

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Cultural Invariance in Musical Communication

### **Permalink**

<https://escholarship.org/uc/item/7hc3762n>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

### **Authors**

Yurdum, Lidya  
Singh, Manvir  
Glowacki, Luke  
et al.

### **Publication Date**

2022

Peer reviewed

# Cultural Invariance in Musical Communication

**Lidya Yurдум**

Department of Psychology, University of Amsterdam,  
Amsterdam 1018WT, Netherlands

**Manvir Singh**

Department of Human Evolutionary Biology, Harvard  
University, Cambridge, MA 02138, USA  
Institute for Advanced Study in Toulouse, 31080,  
Toulouse Cedex 6, France

**Luke Glowacki**

Department of Anthropology, Boston University, Boston,  
MA 02215, USA

**Tom Vardy**

School of Psychology, University of Auckland, Auckland  
0632, New Zealand

**Quentin Atkinson**

School of Psychology, University of Auckland, Auckland  
0632, New Zealand

**Courtney B. Hilton**

Haskins Laboratories, New Haven CT 06511 USA

**Disa Sauter**

Department of Psychology, University of Amsterdam,  
Amsterdam 1018WT, Netherlands.

**Max Krasnow**

Department of Psychology, Harvard University,  
Cambridge, MA 02138, USA  
Division of Continuing Education, Harvard University,  
Cambridge, MA 02138, USA

**Samuel Mehr**

Haskins Laboratories, New Haven CT 06511 USA

## Abstract

Despite the variability of music worldwide, some types of human songs share basic acoustic characteristics. For example, dance songs tend to be loud and rhythmic, whereas lullabies tend to be quiet and melodious. Prior studies with western English-speaking participants have shown that this enables listeners to infer aspects of a singer's behavior, despite being unfamiliar with the singer's culture and language. Here, we test whether these intuitions are shared across a diversity of languages and human societies, with 5524 people from 49 industrialised countries comprising 28 languages, and 116 people in 3 small-scale societies with limited access to global media. Each made inferences about the behavioral contexts of 118 songs from 86 societies. Both groups reliably identified the behavioral functions of dance songs, lullabies, and healing songs. Linguistic and geographical proximity between listeners and singers was minimally predictive of accuracy, demonstrating a degree of cultural invariance in music perception.

**Keywords:** music, cross-cultural, universality, form, function, in-group advantage

Music is a human universal endowed with rich variation within and across cultures (Lomax, 1968; Mehr et al., 2019; Nettl, 1964). Some of the ways that music is used, however, are conspicuously similar around the globe, such as singing to soothe fussy infants, or singing to excite people to coordinate their bodies in the context of dance (Hilton, Crowley, et al., 2021; Mehr et al., 2018, 2019). Music used for specific functions tends to display stereotyped acoustic features: for example, dance songs all around the world tend to share clearly accented and predictable beat structures. Such regularities allow people to reliably infer the behavioral

functions conveyed by unfamiliar foreign music (Mehr et al., 2018, 2019), even in children as young as three years of age (Hilton, Crowley, et al., 2021).

Why should the relationship between the behavioral function of a song and its acoustic features be stable across cultures? Stable patterns in human song may originate from our evolved psychology and unique social environment (Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al., 2021; Savage et al., 2021). Specifically, innate perceptual biases may constrain cultural evolutionary processes such that behaviours ultimately become both culturally specific while grounded in general biological tendencies (Richerson & Boyd, 2008; Sperber & Hirschfeld, 2004). This combination of universals and variation is what is observed in the association between behavioral contexts and acoustic features in music around the world (Mehr et al., 2019).

However, a notable limitation of the prior experiments studying people's perceptions of the relationships between songs' behavioral functions and acoustic features (Hilton, Crowley, et al., 2021; Hilton, Moser, et al., 2021; Mehr et al., 2018, 2019; Trehub et al., 1993) is that the listeners in these studies were from primarily Western, Educated, Industrialized, Rich, and Democratic countries (Henrich et al., 2010). Thus, although the stimuli they listened to were cross-culturally representative, it is unclear how much of the accuracy of their perceptions is accounted for by universal musical behaviour, and how much of it is a product of (Western) enculturation, education, and exposure to world music through globalized media.

The present study provides a strong test of the universality and generalizability of people's inferences of musical functions in a systematic cross-culturally diverse sample of listeners from around the world. This includes both a diverse sample of listeners from industrialized societies, and from three small-scale societies. "Small-scale societies" comprise 1) societies in which individuals interact in a "small" world (where individuals interact with 10s-100s of other individuals, most interactions are face-to-face, and there is a high degree of interdependence), and 2) societies that are less affected by the state, the market, globalisation and world religions. We preregistered this study at <https://osf.io/msvwz> with the following confirmatory hypotheses: Replicating prior work with Western English-speaking participants (Mehr et al., 2018), we predicted that this more diverse sample of listeners would rate lullabies as more likely to be used to "soothe a baby", dance songs as more likely to be used "for dancing" and healing songs as more likely to be used "to heal the sick", relative to the other song types. We did not expect this same pattern for love songs, as no such pattern was found in Mehr et al. (2018). We expected response patterns to be similar across listeners in industrialised countries and for these to hold amongst participants from culturally isolated, small-scale societies, and for listeners in both industrialised and small-scale societies to demonstrate similar response patterns.

To further tease apart universal and culture-specific explanations of these effects, we also conducted exploratory analyses assessing how cultural proximity affects people's inferences. For example, when listening to a song from a familiar culture, is one better at identifying its true behavioral context, relative to when the song is from an unfamiliar culture? We expected listeners' inferences to reflect both a sensitivity to culturally learned cues as well as sensitivity to cues whose origins are general to all human cultures, in line with other similar domains such as people's perceptions of emotion in vocalizations (Elfenbein & Ambady, 2002; Laukka & Elfenbein, 2021).

## Methods

### Procedure

To test if listeners can infer the functions of unfamiliar music, we asked participants to listen to songs recorded in small-scale societies and judge their behavioral functions. We recruited participants from 49 industrialised societies and three small-scale societies.

The industrialised society cohort completed an online Qualtrics survey in their native language wherein each participant listened to 24 songs. The surveys were professionally translated by Qualtrics with a back-translation procedure. Each song originally performed one of four functions in the culture it was recorded in: expressing love for another person, healing the sick, soothing a baby, or accompanying dance. Participants were unaware of the correct function. After each song, they were asked to think about the people making the music and rate how likely it was

that the song was used for six different functions: the four possible functions described above and two additional functions that were not represented in the corpus ("to greet visitors" and "to praise a person's achievements"). The response scale ranged from 1 (Definitely not used for [function]) to 4 (Definitely used for [function]). Participants reported demographics, and musical and online experience.

A similar procedure was carried out in the three small-scale societies in the local languages. The procedures were translated by native speakers and the on-site authors in a back-translation process. On the basis of piloting completed by two of the authors, two adjustments were made to make the task more manageable and comprehensible: we reduced the response scale by one point to include only 3 response options, and the number of songs each participant heard was reduced to 18. The task was carried out on-site: participants were first familiarised with a keypad with three buttons corresponding to the response scale ("yes", "a little" and "no"). Participants listened to 18 songs and used the keypad to answer the verbal prompt "Do you think the people making this music use it for [function]?" Because the participants wore headphones, the experimenters could not hear what song was played on a given trial, and so they were always unaware of the correct song type. At the end of the experiment, participants were asked to re-identify each button to confirm that they remembered the response labels.

### Participants

**Industrialised Societies.** 5,524 listeners from 49 countries (not including the United States) completed a Qualtrics survey, which was translated into 28 different languages to facilitate sampling of participants from these countries without the requirement that they speak English. This provides a strong test of the generalizability of the previous finding (Mehr et al., 2018) that listeners, primarily from the United States and India, reliably discriminated form-function relationships in music from around the world. Participants were paid for their participation. We excluded participants who scored poorly on a headphone check, reported significant difficulties hearing the music, failed an attention check, or completed the study in under a minute, and those whose IP address did not match the country they reported being in.

**Small-Scale Societies.** 116 adult participants were recruited from the Nyangatom in Ethiopia, the Mentawai in Indonesia and Ni-Vanuatu in Vanuatu through word-of-mouth. These three societies were chosen for their limited exposure to music from other cultural traditions and because the authors had direct access to them. All three societies have limited access to TV, radio and the internet and can be assumed to not have significant exposure to these communication channels; the Nyangatom in particular have no access to these technologies. 15 participants were excluded because the experimenter expressed apprehension as to whether they

understood the task; these exclusions were made without the experimenter being aware of the songs the participant heard.

### Materials

Stimuli comprised 118 vocal music excerpts from the Natural History of Song Discography (Mehr et al., 2019) recorded in 86 mostly small-scale societies spanning all 30 world regions (Murdock et al., 2008; Naroll, 1967), 86 languages and a range of subsistence methods. These stimuli consisted of 30 lullabies, 30 dance songs, 30 love songs, and 28 healing songs, and hence, four functions: soothing a baby, dancing, expressing love and healing the sick. The subset of songs each participant heard were drawn from the corpus randomly and without replacement.

### Results

For both the industrialised and small-scale society cohorts, we calculated mean scores for each song on each function scale. These scores were then z-scored to facilitate comparison across song types and across cohorts. Amongst the industrialised cohort listeners, each song was rated between 919 and 1185 times. The number of times each individual song was rated was by necessity much smaller for the small-scale society cohort (min: 8, max: 28); however, this sample still contained between 486 and 541 ratings per song type.

### Accuracy of Listener Inferences

We first tested whether each target song type (e.g., “dance songs”) was rated higher than the average rating across all four song types on its respective scale (e.g., “used for dancing”). Conceptually, this is equivalent to asking whether a target song type is perceived to be more appropriate for a given function than the average song in our study. We ran four separate no-constant multiple regressions, where the z-transformed mean ratings for each song on each scale were regressed onto binary variables denoting the actual song types. As in Mehr et al. (2018), listeners discriminated three of the four song types reliably above chance, confirming the preregistered predictions.

Importantly, this held for listeners from both industrialised and small-scale societies (Figure 1). The industrialised cohort rated dance songs significantly above the base rate of “used for dancing”: Ratings for actual dance songs on this scale were 0.90 standard-deviations above the average rating across all songs (regardless of original function). In contrast, lullabies were rated 0.83 standard-deviations below the base rate ( $ps < 0.0001$ ), indicating that the listeners from industrialised countries could infer that completely unfamiliar dance songs were suitable for dancing, whereas lullabies were not. Indeed, dance songs were higher on average on the “used for dancing” scale compared to each of the other three song types individually ( $ps < 0.001$ ) and compared to the three combined ( $p < 0.0001$ ). Listeners from the three small-scale, culturally isolated societies displayed

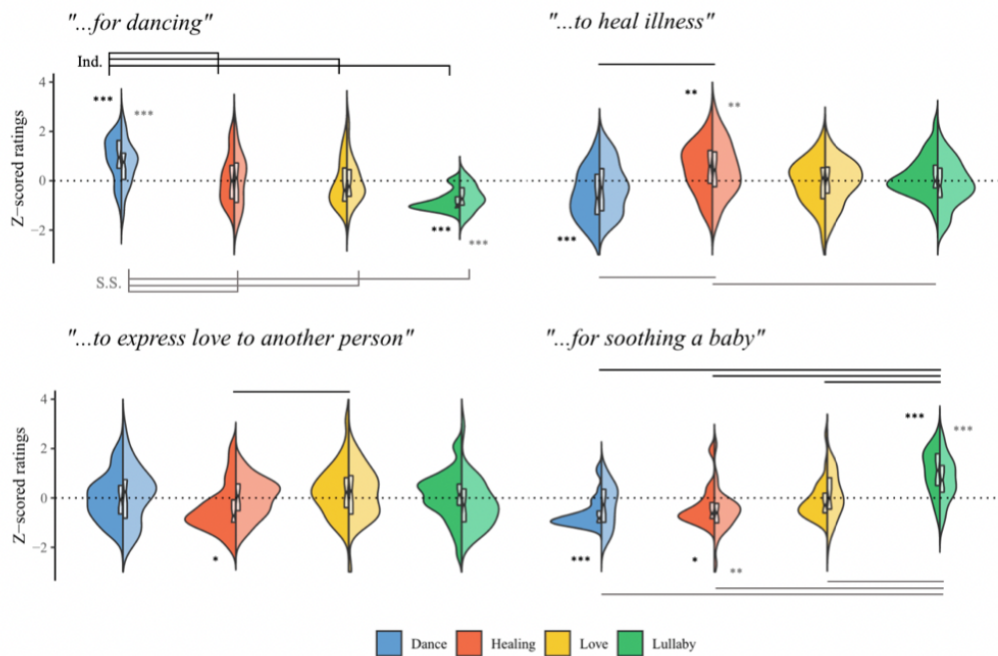


Figure 1. Accuracy of listener inferences in industrialised (left) and small-scale (right) societies. The dotted line represents the baseline (average) rating all four song types on a particular scale. Listener ratings are z-scored to allow comparison across cohorts. For all song types except love songs, listeners were able to infer the primary function of the song.

Significance: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05

similar response patterns, despite the smaller sample size and the differences in the experimental measurements (ratings were made on a simpler 3-point scale rather than a 4-point scale). As with the industrialized sample, dance songs were once again reliably rated above the average for “used for dancing” ( $\beta_{\text{dance}} = 0.67$ ,  $SE = 0.162$ ,  $p < .0001$ ), and lullabies rated below average ( $\beta_{\text{baby}} = -0.68$ ,  $SE = 0.162$ ,  $p < .0001$ ).

Ratings on the “used to soothe a baby” scale were similarly clear for both cohorts of listeners. Listeners from industrialised societies rated lullabies as 1.09 standard deviations above the average rate across all songs, clearly separated from the other three song types both individually and when averaged together ( $ps < 0.0001$ ). Unsurprisingly, dance songs were rated as least appropriate for soothing a baby, with the average rating being 1.71 standard deviations below the average for true lullabies ( $p < .0001$ ). Once again, we observed this same pattern of results amongst listeners from small-scale societies. For the “used for soothing a baby” scale, listeners rated lullabies on average 0.75 standard-deviations above the base rate ( $p < .0001$ ). While both dance songs and healing songs were rated below average on this scale, only the ratings for healing songs were statistically different from the average ( $\beta_{\text{heal}} = -0.55$ ,  $SE = 0.166$ ,  $p = 0.001$ ;  $\beta_{\text{dance}} = -0.32$ ,  $SE = 0.161$ ,  $p = 0.05$ ). The ratings for dance and healing songs were both reliably lower in comparison to the ratings of lullabies ( $ps < 0.05$ ).

The effects for healing songs were more subtle, but still significant in both cohorts. Listeners from industrialised societies rated true healing songs 0.49 standard-deviations above the average on the “used to heal illness” scale ( $p = 0.006$ ). Healing songs were rated higher on this scale compared to dance songs ( $p < .001$ ) but did not reliably differ from either lullabies or love songs ( $p > 0.05$ ). Listeners from small-scale societies similarly identified healing songs as appropriate for healing illness: only the ratings for healing songs differed from the average rate on this scale ( $\beta_{\text{heal}} = 0.48$ ;  $p = 0.01$ ). Healing songs were also rated higher compared to both dance songs and lullabies ( $ps < 0.05$ ), but not love songs ( $p = 0.11$ ).

Finally, as expected, neither of the cohorts differentiated the function of love songs. Only the industrialised society cohort identified healing songs as being inappropriate for expressing love to another person. None of the other song types were rated significantly above (or below) the average for the “used to express love” scale ( $ps > 0.05$ ) in either cohort.

In exploratory analyses, we replicated the findings from the industrialized societies using mixed-effects models to improve their generalizability; all findings repeated with slightly attenuated effect sizes (see Supplementary Material: <https://osf.io/nxwv9/>).

### Within-Cohort Agreement

To test the consistency of the observed effects, we tested how similar the responses of participants within the industrialised cohort were. To do this, we first split the industrialised society sample into 28 “sub-cohorts”, defined by the 28

different languages that the study was translated into. We then bootstrapped randomly selected pairs of these sub-cohorts and tested the distribution of correlations against a null hypothesis of  $r = 0$ . Agreement was high on all four function scales. Listeners from industrialised societies mostly agreed on which songs were (and were not) appropriate for dancing (mean  $r = 0.88$ ), soothing a baby (mean  $r = 0.84$ ), healing the sick (mean  $r = 0.61$ ), and expressing love for another person (mean  $r = 0.59$ ; all  $ps < 0.0001$ ). The high degree of internal consistency of these inferences across industrialised societies is also shown in Figure 2b.

### Comparing Listeners from Industrialised and Small-Scale Societies

So far, the listeners from small-scale societies showed qualitatively similar patterns of inferences to those from industrialised societies. To test their similarity more precisely, we ran a Pearson’s correlation comparing the two cohorts’ average listener ratings for each song on each dimension. The industrialised and small-scale society cohorts’ responses were positively correlated across the 118 songs for all four rating scales (Figure 2a), indicating that participants in both cohorts used the rating scales similarly. However, the magnitude of the correlations differed depending on the scale: correlation was moderate to high on the “used for dancing” ( $r(116) = 0.84$ ) and “used to soothe a baby” ( $r(116) = 0.59$ ) scales ( $ps < 0.0001$ ), and weaker for the “used to heal illness” ( $r(116) = 0.33$ ,  $p < .001$ ) and “used to express love for another person” scales ( $r(116) = 0.25$ ,  $p = 0.007$ ).

To better account for differences in sample size between the cohorts, we re-ran this analysis using a stratified bootstrapping technique: Within each cohort, we sampled 30 observations with replacement for each song and generated a new dataset of song-wise averages for each rating scale. We then correlated the new datasets from the industrialised and small-scale cohorts and stored the resulting correlation coefficient. After 10,000 iterations, we compared the distribution of correlation estimates to a null hypothesis of  $r = 0$ . These correlations were significant although slightly attenuated (dance: mean  $r = 0.72$ ; lullaby: mean  $r = 0.50$ ; healing: mean  $r = 0.22$ ; love: mean  $r = 0.15$ ; all  $ps < 0.0001$ ; Supplementary Figure 1).

### Role of Cultural Proximity

The analyses presented thus far show robust commonalities across cultures and provide strong support for the idea that innate perceptual biases constrain specific types of human song. To further tease apart the extent to which universal and culture-specific tendencies shape the relationship between the behavioral function and acoustic features of a song, we conducted a number of post-hoc exploratory analyses.

First, we investigated the effect of familiarity a particular listener has with the culture of the song they are listening to. If culture-specific learned musical “rules” explain a large portion of the effects described above, then we would expect to see substantial differences between the performance of a

listener on trials where they listen to culturally familiar songs, and those where they listen to culturally unfamiliar songs. To operationalize this hypothesis, we used two convergent measures of cultural proximity between listener and song: linguistic relatedness and geographic proximity. Metadata regarding both the language and geographic location associated with each listener and song allowed us to split trials into those in which language family or geographic region were shared and not shared. We tested the effect of these two proxies for cultural familiarity using mixed-effects models with a categorical fixed effect for whether a participant shares language families or geographical area with the song, and random effects for participant and song.

The results showed statistically significant effects of sharing a language family for discriminating dance, lullaby, and love songs ( $p < 0.05$ ), and not healing songs ( $p = 0.52$ ; Figure 3). However, these effects were very small, the largest of which was found for lullabies, where sharing a language family resulted in an estimated boost to “used for soothing a baby” ratings of 0.06 on the 1-4 rating scale. For context, this is equivalent to only ~2% of the whole rating scale, and only ~5% of the estimated difference between dance songs and lullabies on the “used for dancing” scale: that is, the magnitude of the effect of culture is minimal compared to the variance explained by actual song type and universal regularities in the songs’ musical features. There was a

similar pattern of results for the effect of sharing a geographical area, with marginally larger effects for dance, lullaby, and love songs ( $p < 0.05$ ). Here, the largest effect was found for sharing a geographical area when rating a dance song on the “used for dancing” scale, resulting in a 0.16 increase on the 1-4 scale (equivalent to ~4% of the rating scale); as such, like the effects of linguistic proximity, geographic proximity had a statistically significant but practically nonsignificant effect.

Given that culturally close groups are likely to share both a language and be in close geographic proximity, we also explored potential additive effects of sharing a language family and geographic subregion. To do this, we regressed the target song type ratings (on their relevant scale) onto two binary variables: language family (shared vs. not shared) and geographic subregion (shared vs. not shared). The interaction between the two variables was not significant for any of the four song types, meaning that the effect of sharing a geographic region was no different depending on whether the listener was also more familiar with the language of the song.

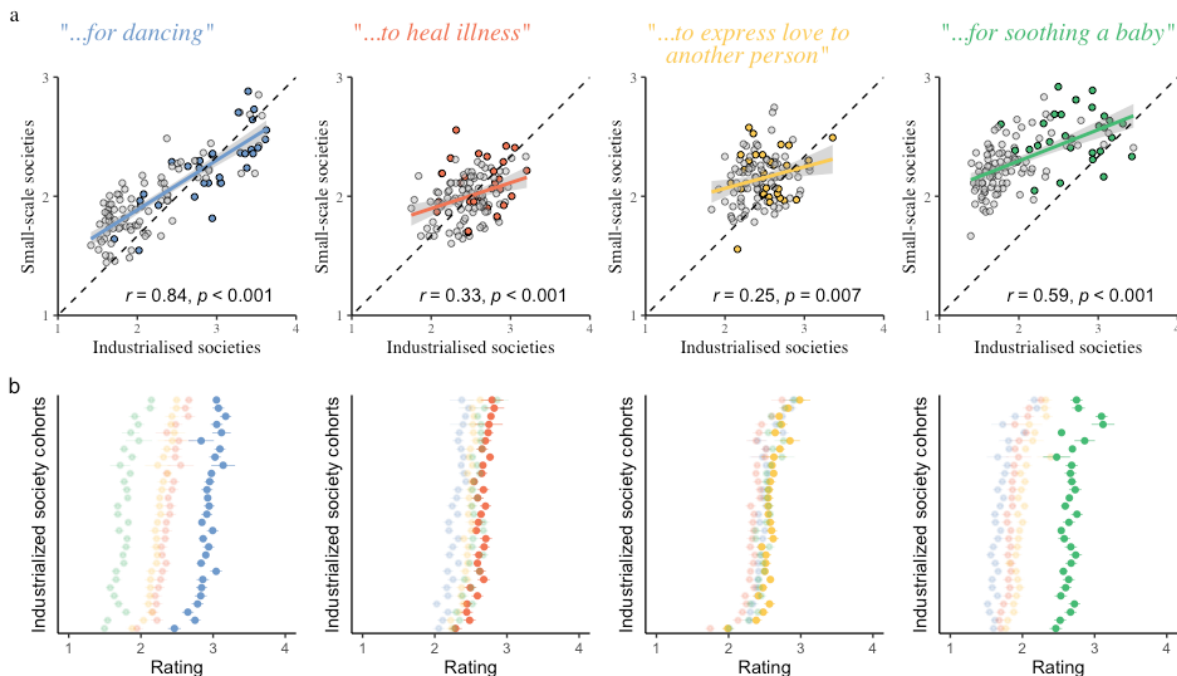


Figure 2. Comparison of listener ratings (a) across cohorts, and (b) within the industrialised society cohort. Panel (a) shows the correlation between the two cohorts’ ratings on a given dimension for all 118 songs. Coloured dots indicate the songs that are of the target song type for a specific dimension, i.e., dance songs in the leftmost plot. Panel (b) shows the ratings of each of the four song types on a given dimension, split by the industrialised society linguistic sub-cohorts, i.e., the leftmost plot shows how dance songs were rated on the four different dimensions (denoted by the four colours), by the 28 linguistic sub-cohorts.

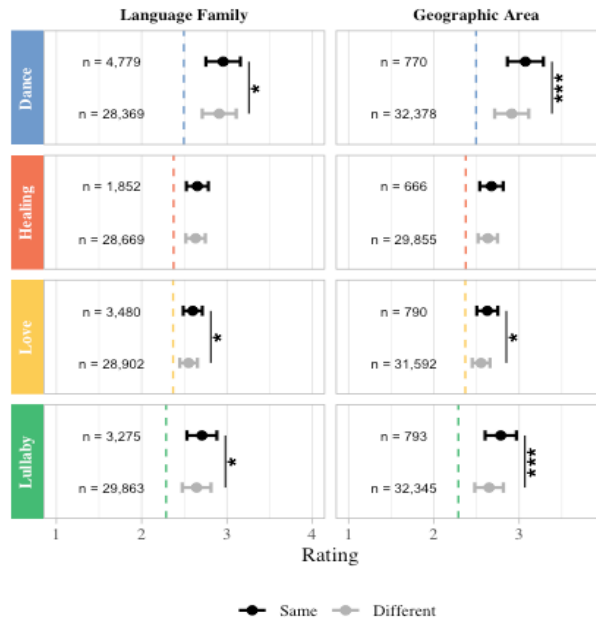


Figure 3. Visualising the results of mixed models testing the effect of shared language family on listeners' inferences. For three out of the four song types (dance songs, love songs and lullabies), shared language family between song and listener significantly increased the rating of a song on its relevant scale. *ns* reported are number of trials per category.

Significance: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05

## Discussion

In experiments conducted in 28 languages online to participants in 31 countries, and conducted in the field for 3 small-scale societies, we showed that people have reliable inferences about the behavioral functions of unfamiliar dance songs, lullabies, and healing songs. For example, songs that were used for dancing in their original cultural context were reliably rated highest on a “used for dancing” scale compared to five other possible behavioral contexts. This match between behavioral context and human inferences was also reliable for lullabies and healing songs, but in line with our predictions, was not reliable for love songs. Averaging across all song types, listener inferences about each behavioral context were remarkably consistent, showing reliable correlations not only within the 28 urban-society sub-cohorts but also between samples from urban societies to those in the three small-scale societies. The most highly correlated inferences were for whether a song was “used for dancing”, and the least correlated inferences were for “used for expressing love to another person”, but all were statistically significant.

The cross-cultural regularities in these particular kinds of musical inferences, and the fact that these were strongest for inferences about dance songs and lullabies, are consistent

with claims that music may have evolved to signal covert information in contexts related to dancing and infant care (Mehr et al., 2021). The plausibility of innate perceptual biases underlying these results is convergently supported by comparison to other domains in which these are well established, such as the cross-cultural intelligibility of emotional expression in vocalizations [Faragó et al. (2014); Laukka & Elfenbein (2021); Scherer 2001], music (Balkwill & Thompson, 1999; Cowen et al., 2020; Fritz et al., 2009; Sievers et al., 2021), and facial expressions (Cowen et al., 2021). Indeed, the communicative intent of some vocal signals are even detectable across species (Filippi et al., 2017; Kamiloğlu et al., 2020). Although we have not studied language here, we speculate that the perceptual and cognitive constraints leading to form-function regularities in music could be similar in kind to those underlying the surprisingly robust form-meaning regularities in the world's languages (Blasi et al., 2016; Cwiek et al., 2021; Imai & Kita, 2014; Sidhu & Pexman, 2018).

What role does culture-specific learning play in listeners' perceptions of foreign songs? Clearly, cultures differ dramatically with regards to their music, and it is reasonable to assume that cultural familiarity would influence the accuracy of listeners' inferences. We conducted a series of exploratory analyses that investigated the role of cultural proximity between listener and song. The two methods of quantifying cultural proximity revealed similar effects: Both linguistic proximity and geographic proximity produced small, but practically nonsignificant advantages for three out of the four song types.

It is notable that love songs, which were not clearly differentiated by listeners, also benefitted from cultural familiarity. We suspect that this category of song is highly culturally specific and ambiguous, given the wide variety of topics and sentiments covered in love songs. While this ambiguity may have made it harder for listeners to accurately differentiate love songs, cultural familiarity still improved listeners' inferences.

Our results speak to the idea that human song is shaped by both biological predispositions as well as the more obvious culture-specific nuances, building on a number of recent studies (Hilton, Crowley, et al., 2021; Hilton, Moser, et al., 2021; Mehr et al., 2019), and consistent with related work in cross-cultural emotion recognition, where meta-analyses have revealed universal recognition of basic emotional facial expressions, in tandem with smaller effects associated with judging members of a cultural or ethnic in-group (Elfenbein & Ambady, 2002; Laukka & Elfenbein, 2021). This study is unique in the variety of stimuli and the diversity of participants included; however, future work should aim to address these research questions with larger musical corpora and more song types. While cultural familiarity with the producer of a song conferred an in-group advantage on our listener's inferences, the magnitude of these effects paled in comparison to the variance explained by actual song type.

## References

- Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, 17(1), 43–64. <https://doi.org/10.2307/40285811>
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823. <https://doi.org/10.1073/pnas.1605782113>
- Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1910704117>
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589(7841), 251–257. <https://doi.org/10.1038/s41586-020-3037-7>
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., & Lippus, P. (2021). Novel vocalizations are understood across cultures. *Scientific Reports*, 11(1), 1–12.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203.
- Faragó, T., Andics, A., Deveszeri, V., Kis, A., Gácsi, M., & Miklósi, Á. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, 10(1), 20130926.
- Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele, M., Ocklenburg, S., Boer, B. de, Sturdy, C. B., Newen, A., & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, 284(1859). <https://doi.org/10.1098/osf.io/rspb.2017.0990>
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19(7), 573–576. <https://doi.org/10.1016/j.cub.2009.02.058>
- Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, 14(1), 21–51. <https://doi.org/10.1007/s12110-003-1015-z>
- Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae*, 13(2 suppl), 291–320. <https://doi.org/10.1177/1029864909013002131>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Bainbridge, C. M., Atwood, S., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., ... Mehr, S. A. (2021). Acoustic regularities in infant-directed speech and song across cultures. *PsyArXiv*. <https://doi.org/10.1101/2020.04.09.032995>
- Hilton, C., Crowley, L., Yan, R., Martin, A., & Mehr, S. (2021). Children infer the behavioral contexts of unfamiliar foreign songs. *PsyArXiv*. <https://doi.org/10.31234/osf.io/rz6qn>
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130298.
- Kamiloglu, R. G., Slocombe, K. E., Haun, D. B., & Sauter, D. A. (2020). Human listeners' perception of behavioural context and core affect dimensions in chimpanzee vocalizations. *Proceedings of the Royal Society B*, 287(1929), 20201148.
- Laukka, P., & Elfenbein, H. A. (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, 13(1), 3–11.
- Lomax, A. (1968). Folk song style and culture. *American Association for the Advancement of Science*.
- Mehr, S. A., Krasnow, M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*. <https://doi.org/10.31234/osf.io/nrqb3>
- Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, 366(6468), 957–970. <https://doi.org/10.31234/osf.io/emq8r>
- Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. *Current Biology*, 28(3), 356–368. <https://doi.org/10.1016/j.cub.2017.12.042>
- Murdock, G. P., Ford, C. S., Hudson, A. E., Kennedy, R., Simmons, L. W., & Whiting, J. W. M. (2008). Outline of cultural materials. *Human Relations Area Files, Inc*.
- Naroll, R. (1967). The proposed HRAF probability sample. *Behavior Science Notes*, 2(2), 70–80. <https://doi.org/10.1177/106939716700200202>
- Nettl, B. (1964). *Theory and method in ethnomusicology*. Collier-Macmillan.
- Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.
- Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, T. (2021). Music as a coevolved



- system for social bonding. *Behavioral and Brain Sciences*.  
<https://doi.org/10.31234/osf.io/qp3st>
- Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*, 33(1), 25–31.
- Sievers, B., Parkinson, C., Kohler, P. J., Hughes, J. M., Fogelson, S. V., & Wheatley, T. (2021). Visual and auditory brain areas share a representational structure that supports emotion perception. *Current Biology*.  
<https://doi.org/10.1101/254961>
- Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, 8(1), 40–46.  
<https://doi.org/10.1016/j.tics.2003.11.002>
- Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Adults identify infant-directed music across cultures. *Infant Behavior and Development*, 16(2), 193–211.  
[https://doi.org/10.1016/0163-6383\(93\)80017-3](https://doi.org/10.1016/0163-6383(93)80017-3)