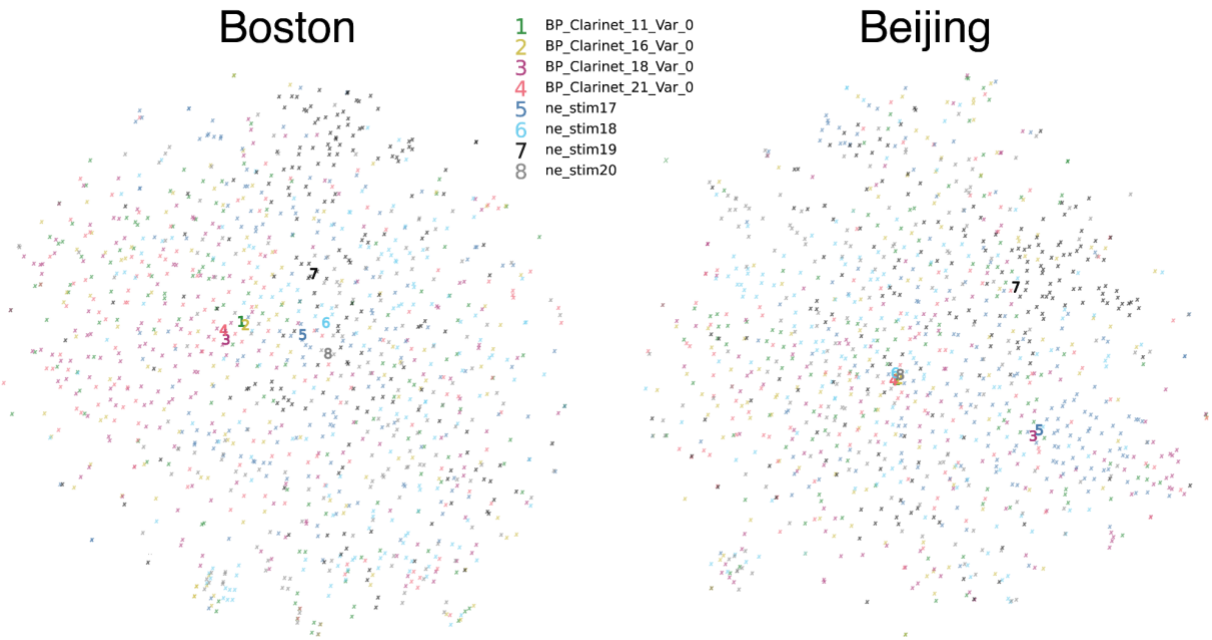

Figure 2. Visualizations of imagined stories in semantic space for the Boston and Beijing cohorts.

The same visualization applied to (translated) stories from the Beijing cohort revealed a different pattern of clustering across excerpts (Fig. 2 right panel). Here, excerpts 3 and 5 clustered together, and excerpt 7 occupied a space that was separate from all the other excerpts, again confirmed by the heatmap showing pairwise semantic similarity across excerpts in Figure 3 (upper left quadrant). Since excerpt 7's primary difference from the other excerpts was the inclusion of slower rhythmic gestures, this clustering implies that rhythmic gestures influence the narratives of the Beijing cohort more than the Boston cohort.

These cross-cultural differences in relative distances between excerpts suggests that the same set of comparably unfamiliar excerpts prompted participants in Beijing and Boston to imagine different patterns of stories. This was confirmed by directly comparing the average cosine similarity for all participants across pairs of excerpts and within and between the cultures, as

shown in Figure 3. This showed on average lower cosine similarity between cultures (Boston and Beijing, the off-diagonal quadrants) than within cultures (Boston-Boston and Beijing-Beijing, the diagonal quadrants). The finding of higher cosine similarity in imagined narratives for within-culture than for between-culture comparisons is statistically significant for all excerpts as confirmed by a paired t-test comparing the distributions of average cosine similarity within- and between-cultures: $t(127) = 16.16$, $p < .0001$. This replicates the key finding in Margulis et al (2022) and extends it to the present unfamiliar stimuli.

<insert Figure 3 here>

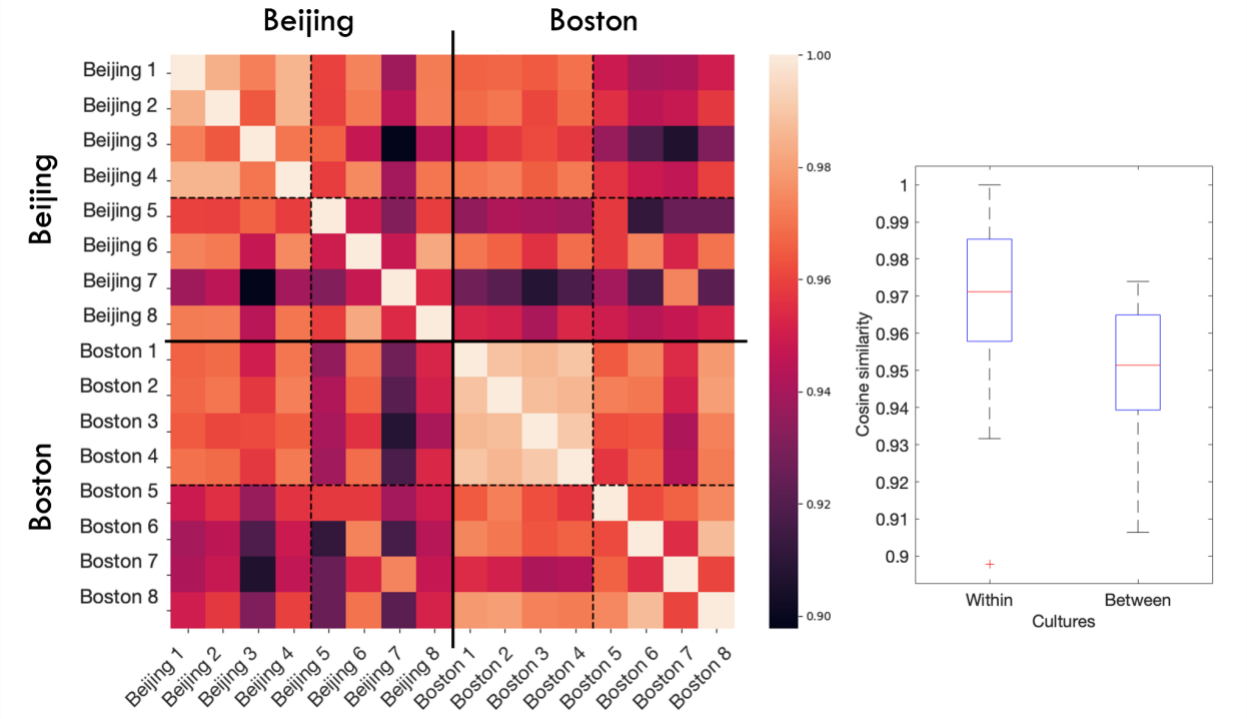

Figure 3. Cosine similarity between pairs of average stories within-culture and between-cultures, showing higher within-culture similarity than between-culture similarity (higher values on the diagonal than the off-diagonal). Averaged within-culture and between-culture similarity are summarized on the right panel.

Computing the cosine similarity (ranging from 0, which is completely dissimilar, to 1, which is exactly the same) between each participant's generated story and the average story allowed us to quantify the individual differences in semantic conventionality for each participant's generated story for each excerpt. Comparing the distribution of cosine similarities across excerpts revealed that for every excerpt, the Beijing cohort showed consistently lower cosine similarities between individual stories and the excerpt's average compared to the Boston cohort ($t(447) = 7.739$, $p < .001$; Fig. 4a-b).

This difference in levels of consensus arose despite similar narrative engagement across cultures: NES scores did not differ between the Boston and Beijing cohorts ($t(433) = 0.32$, $p = .749$; Fig. 4e). Consistent with prior work (Margulis et al., 2022), this reinforces the notion that narrative engagement and narrative consensus can operate independently. Individual excerpts can be highly narratively engaging even when they do not inspire consensus around the details of the imagined story, and vice versa.

Cross-cultural differences in the semantic conventionality, i.e. cosine similarity to the within-culture consensus narrative, were further quantified as a difference in skew and kurtosis in the trial-by-trial data between the cohorts: the Chinese cohort showed less of a rightward skew (less negative skewness) and a flatter distribution (lower kurtosis; more platykurtic) in cosine similarity (Fig. 4c-d). This difference suggests that individual excerpts prompted stories that traversed a broader semantic space across excerpts in the Beijing participants compared to the Boston participants.

<insert Figure 4 here>

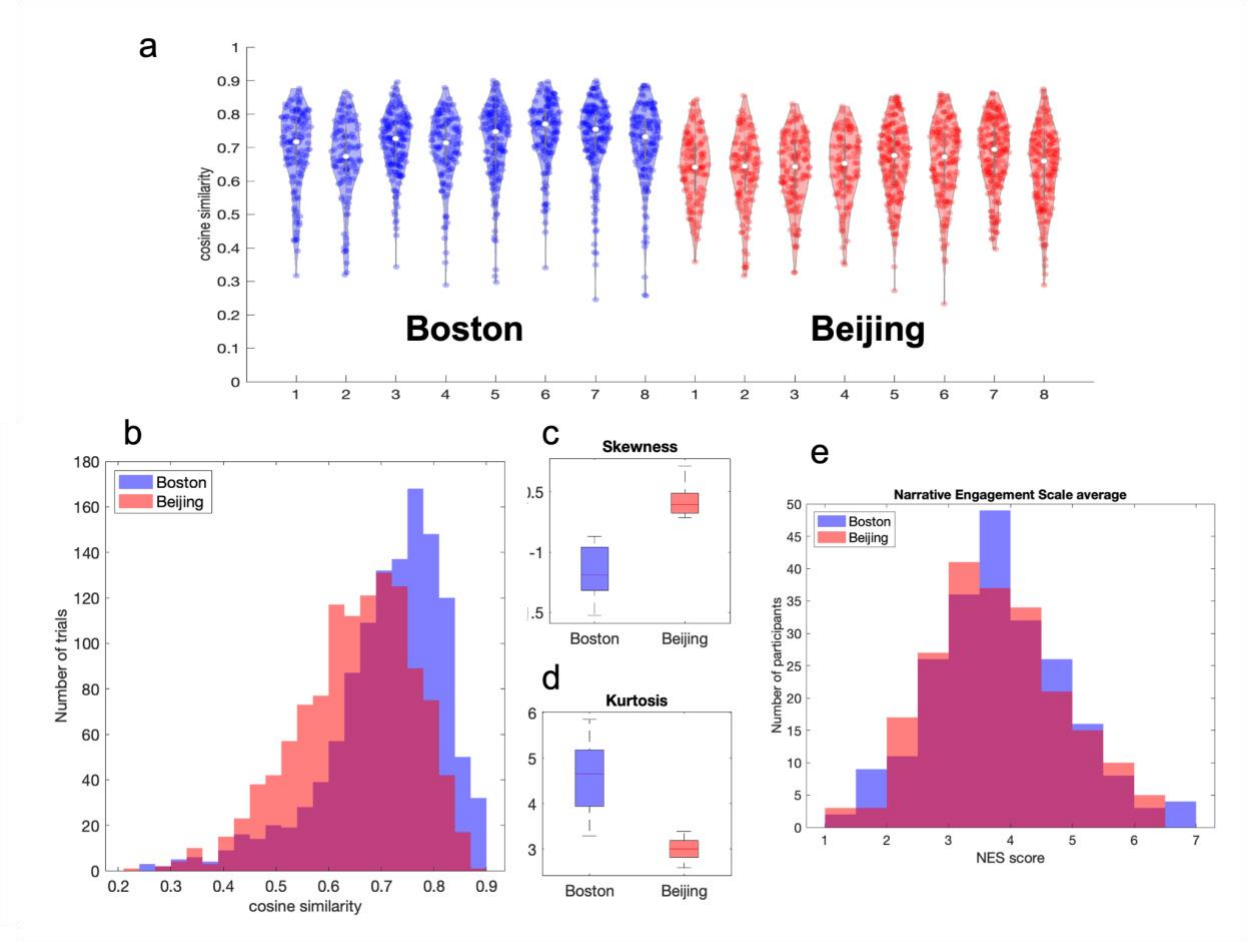

Figure 4. Cross-cultural differences in distribution of semantic conventionality, i.e. cosine similarity to consensus narrative. **4a.** Cosine similarity to within-culture consensus narrative for each excerpt. Boston cohort data are shown in blue; Beijing cohort in red. **4b**. Distribution of cosine similarity for the Boston and Beijing cohorts, with the Beijing cohort showing less skew and flatter distribution as indicated by the skew (**4c**) and kurtosis (**4d**) for the two cohorts. **4e**. Distribution of narrative engagement scale scores do not differ between the cohorts, confirming that the cross-cultural differences in generating consensus are not explained by differences in narrative engagement to the task.

Semantic conventionality, i.e. the average cosine similarity between individual participants' stories and the within-culture consensus story across excerpts, was predicted positively by narrative engagement (NES score) in both cohorts (Boston: standardized beta = .14, t = 2.19, p = .029; Beijing: standardized beta = .26, t = 3.74, p = <.001). In other words, participants of both cohorts who reported imagining more vivid and immediate stories also tended to report stories that were closer to the consensus narrative. In the Beijing cohort, semantic conventionality was also positively predicted by the emotion subscale of the Gold-MSI (beta = 0.25, t = 2.17, p = .031) and negatively predicted by age (beta = -0.19, t = -2.96, p = .004). Thus, among Beijing participants, those who were older wrote less conventional stories, whereas those who reported being more emotionally engaged by music wrote more conventional stories. In the Boston cohort, semantic conventionality was negatively predicted by the absorption subscale of the eBMRQ (beta = -0.174, t = -2.15, p = .033): those who reported being more absorbed by the music wrote less culturally conventional stories.

Considering narrative engagement as the outcome measure, narrative engagement (average NES score) was positively predicted by semantic conventionality in both cohorts (Boston: beta = 0.143, t = 2.19, p = .029; Beijing: beta = 0.25, t = 3.74, p < .001). In the Boston cohort, narrative engagement was also positively predicted by age (beta = 0.14, t = 2.18, p = .03), i.e. Boston participants who self-reported imagining clearer and more vivid stories (high average NES score) also tended to be older, and to write more conventional stories.

In the Beijing cohort, narrative engagement was also negatively predicted by the perceptual ability subscale of the Gold-MSI (beta = -0.25, t = -2.1, p = .037). In other words, Beijing participants who reported being more narratively engaged (higher NES score) tended to write

more conventional stories, but also self-identified as having lower ability to recognize fine-grained perceptual differences in music.

Discussion

We used an unfamiliar musical system, combined with cross-cultural testing and analytical tools from natural language processing, to probe cross-cultural and individual differences in mental imagery during music listening. We observed some cultural differences in the semantic content of imaginings given this unfamiliar music, with imagined stories showing different clusterings in semantic space for the Beijing and Boston cohorts. The same set of excerpts prompted different stories in the Beijing cohort and the Boston cohort, suggesting that the mapping between acoustic features and imagined stories might differ across cultures. Semantic conventionality of the imaginings to the within-culture average story, as determined by cosine distances to the within-culture consensus narrative, were high overall but were lower in the Beijing cohort than in the Boston cohort. Semantic conventionality was predicted by narrative engagement in both cultures, but was additionally inversely predicted by age in the Beijing cohort and by absorption in the Boston cohort. Narrative engagement was positively related to age in the Boston cohort, and negatively related to self-reported perceptual abilities in the Beijing cohort. We interpret each of these findings below.

Cross-cultural differences in consensus and narrative engagement

We observed different levels of semantic conventionality despite similar narrative engagement across cultures: notably a larger, more varied semantic space among Beijing participants, and a narrower, less varied semantic space among Boston participants. Although narrative engagement and semantic conventionality were significantly predictive of each other in both cultures, the Beijing cohort showed lower cosine similarities to the Beijing consensus narratives than the

Boston cohort to the Boston consensus narratives, whereas narrative engagement was similar across the two samples. While narrative engagement reflects self-reported vividness, ease, and clarity of imaginings, it is agnostic of the content of the imagining itself. In contrast, cosine similarity from consensus narrative reflects semantic conventionality, i.e. the semantic distance of each participant's imaginings from their cultural norm as derived from these samples. Thus, individual participants' semantic conventionality may reflect their adherence to cultural norms based on the semantic content of imaginings, which past studies have shown to be rooted in culture in multiple domains including music (Margulis, Wong, et al., 2022; Zittoun & Gillespie, 2015). Future work is needed to determine the extent to which narrative engagement and semantic conventionality tap into overlapping constructs.

Here, we see lower adherence to cultural norms in the semantic content of music-evoked imaginings in the Beijing cohort than in the Boston cohort. This may be driven by heterogeneity across our samples on individual differences measures including age (Table 2). Future studies are needed to recruit more similar samples in age across cultures. If the differences could replicate with more similar age groups across cultures, it could reflect broader cultural differences in music-evoked imaginings. Previous work has noted epistemological differences across Eastern and Western cultures, specifically the idea that Euro-American culture emphasizes analytic modes of thought, whereas Chinese culture emphasizes holistic thought (Nisbett et al., 2001). Holistic and analytic epistemologies may differentially influence the music-driven imagination. While Greco-Roman epistemology emphasizes the use of logic and reasoning as ideal ways to capture the truth, East Asian thought is heavily influenced by a Confucian epistemology, which emphasizes dialectic thinking, and remaining open to multiple interpretations of truths (as iconized for example by the yin-yang as a symbol of harmony). Perhaps because of these cultural

differences, Boston participants may have tended to view the task of imagining stories to music as a problem to be solved, producing a single, more consistent and coherent, "correct" type of story. In contrast, Beijing participants may have regarded imagining stories as more of a creative task, an opportunity to explore possible scenarios, thus resulting in a looser interpretation of the musical surface to derive a wider, more diffusely distributed semantic space. It is possible that the Chinese participants were treating the task as more of an opportunity to expound upon their poetic skills whereas the US participants were treating the task as more of a problem to be solved. Alternatively, these results may reflect other differences between the Boston and Beijing cohorts: although both groups of participants were university students, the Boston cohort participated in return for course credit whereas the Beijing cohort participated for payment (see Participants). While these recruitment strategies were chosen to maximize comparability across the samples given that participation in return for course credit was not an option for the Beijing cohort in this particular study, this difference may have nevertheless affected the motivational structures through which the two cohorts of participants viewed the study. Future studies should aim for recruitment strategies that maximize meaningful and reproducible comparisons across cultures, while continuing to include diverse cross-cultural samples for a more holistic understanding of music and imagination (Savage et al., n.d.).

Narrative engagement and perceptual acuity

Narrative engagement was inversely predicted by self-identified perceptual ability in Chinese participants. The perceptual ability subscale of the Gold-MSI asks for one's ability to judge the intonation and rhythmic accuracy of musical performances (some example items include "I can tell when people sing or play out of tune" or "I can tell when people sing or play out of time with the beat" (Müllensiefen et al., 2014)). The finding that Chinese participants with high perceptual

acuity are less likely to report being narratively engaged provides further support for the idea that there are multiple ways to engage with music listening (Thompson et al., 2022): while participants who report higher perceptual acuity could be listening for more technical aspects of the excerpts, such as tuning, phrasing, or instrumentation, it is the participants with lower perceptual acuity who tend to automatically and simultaneously imagine a vivid story with ease.

Individual differences in conventionality and creativity

Semantic conventionality was positively predicted by narrative engagement in both cohorts. We observed evidence that age carries predictive value for music induced mental imagery; however, the direction of effects differed between the two cohorts. While age predicted narrative engagement, but not conventionality, in the Boston cohort, age predicted conventionality, but not narrative engagement in the Beijing cohort. Furthermore, the effect directionalities were inverted. Specifically, narrative engagement increased with age in the Boston cohort, but semantic conventionality decreased with age in the Beijing cohort: older Boston participants more readily identified themselves as imagining vivid stories, whereas older Beijing participants provided more semantically unconventional stories. This finding could be interpreted by conceptualizing low semantic conventionality as higher originality or even higher creativity. As the present task calls for linguistic knowledge as well as some degree of creativity and imagination, it evokes divergent thinking as well as convergent thinking processes (Zhang et al., 2020). Like other creative tasks, the ability to imagine stories to music leads to some new ideas that can arise as novel associations, as well as other ideas that are generated by exploiting structural features of existing conceptual spaces for each individual within their culture (Boden, 1994). Along similar lines, higher semantic distance between words has been linked to verbal creativity (Olson et al., 2021). The simultaneous importance of creativity and conventionality is

especially apparent in cultural tasks that require social learning (Jagiello et al., 2022), such as music perception and production (Loui & Margulis, 2022; Savage et al., 2021). One speculative interpretation of these results is that within an individualistic culture such as the US, higher absorption to music is related to higher creativity. In contrast, in a collectivistic culture such as in China, individuals who react more emotionally to music may more readily pick up on its structural features to generate a more conventional story, whereas older adults may be less compelled to conform to conventions and more compelled to be creative. Although speculative at this point, if these differences between individualistic and collectivistic cultures hold, one prediction is that they would also extend to other countries and other cultures that vary by collectivism scale (Hofstede, 2010) independent of language use. While the present results are first to observe individual differences to music-evoked imaginings, future work could disentangle the effects of age and emotional sensitivity in other creative tasks, while also following up on the observed cultural differences in other creative tasks.

Limitations

There are methodological challenges in this study, many of which are consistent with psychological research on culture more broadly (Van De Vijver & Leung, 2000). One main limitation is that as a result of the recruitment process, the Boston cohort is necessarily more homogeneous: the Boston participants were all university students from the same university, pursuing an introductory Psychology course likely towards their degree. This can be seen in a smaller spread of age in the Boston cohort, as the SD in the Beijing cohort is larger than that of the Boston cohort. This difference in recruiting confounds the findings; as such, the findings would need further replication to increase our confidence in cross-cultural differences as well as within-group individual differences that differ between cultures. Future studies are needed to

recruit more similar samples in age across cultures. Additionally, the NLP tools relied on in the study are currently limited to the English language; translation from Chinese introduced extra noise into the data. Researchers are currently working to expand linguistic diversity in natural language processing and future studies may be able to improve power by performing analyses directly in the original languages. Despite these limitations, the finding that individuals reported imagined stories in both cultures is robust and replicates previous work (Margulis et al, 2022), and the relationship between narrative engagement and semantic conventionality is relatively robust and replicated in the current samples from both cultures.

Conclusions

Taken together, this paper shows that music composed using the unfamiliar Bohlen-Pierce scale can elicit imagined stories that are rich in narrative content. Participants from different cultures make use of different musical cues to generate imagined stories. While both cultures were similarly narratively engaged by the story response task, the Boston participants showed higher semantic conventionality whereas the Beijing participants showed a wider spread of semantic content in the stories. These results reveal individual as well as cultural differences in modes of narrative engagement and imagination during music listening.

Acknowledgements

Supported by NSF-CAREER 1945436, NSF-BCS 2240330, NIH R01AG078376 and R21AG075232, and startup funds and Covid-19 internal seed funding from Northeastern University. We thank Nick Kathios for the norming study, Jethro Lee for data collection and Lei Chen for help with translations and Steffen A. Herff for helpful comments on a prior version of this manuscript.

References

Arora, S., Liang, Y., & Ma, T. (2017). A simple but tough-to-beat baseline for sentence embeddings. *International Conference on Learning Representations*.

Barrett, F. S., & Janata, P. (2016). Neural responses to nostalgia-evoking music modeled by elements of dynamic musical structure and individual differences in affective traits. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2016.08.012

Belfi, A. M., & Jakubowski, K. (2021). Music and Autobiographical Memory. *Music & Science, 4*, 20592043211047124. https://doi.org/10.1177/20592043211047123

Bennett, A. (2017). *Music, space and place: Popular music and cultural identity*. Routledge.

Bierwiaczonek, K., & Kunst, J. R. (2021). Revisiting the Integration Hypothesis: Correlational and Longitudinal Meta-Analyses Demonstrate the Limited Role of Acculturation for Cross-Cultural Adaptation. *Psychological Science, 32*(9), 1476–1493. https://doi.org/10.1177/09567976211006432

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

Boden, M. A. (1994). *Dimensions of creativity*. The MIT Press.

Busselle, R., & Bilandzic, H. (2009). Measuring Narrative Engagement. *Media Psychology*, *12*(4), 321–347. https://doi.org/10.1080/15213260903287259

Cardona, G., Ferreri, L., Lorenzo-Seva, U., Russo, F. A., & Rodriguez-Fornells, A. (2022). The forgotten role of absorption in music reward. *Ann N Y Acad Sci*, *1514*(1), 142–154. https://doi.org/10.1111/nyas.14790

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology*, *85*(4), 374.

Cheung, B. Y., Chudek, M., & Heine, S. J. (2011). Evidence for a Sensitive Period for Acculturation: Younger Immigrants Report Acculturating at a Faster Rate. *Psychological Science, 22*(2), 147–152. https://doi.org/10.1177/0956797610394661

Davidow, J. Y., Foerde, K., Galván, A., & Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron, 92(1), 93–99. https://doi.org/10.1016/j.neuron.2016.08.031

Finn, E., & Wylie, R. (2021). Collaborative imagination: A methodological approach. *Futures*, *132*, 102788. https://doi.org/10.1016/j.futures.2021.102788

Floridou, G. A., Peerdeman, K. J., & Schaefer, R. S. (2022). Individual differences in mental imagery in different modalities and levels of intentionality. *Memory & Cognition, 50*(1), 29-44. https://doi.org/10.3758/s13421-021-01209-7

Hajdu, G. (2015). *Dynamic Notation–A Solution to the Conundrum of Non-Standard Music Practice*. 241–248.

Henrich, J., Blasi, D. E., Curtin, C. M., Davis, H. E., Hong, Z., Kelly, D., & Kroupin, I. (2022). A cultural species and its Ccognitive phenotypes: Implications for philosophy. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021-00612-y

Herff, S. A., Cecchetti, G., Taruffi, L., & Déguernel, K. (2021). Music influences vividness and content of imagined journeys in a directed visual imagery task. *Scientific Reports, 11*(1), 1-12.

Herff, S. A., McConnell, S., Ji, J. L., & Prince, J. B. (2022). Eye Closure Interacts with Music to Influence Vividness and Content of Directed Imagery. *Music & Science, 5*, 20592043221142711.

Hofstede, G. (2001). Culture's recent consequences: Using dimension scores in theory and research. *International Journal of Cross Cultural Management*, *1*(1), 11–17.

Jagiello, R., Heyes, C., & Whitehouse, H. (2022). Tradition and invention: The bifocal stance theory of cultural evolution. *Behavioral and Brain Sciences*, *45*, e249. Cambridge Core. https://doi.org/10.1017/S0140525X22000383

Jakubowski, K. (2020). Musical Imagery. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (pp. 187–206). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781108580298.013

Johnson, C. (2010, March 7). Symphony in J flat: The curious quest to re-invent music. *Boston Globe*.

Kathios, N., Sachs, M. E., Zhang, E., Ou, Y., & Loui, P. (2022). *Generating New Musical Preferences from Hierarchical Mapping of Predictions to Reward*. Submitted.

Koelsch, S., Bashevkin, T., Kristensen, J., Tvedt, J., & Jentschke, S. (2019). Heroic music stimulates empowering thoughts during mind-wandering. *Scientific Reports,* 9(1), 1-10.

Krumhansl, C. L. (1987). General properties of musical pitch systems: Some psychological considerations. In J. Sundberg (Ed.), *Harmony and Tonality* (Vol. 54, pp. 33–52). Royal Swedish Academy of Music.

Lancashire, R. (2017). *An experience-sampling study to investigate the role of familiarity in involuntary musical imagery induction*. 10th International Conference of Students of Systematic Musicology (SysMus17) London, England.

Liao, S.-Y. & Gendler, T. (2019). Imagination. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/imagination/

Lie, J. (2012). *What is the K in K-pop? South Korean popular music, the culture industry, and national identity*.

Lin, H.-R., Kopiez, R., Müllensiefen, D., & Wolf, A. (2021). The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication in a Taiwanese sample. *Musicae Scientiae*, *25*(2), 226–251.

Loui, P., Ou, Y., Margulis, E. H., & Kubit, B. M. (2023, May 30). BP imagination x culture. Retrieved from osf.io/qjy8h

Loui, P. (2022). New music system reveals spectral contribution to statistical learning. *Cognition*, *224*, 105071. https://doi.org/10.1016/j.cognition.2022.105071

Loui, P., & Margulis, E. H. (2022). Creativity and tradition: Music and bifocal stance theory. *The Behavioral and Brain Sciences*, *45*, e262. https://doi.org/10.1017/S0140525X22001212

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans Rapidly Learn Grammatical Structure in a New Musical Scale. *Music Perception*, *27*(5), 377–388. https://doi.org/10.1525/mp.2010.27.5.377

Loui, P., Wu, E. H., Wessel, D. L., & Knight, R. T. (2009). A Generalized Mechanism for Perception of Pitch Patterns. *Journal of Neuroscience*, *29*(2), 454–459. https://doi.org/10.1523/jneurosci.4503-08.2009

Margulis, E. H. (2017). An Exploratory Study of Narrative Experiences of Music. *Music Perception*, *35*(2), 235–248. https://doi.org/10.1525/mp.2017.35.2.235

Margulis, E. H., & McAuley, J. D. (2022). Using music to probe how perception shapes imagination. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.07.011

Margulis, E. H., Williams, J., Simchy-Gross, R., & McAuley, J. D. (2022). When did that happen? The dynamic unfolding of perceived musical narrative. *Cognition*, *226*, 105180. https://doi.org/10.1016/j.cognition.2022.105180

Margulis, E. H., Wong, P. C. M., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications*, *5*(1), 146. https://doi.org/10.1057/s41599-019-0363-1

Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences*, *119*(4), e2110406119. https://doi.org/10.1073/pnas.2110406119

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual differences in Music Reward experiences. *Music Perception: An Interdisciplinary Journal*, *31*(2), 118–138.

Mathews, M. V., Pierce, J. R., Reeves, A., & Roberts, L. A. (1988). Theoretical and experimental explorations of the Bohlen-Pierce scale. *J Acoustical Soc Am*, *84*, 1214–1222.

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. https://doi.org/10.1371/journal.pone.0089642

Nanay, B. (2023). *Mental Imagery: Philosophy, Psychology, Neuroscience*. Oxford University Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. APA PsycArticles®. https://doi.org/10.1037/0033-295X.108.2.291

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

Parker, S. (n.d.). *Claude Debussy's Gamelan—College Music Symposium*. Retrieved November 16, 2022, from https://symposium.music.org/index.php?option=com_k2&view=item&id=22:claude-debussys-gamelan

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-b45.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Pelaprat, E., & Cole, M. (2011). "Minding the Gap": Imagination, Creativity and Human Cognition. *Integrative Psychological and Behavioral Science*, *45*(4), 397–418. https://doi.org/10.1007/s12124-011-9176-5

R Rehurek & P Sojka. (2010). *Software Framework for Topic Modelling with Large Corpora*. https://doi.org/10.13140/2.1.2393.1847

Sadakata, M., Yamaguchi, Y., Ohsawa, C., Matsubara, M., Terasawa, H., von Schnehen, A., Müllensiefen, D., & Sekiyama, K. (2022). The Japanese translation of the Gold-MSI:

Adaptation and validation of the self-report questionnaire of musical sophistication. *Musicae Scientiae*, 102986492211100. https://doi.org/10.1177/10298649221110089

Saliba, J., Lorenzo-Seva, U., Marco-Pallares, J., Tillmann, B., Zeitouni, A., & Lehmann, A. (2016). French validation of the Barcelona music reward questionnaire. *PeerJ*, *4*, e1760.

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. Cambridge Core. https://doi.org/10.1017/S0140525X20000333

Savage, P. E., *Nori Jacoby (Max Planck Institute for Empirical Aesthetics, Germany), *Elizabeth H. Margulis (Princeton University, USA), Hideo Daikoku (Keio University, Japan), Manuel Anglada-Tort (Max Planck Institute for Empirical Aesthetics, Germany), Salwa El-Sawan Castelo-Branco (Ethnomusicology Institute - Center for Studies in Music and Dance, NOVA University of Lisbon, Portugal), Florence Ewomazino Nweke (University of Lagos, Nigeria), Shinya Fujii (Keio University, Japan), Shantala Hegde (National Institute of Mental Health and Neurosciences, India), Hu Chuan-Peng (Nanjing Normal University, China), Jason Jabbour (UN Environment Program, USA), Casey Lew-Williams (Princeton University, USA), Diana Mangalagiu (University of Oxford, UK and Neoma BS, France), Rita McNamara (Victoria University of Wellington, New Zealand), Daniel Müllensiefen (Goldsmiths, University of London, UK), Patricia Opondo (University of KwaZulu-Natal, South Africa), Aniruddh D. Patel (Tufts University, USA), & Huib Schippers. (n.d.). Building sustainable global collaborative networks: Recommendations from music studies and the social sciences. In *The Science-Music Borderlands: Reckoning with the past, imagining the future.* MIT Press.

Taruffi, L., & Küssner, M. B. (2019). A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 62.

Thompson, W. F., Bullot, N. J., & Margulis, E. H. (2022). The psychological basis of music appreciation: Structure, self, source. *Psychological Review*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/rev0000364

Van De Vijver, F. J. R., & Leung, K. (2000). Methodological Issues in Psychological Research on Culture. *Journal of Cross-Cultural Psychology, 31*(1), 33–51. https://doi.org/10.1177/0022022100031001004

Vygotsky, L. S. (2004). Imagination and Creativity in Childhood. *Journal of Russian & East European Psychology*, *42*(1), 7–97. https://doi.org/10.1080/10610405.2004.11059210

Wang, J., Xu, M., Jin, Z., Xia, L., Lian, Q., Huyang, S., & Wu, D. (2021). The Chinese version of the Barcelona Music Reward Questionnaire (BMRQ): Associations with personality traits and gender. *Musicae Scientiae*, 10298649211034548. https://doi.org/10.1177/10298649211034547

Zhang, W., Sjoerds, Z., & Hommel, B. (2020). Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage*, *210*, 116572. https://doi.org/10.1016/j.neuroimage.2020.116572

Zittoun, T., & Cerchia, F. (2013). Imagination as Expansion of Experience. *Integrative Psychological and Behavioral Science*, *47*(3), 305–324. https://doi.org/10.1007/s12124-013-9234-2

Zittoun, T., & Gillespie, A. (2015). Internalization: How culture becomes mind. *Culture & Psychology*, *21*(4), 477–491. https://doi.org/10.1177/1354067X15615809