Mutual intelligibility in musical communication

Lidya Yurdum[1,2,*], Manvir Singh[3], Luke Glowacki[4], Tom Vardy[5], Quentin D. Atkinson[5], Courtney B. Hilton[2,5], Disa Sauter[1], Max M. Krasnow[6], & Samuel A. Mehr[2,5,*]

[1]Department of Psychology, University of Amsterdam, Amsterdam 1018WT, Netherlands.
[2]Haskins Laboratories, Yale University, New Haven, CT 06511, USA.
[3]Institute for Advanced Study in Toulouse, 31080 Toulouse Cedex 6, France.
[4]Department of Anthropology, Boston University, Boston, MA 02215, USA.
[5]School of Psychology, University of Auckland, Auckland 1010, New Zealand.
[6]Division of Continuing Education, Harvard University, Cambridge, MA 02138, USA.

*Corresponding authors. E-mail: lidya.yurdum@yale.edu, sam@yale.edu

Despite the variability of music across cultures, some types of human songs share acoustic characteristics with one another. For example, dance songs tend to be loud and rhythmic and lullabies tend to be quiet and melodious. Human perceptual sensitivity to the behavioural contexts of songs on the basis of these acoustic features raises the possibility that basic properties of music are mutually intelligible, independent of linguistic or cultural content. Whether these effects reflect a universal perceptual phenomenon, however, is unclear, because prior studies focus almost exclusively on English-speaking participants, a group that is not representative of humans, writ large. Here we report shared intuitions concerning the behavioural contexts of unfamiliar songs produced in unfamiliar languages, in participants living in Internet-connected industrialised societies ($n = 5,516$ native speakers of 28 languages) or smaller-scale societies with limited access to global media ($n = 116$ native speakers of 3 non-English languages). Participants listened to songs randomly selected from a representative sample of human vocal music, originally used in four behavioural contexts, and rated the degree to which they believed the song was used for each context. Listeners in both industrialised and smaller-scale societies reliably inferred the contexts of dance songs, lullabies, and healing songs, but not love songs. Within and across the cohorts, inferences were mutually consistent. Further, increased linguistic or geographical proximity between listeners and singers only minimally increased the accuracy of the inferences. These results demonstrate that the behavioural contexts of three common forms of music are mutually intelligible across cultures and imply that musical diversity, shaped by cultural evolution, is nonetheless grounded in some universal principles.

Keywords: music, cross-cultural, universality, form, function, cultural evolution

Like many other animals, humans use vocalisations to convey their intentions and affective states (1, 2). Such vocalisations would be meaningless in a world where members of one's own — or other — species could not interpret these signals in a useful way. Indeed, many animal and human vocalisations are not arbitrary but instead display systematic relationships between their acoustic form and their behavioural function (2–4). For instance, the human scream is unlikely to have evolved arbitrarily as a means of communicating distress and urgency: rather, a scream involves extreme high frequencies (5) and acoustic roughness (6) that set it apart from regular verbal communication, and make it explicitly appropriate for the behavioural function of grabbing attention.

Such form-function relationships in human vocalisations allow listeners to infer a range of information about others, such as intention (7), emotion (8, 9), and physical prowess (10, 11). Form-function relationships in vocalisations even appear to be preserved across species: for instance, humans can infer the behavioural context and affect of chimpanzee vocalisations (12), and deer mothers are sensitive to the distress calls of a variety of mammals (13).

Systematic form-function relationships also apply to more complex vocalisations. Song is a human universal characterised by rich variability within and across cultures (14–16). Some of the behavioural contexts in which songs are used, however, are conspicuously similar around the globe, such as singing to soothe fussy infants, or singing to coordinate dancing (14, 17–22). Songs used for specific functions in specific behavioural contexts tend to display stereotyped acoustic features: for example, dance songs tend to share clearly accented and predictable beat structures. As with other types of vocalisations, form-function patterns in human song may originate from our evolved psychology, perceptual biases, or unique social environment (23–26). These constraints on cultural-evolutionary processes result in musical behaviours that show elements of cultural specificity while still remaining grounded in general biological tendencies (27, 28). The resulting regularities enable listeners to reliably infer the behavioural contexts of unfamiliar foreign music (14, 19), even young children, who have less musical experience relative to adults (21).

Notably, while prior experiments demonstrate that people can infer the behavioural contexts of songs from different cultures using their acoustic features, these studies frequently have sampling limitations. For instance, some studies rely primarily on English-speaking Western participants (17), and those that have reached participants around the world still rely on English speakers who have access to the Internet (14, 19, 20); n.b., this important problem affects many areas of the cognitive sciences (29). Thus, although the stimuli participants in these studies listened to were cross-culturally representative, it is unclear how much of the accuracy of listener inferences is accounted for by universal form-function links in musical behaviour, and how much is a product of (Western) enculturation, education, and exposure to world music through globalised media.

Here, we test the prediction that the behavioural contexts of songs are mutually intelligible to listeners across cultures. We study a large and diverse sample of listeners recruited worldwide in many languages, from both industrialised societies and smaller-scale societies. We use *smaller-scale* to refer to (i) societies in which individuals interact in a "small" world (i.e., 10-100 other individuals but not more), most interactions are face-to-face, and there is a high degree of interdependence); and (ii) societies less affected by states, markets, globalization, and/or world religions.

We predicted that listeners in both industrialised and smaller-scale societies would correctly infer the behavioural contexts of three types of unfamiliar songs (dance, lullaby, healing), reflecting a sensitivity to acoustic cues shared in these contexts across cultures (the preregistration is at https://osf.io/msvwz). In exploratory analyses, we asked whether culturally learned cues would give listeners an advantage when inferring the behavioural contexts of songs that are more closely related to their own culture, in line with other domains, such as the perception of emotion in vocalizations (9, 30).

Methods

Participants

Industrialised Societies (n = 5,516)

We partnered with Qualtrics Panels to recruit a global sample of participants that maximized linguistic and geographic diversity. We aimed for a minimum of 100 participants in each of 45 countries, who were native speakers of an official language of their country of residence, and who would complete the study in that language. In countries where official languages included both English and at least one non-English language, we planned to recruit only in the non-English language. For example, Zulu and English are both official languages of South Africa, but our goal was to recruit only South Africans who were native Zulu speakers and who would complete the study in Zulu.

As such, the participants studied included many native speakers of many non-English languages, along with native English speakers from countries where English is the primary official language, such as Australia (we did not recruit in the United States because prior work included many United States participants, 14, 21). The full list of languages and countries represented in the sample (after exclusions; see below) is in Table 1 and the approximate locations of the participants are visualised in Figure 1.

Note that in the cases of countries with multiple official languages, we were not always successful in our goal of only recruiting native speakers of non-English languages, due to recruitment difficulties. As a result, some participants in some countries were split across native language groupings. For example, the South

African sample included native speakers of both Zulu and English (contrary to our plan to include only native speakers of Zulu), whereas the Kenyan sample included only native speakers of Swahili (as planned). Further details on deviations from the preregistered recruitment plan are in SI Text 1.1.

We aimed to maximise data quality with eight planned exclusion criteria: we excluded participants who (i) performed poorly on a commonly used headphone detection task (31); (ii) reported difficulties hearing the audio on at least 4 of 24 trials (e.g., because of poor connectivity); (iii) had an IP address that did not geolocate to the same country they reported as their location; (iv) failed a simple attention check; (v) completed the survey more rapidly than would be possible; (vi) reported not wearing headphones; (vii) reported being in a noisy environment; or (viii) reported not being careful in completing the study. After exclusions, the sample included n = 5,516 native speakers of 28 languages, located in 49 countries.

Qualtrics Panels compensated each participant directly in the local currency, with rates varying across countries as a function of local payment standards for survey participation. All participants provided informed consent, under a protocol approved by the Harvard University Committee on the Use of Human Subjects (Ethics ID: IRB16-1080).

Language family	Language	Total n	Subregion	Country	Country-wise n
Afro-Asiatic	Amharic	33	Africa	Ethiopia	33
	Arabic	534	Africa	Egypt	133
				Morocco	133
			Middle East	Oman	2
				Saudi Arabia	133
				United Arab Emirates	131
			Western Europe	Belgium	2
Atlantic-Congo	Zulu	66	Southern Africa	South Africa	66
	Swahili	132	Eastern Africa	Kenya	132
Austroasiatic	Vietnamese	135	Southeast Asia	Vietnam	135
	Filipino	132	Southeast Asia	Philippines	132
	Indonesian	133	Southeast Asia	Indonesia	133
Indo-European	Bengali	133	South Asia	Bangladesh	27
				India	106
	Czech	133	Central Europe	Czech Republic	133
	Danish	133	Scandinavia	Denmark	133
	Dutch	178	Western Europe	Belgium	45
				Netherlands	133
	French	257	Western Africa	Benin	1
				Burkina Faso	4
				Cameroon	17
			Western Europe	Belgium	102
				France	133
	German	136	Central Europe	Austria	133
			Western Europe	Belgium	3
	English	819	Arctic and Subarctic	Canada	133
			Australia	Australia	133
			British Isles	United Kingdom	133
			Polynesia	New Zealand	133
			Southeast Asia	Singapore	133
			Southern Africa	Namibia	5
				South Africa	87

				Zambia	14
			Western Africa	Ghana	46
			Western Europe	Belgium	2
	Italian	125	Southern Europe	Italy	124
			Western Europe	Belgium	1
	Greek	133	Southeastern Europe	Greece	133
	Norwegian	133	Scandinavia	Norway	133
	Portuguese	297	Southern Europe	Portugal	134
			Southern South America	Brazil	163
	Romanian	135	Southeastern Europe	Romania	135
	Russian	141	Eastern Europe	Russian Federation	141
	Spanish	533	Northern Mexico	Mexico	133
			Northwestern South America	Colombia	133
			Southern Europe	Spain	133
			Southern South America	Argentina	134
	Ukrainian	133	Eastern Europe	Ukraine	133
	Urdu	133	South Asia	Pakistan	133
Japonic	Japanese	134	East Asia	Japan	134
Koreanic	Korean	134	East Asia	South Korea	134
Sino-Tibetan	Mandarin	266	East Asia	China	133
				Hong Kong	133
Turkic	Turkish	132	Southeastern Europe	Turkey	131
			Western Europe	Belgium	1
Uralic	Finnish	133	Scandinavia	Finland	133

Table 1. Linguistic and geographic information about the participants in the industrialised societies. Language information refers to the native language of the participant; languages were classified via Glottolog. Geographic information refers to the country of residence of the participant; world subregions were classified via the Human Relations Area Files.

Smaller-scale Societies (n = 116)

We recruited adult participants from the Nyangatom in Ethiopia (n = 35), the Mentawai in Indonesia (n = 30), and the Tannese Ni-Vanuatu in Vanuatu (n = 56), via word-of-mouth sampling. The approximate locations of each of these smaller-scale societies are visualised in Figure 1 and summary information about each is in Table 2. The societies were chosen for their reduced exposure to music from other cultural traditions. At the time of data collection (2017 to 2019), all three societies had somewhat limited access to TV, radio and the Internet and could not be assumed to have had significant exposure to these communication channels.* In each society, indigenous music continues to be widespread and central to cultural identity.

In the cases of 5 participants, an experimenter expressed concern as to whether the participant understood the task; these participants were excluded without the experimenter being aware of the songs heard. As in the industrialised cohort, participants were compensated directly in the local currency, with rates determined by the Principal Investigator at each site and in keeping with norms across other research projects conducted in the area. Ethics approval was granted by the Pennsylvania State University Office for Research Protections

*At the time of data collection, the three societies had somewhat limited access to TV, radio and the Internet and they varied in their familiarity with these technologies. The Nyangatom communities had little exposure to these technologies when the experiment was conducted, although exposure has since expanded considerably. The Ni-Vanuatu communities were exposed to Christian music in church, as well as reggae and other foreign music through battery powered radios and, over the last five years, increasing access to the Internet via cell phones. Nonetheless, traditional Kastom music is still widely performed in local religious and civil ceremonies and is an important part of Ni-Vanuatu culture and identity. The Mentawai communities studied encountered non-Mentawai music, particularly Indonesian and Bollywood music, through both radios and memory sticks purchased in the port-town, although both cell phone and radio ownership were rare.

(Ethics ID: STUDY00012265) for data collection in Ethiopia; the Institute for Advanced Study in Toulouse (Ethics ID: 2017-09-001) for data collection in Indonesia; and the University of Auckland Human Participants Ethics Committee (Ethics ID: 021538) for data collection in Vanuatu.

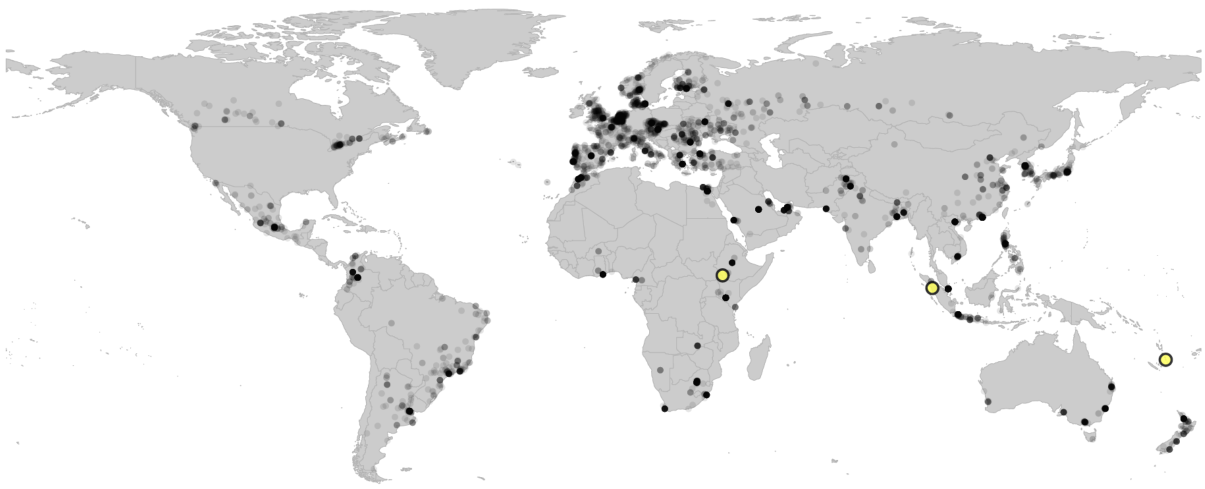

Figure 1 | Geographic distribution of participants. We recruited participants in industrialised societies and in three smaller-scale societies. The grey dots indicate the approximate locations of the participants in industrialised societies, as measured via IP geolocation. The yellow dots indicate the approximate locations of the three smaller-scale societies (from left to right, the Nyangatom, Mentawai Islanders, and Tannese Ni-Vanuatu).

Region	Sub-Region	Society	Language	Language family	Subsistence type	Approx. Community Size	Distance to city (km)	Final n
Africa	Eastern Africa	Nyangatom	Nyangatom	Nilotic	Pastoralist	155	180	34
Asia	Southeast Asia	Mentawai Islanders	Mentawai	Austronesian	Horticulturalist	260	120	27
Oceania	Melanesia	Tannese Ni-Vanuatu	Bislama	Indo-European Creole	Horticulturalist	6000	224	55

Table 2. Information about the three smaller-scale societies.

Materials

The stimuli were excerpts of each of the 118 songs in the *Natural History of Song Discography* (14), originally recorded in 86 mostly smaller-scale societies spanning 30 world regions (32, 33), over 75 languages, and a range of subsistence methods. The songs were originally used in four behavioural contexts: soothing a baby, dancing, expressing love and healing the sick. The contexts of the songs were determined on the basis of ethnographic descriptions alone (see 14 for full methods).

The excerpts were randomly selected 14-second segments of each song that contained singing (i.e., not instrumental-only sections), used in prior work (19). Readers can explore the *Discography* graphically at https://themusiclab.org/nhsplots or download the excerpts from https://doi.org/10.5281/zenodo.7265514.

Procedure

For each trial of the listening task, participants first listened to the full 14-second song excerpt. Afterward, they were prompted with the text "Think of the people making this music. I think that they. . . ", to which they could respond on a scale from 1 ("Definitely do not use the music. . . [context]") to 4 ("Definitely use the music. . . [context]"), where [context] referred to each of the four behavioural contexts represented in the corpus, i.e., "for dancing", "to soothe a baby", "to heal illness" and "to express love for another person". Two additional contexts that were not represented in the corpus were also included, as distractors ("to greet visitors" and "to praise a person's achievements"). The text was always presented in the participant's native language (see Translations, below).

Each participant heard a set of excerpts drawn from the corpus randomly and without replacement. In the industrialised cohort, participants heard 24 excerpts; in the smaller-scale societies, the experiment was shorter, with only 18 excerpts.

In the industrialised societies, participants completed the listening task via a Qualtrics survey displayed in their native language. It also included questions on the participants' gender, age, country, native language, the amount of time they spent per day on the Internet or listening to music, their perception of their own musical skills, and their familiarity with traditional music from around the world. The survey could be completed on a desktop computer or mobile device, but required participants to wear headphones (see Participants). Responses were collected by keypresses, screen taps, and/or mouse clicks.

In the smaller-scale societies, participants sat with an experimenter, who read instructions aloud in the participant's native language (Nyangatom, Mentawai, or Bislama) and recorded their responses on a laptop (see Figure S1). During the listening task, participants listened to the song excerpts on headphones (ensuring the experimenter was unaware of which stimuli were heard) and entered their responses by pressing one of three large buttons on a custom button box. The buttons were labelled with a sequence of circles in ascending size, to help participants remember the direction of the scale. Participants were first familiarised with the box, identifying the three buttons corresponding to the possible responses. At the end of the experiment, participants were asked to re-identify each button to confirm that they remembered the response labels. The experiment was controlled via E-Prime 2.0.10.356 (Psychology Software Tools, Inc.). The participants sat opposite the experimenter and could not view the laptop screen. Participants reported their gender before the listening task, but no further data were collected.

On the basis of piloting in the field, we simplified the task used in the smaller-scale societies by reducing the number of response options from 4 points to 3 points, and rephrased the prompt as a question (i.e., "Do you think they use the music for [context]?" with response options "no", "a little", and "yes"; see Translations). We also opted to include two additional distractor contexts, for a total of eight contexts per song (the six reported above along with the two distractors from (19): "to mourn the dead" and "to tell a story").

Translations

For the online experiment, all text was professionally translated by partners hired by Qualtrics Panels. These individuals and organisations hold two ISO certifications (ISO 17100:2015, ISO 9001:2008), which require that all translation processes and resources undergo regular external audits. We delivered an English-language survey to Qualtrics, whose partners translated the surveys using a standardised glossary. The translated files were then reviewed by a senior editor, whose native language was the same as that of the translation, before being returned to us. We and our collaborators and students manually reviewed the translated materials in the languages that we ourselves were fluent in, seeking out native speakers of as many of the languages as we were able to find through our university networks to provide an additional check on the translation quality. For all noted discrepancies, we worked with Qualtrics and their partners to re-evaluate and update the translation.

The translation procedures were similar for the smaller-scale societies, but our on-site researchers worked with local collaborators (who were native speakers of the local language) rather than third parties. In Ethiopia, the materials were translated into Nyangatom by two native speakers who work as translators, working together to reach consensus. In Indonesia, M.S. prepared the Mentawai translation with the aid of a research assistant competent in English and Mentawai; together they then discussed and corrected the translation with other native Mentawai speakers, and it was then back-translated into English by a third-party, with any remaining differences discussed until reaching agreement. In Vanuatu, a research assistant translated the English script into Bislama and a second research assistant then translated it back into English; discrepancies were discussed with both research assistants until reaching agreement. In all three smaller-scale societies, the English prompt that was translated took the form of a question (i.e., "Do you think they use the music for…" rather than "I think that they…"), as the prompt was read aloud to the participant rather than read on a screen.

Results

For both cohorts, we calculated song-wise mean scores across all participants on each behavioural context dimension. These scores reflected, on average, how likely the participants thought it was that each song was used in each of the six behavioural contexts. These song-wise averages were then z-scored.

Because each participant only heard a randomly selected subset of the corpus, the number of ratings averaged for each song in each cohort varied (industrialised societies: median = 1094.5 ratings, range = 917-1183 times; smaller-scale societies: median = 18, range = 8-28).

Three forms of song are mutually intelligible

First, we asked whether listeners could accurately infer the behavioural contexts of the songs, using the same analysis strategy as (19), which included similar data types: we tested whether each behavioural context (e.g., all the dance songs) was rated higher than the average rating across all songs, on its corresponding dimension (e.g., "…for dancing"), with multiple regressions with an intercept fixed at zero, where the z-transformed mean ratings for each song in each context were regressed onto binary variables denoting the actual behavioural contexts. This approach measures whether songs originally used in a given behavioural context were perceived to be *more* appropriate for that context than the average song in the corpus.[†]

Listeners from both the industrialised and smaller-scale societies discriminated three of the four behavioural contexts reliably above chance (Figure 2). This confirms the primary preregistered prediction and replicates prior findings in a much narrower sample (i.e., English-speaking Amazon Mechanical Turk participants; 19).

Response patterns across behavioural contexts were informative in both positive and negative directions. For example, the industrialised cohort rated dance songs 0.90 standard-deviations above the base rate of "…for dancing" responses (β_{danc} = 0.90, SE = 0.145, $p < .0001$), but rated lullabies 0.83 standard-deviations *below* the base rate (β_{baby} = -0.83, SE = 0.145, $p < .0001$). This suggests listeners inferred that completely unfamiliar dance songs were suitable for dancing, *but also that lullabies were not.* The reverse pattern was evident for "…to soothe a baby" responses, with lullabies rated 1.09 standard deviations above the base rate (β_{baby} = 1.09, SE = 0.139, $p < .0001$) and dance songs well below the base rate (β_{danc} = -0.62, SE = 0.139, $p < .0001$).

Despite the smaller sample sizes and minor differences in the method, similar patterns were evident in data from the smaller-scale societies. Dance songs were rated above the base rate of "…for dancing" (β_{dance} = 0.66, SE = 0.162, $p < .0001$), with lullabies below it (β_{baby} = -0.68, SE = 0.162, $p < .0001$); and lullabies were rated 0.75 standard-deviations above the base rate of "…to soothe a baby" (β_{baby} = 0.75, SE = 0.161, $p < .0001$).

In both cohorts, effects in healing songs were smaller, but still indicated reliable inferences, with ratings on "…to heal illness" above the base rate in both industrialised societies (β_{heal} = 0.49, SD = 0.18, p = 0.007) and smaller-scale societies (β_{heal} = 0.47, SD = 0.18, p = 0.01). Consistent with (19), neither of the cohorts' ratings of love songs on "…to express love for another person" were higher than the base rate, suggesting an

[†]For an alternative analysis approach using mixed models in the industrialised societies, see SI Text 1.2.

inability to accurately identify this behavioural context.[‡] (Industrialised societies: $\beta_{love} = 0.30$, $SD = 0.18$, $p = 0.1$; Smaller-scale societies: $\beta_{love} = 0.15$, $SD = 0.18$, $p = 0.41$).

In sum, these findings indicate that the behavioural contexts of dance songs, lullabies and healing songs recorded worldwide are intelligible to listeners in both industrialised and smaller-scale societies.

Listeners' intuitions about songs are similar, worldwide

We compared listeners' intuitions to one another in two ways. First, we compared the responses of listeners in the industrialised cohort to listeners in the smaller-scale society cohort. Second, we measured variation in listener responses varied across linguistic subgroups of the industrialised cohort.

Comparison of listeners across industrialised and smaller-scale societies

As an overall test of cross-cohort similarity, we simply computed Pearson correlations of the song-wise mean ratings on each dimension. The four correlations were positive and statistically significant (Figure 3a), but varied in magnitude, with the highest correlations in "…for dancing" ($r = 0.84$) and "…to soothe a baby" ($r = 0.59$).

As a robustness check, we repeated this analysis with an alternate approach, using stratified bootstrapping to estimate the variability in each correlation, given the much larger heterogeneity of the industrialised cohort (Figure S2). The findings repeated, with modestly attenuated effect sizes.

For a more conservative test for differences between the intuitions of listeners in the two cohorts, we compared the z-scored ratings of the industrialised cohort for each behavioural context on each dimension to those of the smaller-scale society cohorts, with t-tests (i.e., testing for mean differences of each of the 16 half-violins in Figure 2: 4 behavioural contexts × 4 dimensions). None of the 16 comparisons were statistically significant; the largest cohort-wise difference had $p = .09$, above the conventional alpha of .05 (and well above a more conservative Bonferroni-adjusted alpha for 16 comparisons of .003).

Thus, we found little evidence for cohort-wise differences in listener intuitions, and good evidence for cohort-wise similarities.

Internal consistency of the industrialised cohort

We measured how similar the responses of participants *within* the industrialised cohort were to one another with two approaches. In both cases, we split the industrialised society sample into 28 subgroups, based on the 28 different native languages spoken by the participants.

First, we re-ran the main song-wise analysis within each subgroup, providing (in effect) a 28-fold replication attempt of the main analysis for each of the four dimensions. The replications were generally successful (Figure 3b).

In 27 of the 28 linguistic subgroups, dance songs were rated significantly above the base rate of "…for dancing" (ps $< .001$); only the Korean-language subgroup did not rate dance songs significantly above the base rate across all songs ($p = 0.13$), but nevertheless rated the other three groups of songs as inappropriate for dancing (ps $< .0001$). All 0 linguistic subgroups rated lullabies above the base rate of "…to soothe a baby" (ps $< .0001$).

As in the main effects, results in healing songs were somewhat weaker, identified as most appropriate in the context of "…to heal illness" by 20 of the 28 subgroups (ps $< .05$). Only 12 subgroups rated love songs significantly higher (ps $< .05$) than the base rate of "…to express love for another person" across all songs.

[‡]In a forced-choice version of this task, English-speaking citizen-science participants *did* reliably identify love songs (14), albeit with a small effect size. Love songs are apparently a rather ambiguous category, worldwide.

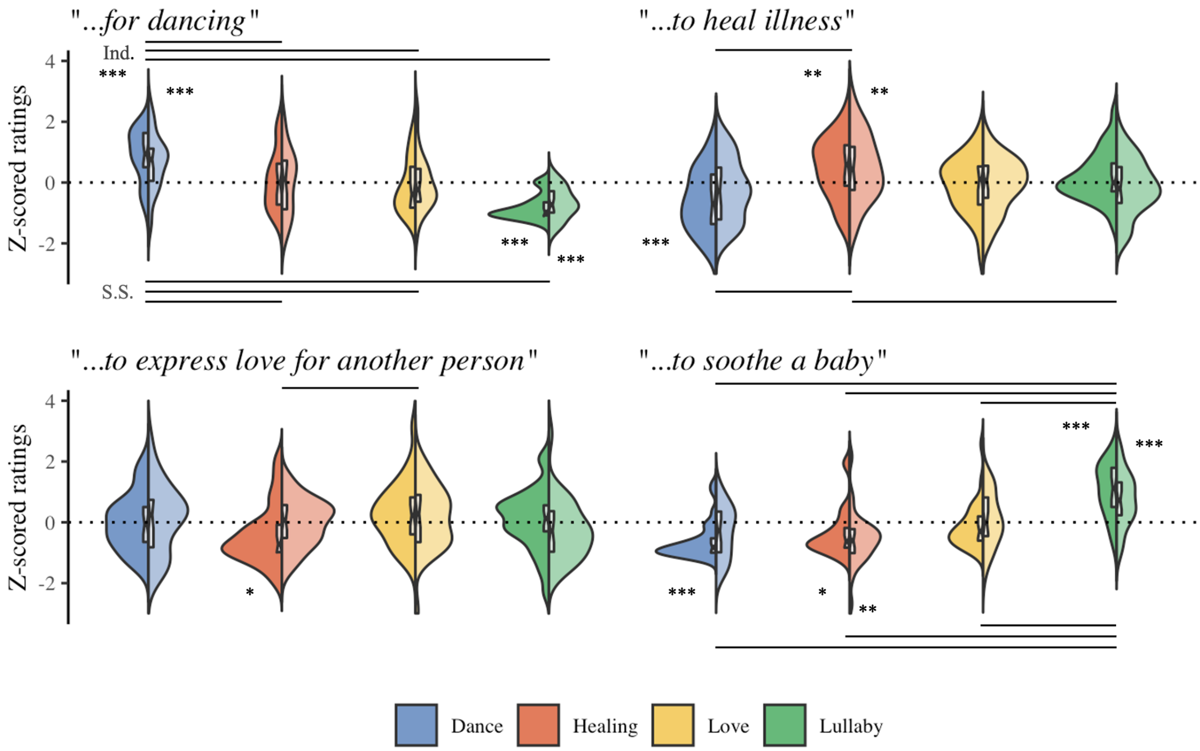

Figure 2 | The behavioural contexts of songs found worldwide are detectable by listeners recruited worldwide. Listeners heard a random selection of songs originally produced in one of four behavioural contexts: songs that were used "for dancing", "to heal illness", "to express love for another person", or "to soothe a baby". For each song, they were unaware of the culture or the behavioural context in which it was recorded. Each of the four plots visualises the distributions of mean song-wise ratings for a particular behavioural context dimension (e.g., "...for dancing"). The paired half-violins in each plot correspond to the four behavioural contexts, i.e., the actual behavioural contexts in which the songs originally appeared, denoted by colour. Each of the half-violins corresponds to the mean song-wise ratings from each of the two types of participants (i.e., from industrialised societies or smaller-scale societies). All ratings were z-scored, with a score of 0 indicating the average rating on a given dimension, across all songs, regardless of the songs' original behavioural context. For dance songs, lullabies, and healing songs, the ratings of listeners in both types of societies accurately reflected the original behavioural context of the songs (e.g., dance songs, but not the other three behavioural contexts, were rated significantly above average on the dimension "...for dancing"), indicated by the stars on either side of a violin, which compare the z-scored rating to the value 0. The horizontal lines beetween violin plots denote significant differences in ratings between behavioural contexts, and are split by cohort type, indicated by "Ind." (industrialised) or "S.S." (smaller-scale). For example, participants in the industrialised cohort (Ind.) rated healing songs as significantly lower on the "used to express love for another person" dimension, relative to love songs; whereas participants in the smaller-scale society cohort (S.S.) did not rate the two differently. The shaded area in the half-violins represent kernel density estimates; the vertical boxplots denote the median (horizontal line), 95% confidence interval (notches), and interquartile range (edges of the boxes), all computed cohort- and song-wise within each plot. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

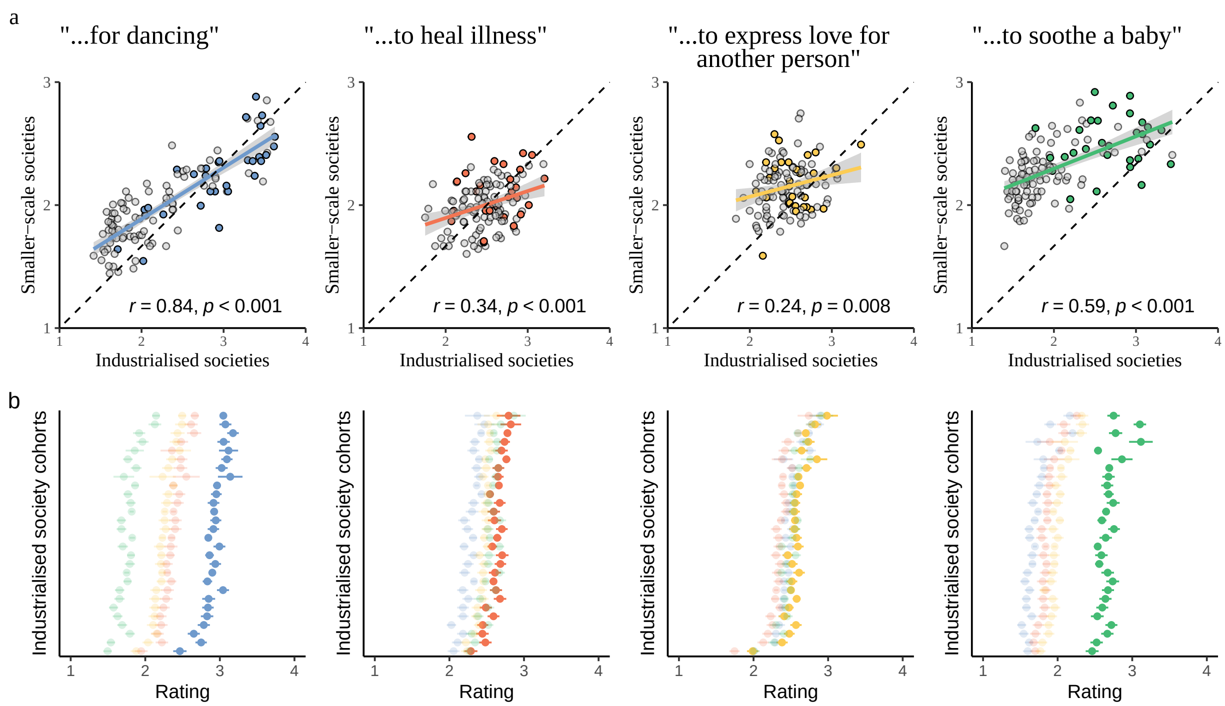

Figure 3 | Consistency of listeners' intuitions across cohorts and across languages. a, The mean song-wise ratings of listeners in the industrialised and smaller-scale societies, across the full corpus of songs, correlated with one another, on each of the four dimensions of interest. In the scatterplots, each point denotes a song-wise mean plotted in terms of its rating by participants in the industrialised societies (x-axis) and participants in the smaller-scale societies (y-axis). The highlighted dots denote songs whose behavioural context corresponds with the dimension of that plot (e.g., the blue points in the left-most plot, "...for dancing", denote dance songs). The line, shaded 95% confidence band, and associated statistics in each plot are computed via simple linear regressions. The diagonal dashed line indicates a hypothetical 1:1 relationship between the two cohorts. Note that participants in the smaller-scale societies used a 3-point scale rather than a 4-point scale; see Methods. **b** Within each linguistic subgroup of the industrialised societies, the main effects repeated consistently. The forest plots show the mean ratings of songs originally used in each of the four behavioural contexts, on each of the dimensions (one per plot), within each of the 28 linguistic subgroups (i.e., each row of points summarises data from one subgroup, such as native speakers of Urdu). For instance, the rightmost plot shows that lullabies (in green) were rated higher on the dimension "...to soothe a baby" in all 28 subgroups. The colours of the points correspond to the behavioural contexts, using the same scale as Figure 2 (dance songs in blue, healing songs in red, love songs in yellow, and lullabies in green).

Second, we used a similar correlation approach to the one reported above to measure the range of similarities. We built bootstrap samples of correlations between randomly selected pairs of linguistic subgroups, and tested the distribution of correlations against a null hypothesis of mean $r = 0$. The correlations were high for all four dimensions ("…for dancing": mean $r = 0.88$; "…to soothe a baby": mean $r = 0.84$; "…to heal illness": mean $r = 0.61$; "…to express love for another person": mean $r = 0.59$; all $ps < 0.0001$).

In sum, the intuitions of listeners worldwide (both across industrialised and smaller-scale societies and within industrialised societies) were similar to one another.

Cultural proximity is relatively uninformative to listeners

While we have shown a number of similarities across the intuitions of listeners worldwide, last, we explored a possible factor that could explain *differences* between them: cultural proximity between listener and singer.

If culture-specific musical "rules" explain differences in a given song from the worldwide "norm" for songs in a given behavioural context (i.e., leading to variability in listener intuitions in the effects reported above) then one might expect clear relations between cultural familiarity and listener accuracy. Specifically, when listeners hear songs from cultures that are *more similar* to theirs, their intuitions about behavioural context in a song should more closely match that song's *actual* behavioural context.

To operationalise this hypothesis, we used two measures of cultural proximity between listener and song: linguistic and geographic distance. Phylogenetic distance between languages is often used to model cultural transmission of behaviours, such as camel-herding practices (34), linguistic features (35), or vocalisation styles (20). Research on universalities in emotional facial expressions, for instance, has found that cross-cultural emotion recognition is higher when the judge's native language is closer to that of the poser's (36). Complementing the linguistic-distance approach, we also used geographical distance as a proxy for cultural distance and between-group exposure, as physical distance may predict cultural similarity (30, 37). We used Glottolog (38) to classify local languages into language families and the Human Relations Area Files (ehrafworldcultures.yale.edu) World Sub-region typology to classify geographic location for each culture, as in previous research (14).

We split each participant's data into two sets of trials: (i) trials where the participant rated a song sung in a language from their own language family; and (ii) trials where the participant rated songs that were sung in a language from a different language family (for a full list of language families, see Table 1). For the geographic analysis, we did the same, but using world subregions.

For example, in a participant recruited in Istanbul, Turkey who speaks Turkish, a trial with a song sung in Turkmen would be marked as linguistically "shared", since both Turkmen and Turkish belong to the Turkic language family. A song sung in Greek would be marked as linguistically "different", since Greek is an Indo-European language. On the other hand, a trial with a song recorded in Greece would be marked as geographically "shared", since the song and participant belong to the same geographic subregion (Southeastern Europe). Linguistic and geographic markers of proximity can, but do not necessarily have to overlap, as in the case of the Turkish listener and Greek song.

We then tested the effect of these two proxies for cultural familiarity using mixed-effects models, with a categorical fixed effect for whether a participant shared a language family 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 ($\beta_{shared} = 0.05$, $SE = 0.022$, $p = 0.03$), lullaby ($\beta_{shared} = 0.06$, $SE = 0.028$, $p = 0.03$), and love songs ($\beta_{shared} = 0.05$, $SE = 0.024$, $p = 0.04$), but not healing songs ($\beta_{shared} = 0.02$, $SE = 0.032$, $p = 0.5$; Figure 4).

These effects were very small, however: the largest, found for lullabies, showed that sharing a language family resulted in an estimated boost to "used to soothe a baby" ratings of 0.06 on a 4-point scale — equivalent to only ~2% of the whole scale and only ~5% of the estimated difference between dance songs and lullabies on the "…for dancing" dimension. The magnitude of the effect of cultural proximity was minimal compared to the variance explained by the actual behavioural context and universal regularities in the songs' musical features.

Results were comparable for geographic proximity, with marginally larger effects for dance ($\beta_{shared} = 0.16$, $SE = 0.036$, $p < .0001$), lullaby ($\beta_{shared} = 0.14$, $SE = 0.039$, $p < .001$), and love songs ($\beta_{shared} = 0.07$, $SE = 0.035$, $p = 0.04$), and no significant effect for healing songs ($\beta_{shared} = 0.04$, $SE = 0.040$, $p = 0.27$). Here, the largest effect was found for sharing a geographical area when rating a dance song on the "used for dancing" dimension, resulting in a 0.16 increase on a 4-point scale (equivalent to ~4% of the scale). Like the effects of linguistic proximity, geographic proximity had a statistically significant but practically nonsignificant effect.

Because 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. Studying each of the four behavioural contexts in isolation, we regressed the listeners' ratings (from the dimension corresponding to that behavioural context, e.g., for dance songs, we studied the dimension "... for dancing") on 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 behavioural contexts, however, 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 (statistical reporting is in Table S1).

Discussion

In a global sample of people, residing in both industrialised and smaller-scale societies, and tested predominantly in non-English languages, we find that listeners' inferences about the behavioural contexts of unfamiliar, foreign songs are *accurate*, *similar to one another*, and *relatively uninfluenced by cultural proximity*. Some basic aspects of musical communication are therefore mutually intelligible. These findings generalise prior findings reporting the ability of English-speaking participants recruited online to reliably infer the behavioural contexts of dance, lullaby, and healing songs (14, 19), thereby providing strong evidence for the generality of the effects and for the universality of the phenomenon.

The practice in cognitive science of focusing solely on English speakers is all-too-common (29). We note that the use of multiple samples of non-English speakers in the same experiment affords the ability to conduct mini-meta-analyses of key effects. Here, in the case of the participants in industrialised societies, for example, the approach enabled a 28-fold replication of the main analysis, in each linguistic subgroup. The approach also afforded tests of the cross-linguistic consistency of listeners inferences (for the fit of songs to the contexts "... for dancing" and "... to soothe a baby", consistency was especially high), justifying claims about *human* psychology, as opposed to the psychology of a Western, educated, industrialized, rich and democratic (WEIRD) subset of humans (39, 40).

Universal musical inferences, strongest for the contexts of dance and infant care, support theories that music may evolved to signal covert information in these particular contexts (23, 24, 26), united by the idea that music is a credible signal of a similar kind to vocal signals found across the animal kingdom (25).

The possibility of universal perceptual mechanisms for musical communication is bolstered by comparisons to other domains, where such mechanisms are already well-established, such as the cross-cultural intelligibility of emotional expression in vocalizations (e.g., 9, 36), including across species (12, 41, 42); facial expressions (e.g., 43); and non-referential information in music (44–47). 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 strikingly robust form-function relations in speech worldwide (20, 48–51).

The finding that positive effects of culturally learned cues were detectable in our data — but only with fleeting effect sizes — provides further evidence that, at least at a basic level of listeners decoding the *functions* of singers' vocalisations, music operates in a fashion similar to these other communicative domains. We note, however, that significant cultural variability nevertheless still exists among cultures that share the same language family or geographic subregion; a stronger test of the role of culture in mediating the intelligibility of music would involve comparing performance on songs from one's *own* culture to those from distant cultures. More cross-cultural experiments, perhaps relying on music with obscured or masked lyrics (because linguistic

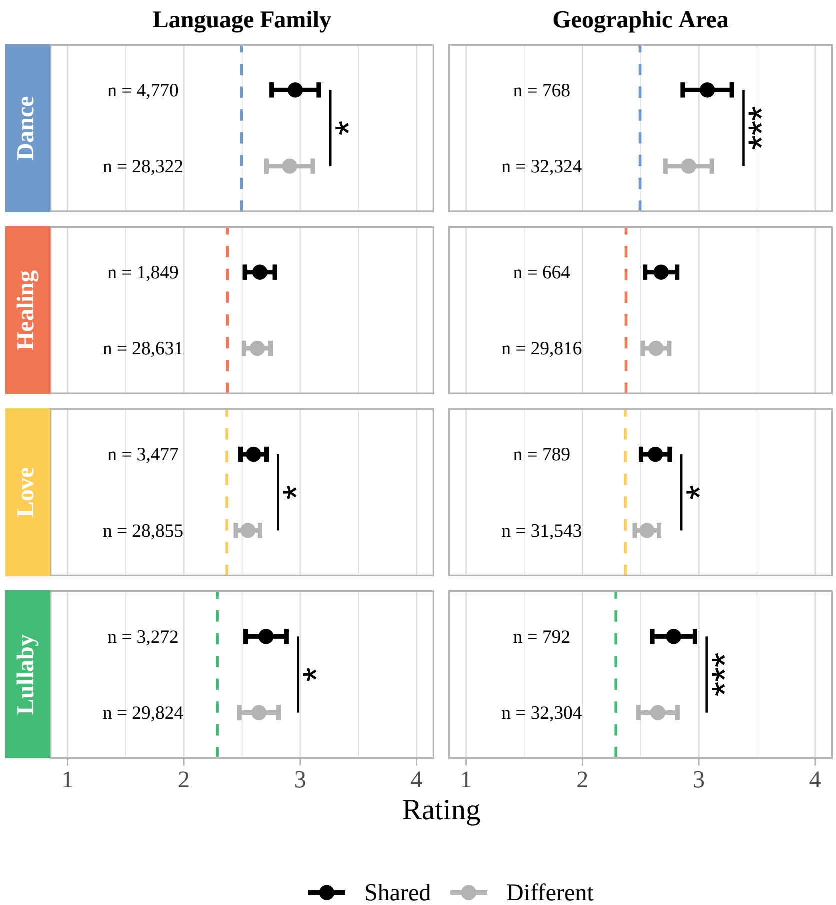

Figure 4 | Increased linguistic or geographic proximity between listeners and singers does not substantially improve performance. Because both songs and listeners came from global samples, in some cases, the culture of the listener is *more related* to the culture of the singer than others. This could, in principle, make it easier for listeners to make inferences concerning the behavioural context of unfamiliar songs. We found little evidence for such an effect, however. Each panel plots the estimated rating of a behavioural context on its corresponding dimension (e.g., dance songs on the "for dancing" dimension). The black point denotes the estimated rating when the listener and song *share* a linguistic family (left) or geographic sub-region (right), and the grey point is the estimated rating when the listener and song *do not share* a linguistic family (left) or geographic sub-region (right). The error bars denote 95% confidence intervals. In three out of the four behavioural contexts (dance songs, love songs and lullabies), both proxies for cultural familiarity with the song increased listeners' ratings of the correct behavioural context dimension by a statistically significant, but practically nonsignificant amount. The ns denote numbers of trials per category, not numbers of participants. The vertical dashed lines indicate the average rating across all songs, regardless of original behavioural context. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

content is a strong cue to behavioural context in music), may further explore the roles that culture plays in shaping music perception.

One area of evident ambiguity in the data reported here is listeners' difficulty, in both cohorts, of recognizing when music was being used in the context of "expressing love for another person". Our previous studies have provided conflicting evidence for this ability, apparently varying as a function of the task design (with negative effects on a rating scale, 19, and small, but positive effects in a forced-choice task, 14). Perhaps love songs are a particularly fuzzy category of music when sung in an unfamiliar language. In the present results, despite not reliably identifying love songs, listeners did perform slightly better when listening to songs of higher linguistic or geographic proximity, suggesting that cultural familiarity can shape listeners' intuitions in ambiguous music. The widespread prevalence of love songs in modern popular music presents a puzzle, given this context, of potential interest to music researchers.

This work speaks to the idea that music is shaped by both biological predispositions as well as culture-specific nuances, building on a number of recent studies (14, 19–22) and consistent with related findings in the domains of emotion (9, 30) and language (20, 48–51).

End notes

Data, code, and materials availability

A reproducible R Markdown manuscript is available at https://github.com/themusiclab/intelligible-music, with all associated data and materials. The same repository includes code for running the listener task in Qualtrics (for the industrialised societies) and E-Prime (for the smaller-scale societies), including translations of all experiments. This repository will be archived on Zenodo upon publication of this manuscript. The excerpted audio corpus (the Natural History of Song Discography) is available at https://doi.org/10.5281/zenodo.7265514.

Acknowledgments

This research was supported by the Harvard University Department of Psychology (M.M.K. and S.A.M.); the Harvard Data Science Initiative (S.A.M.); the National Institutes of Health Director's Early Independence Award DP5OD024566 (L.Y., C.B.H., and S.A.M.); the Institute for Advanced Study in Toulouse, under an Agence nationale de la recherche grant, Investissements d'Avenir ANR-17-EURE-0010 (M.S. and L.G.); and the Royal Society of New Zealand Te Aparangi Rutherford Discovery Fellowships RDF-UOA1101 (T.A.V. and Q.D.A.) and RDF-UOA2103 (S.A.M.). We thank the participants; J. Stieglitz and C. Scaff for their efforts at additional data collection; S. Atwood and C. Bainbridge for research assistance; and members of The Music Lab for feedback on the paper.

Author contributions

- S.A.M. and M.M.K. conceived of the research, hired and supervised research assistants, and coordinated the research team.
- S.A.M. and M.M.K. designed the protocol for running the study both online and in the three field sites, with input from M.S. and L.G., who piloted it in the field.
- S.A.M. and M.M.K. provided funding, coordinated the translation of materials, and supervised data collection in the industrialised societies.
- M.S., L.G., T.V., and Q.D.A. provided funding, translated the experiment materials, coordinated recruitment, and collected data in the smaller-scale societies.
- L.Y. led analyses, with contributions from C.B.H.
- C.B.H. conducted code review.
- L.Y., C.B.H., and S.A.M. designed the figures.

- L.Y. wrote the manuscript with contributions from S.A.M., D.S., M.S., and C.B.H.
- All authors edited the manuscript and approved it.

Supplementary Text

1.1 Deviations from the preregistration

We preregistered the study in November 2017 at https://osf.io/msvwz. The data collected and analyses reported deviate from the preregistration in five ways.

1. In the industrialised societies, we planned to collect data from 100 participants in each of 45 countries. Recruitment difficulties in some countries led us to increase the sampling range to include nearby countries where the same targeted language was also an official language. For instance, while we initially intended to recruit native speakers of English in Zambia, our sample from this region also included native speakers of English in nearby Namibia. This approach primarily affected African countries where internet access was limited relative to, e.g., the East Asian countries where we collected data via the same method.

2. In the smaller-scale societies, we planned to collect data in six communities, but due to the COVID-19 pandemic, we were only able to collect data in three.

3. For all participants, we planned the listening task to include 36 songs. This proved to be too long; in industrialised societies we shortened it to 24 songs, and in smaller-scale societies, to 18 songs.

4. To further shorten the task in industrialised studies, we reduced the number of dimensions on which participants rated each song; we planned to use four distractor dimensions but included only two in the full sample. The two we omitted ("...to tell a story" and "...to mourn the dead") had previously been studied in (19). Participants in the smaller-scale societies completed all four distractor dimensions for each song, however.

5. We planned to collect data concerning listeners' intuitions surrounding two forms of songs: the original, naturalistic recordings from the *Natural History of Song Discography* as well as artificially produced (i.e., synthesised) versions of the songs, created using transcriptions of them reported in (14). Due to limitations on the amount of data we could collect, we obtained far less data on listeners' responses to the synthesised songs than the naturalistic recordings. As such, we leave those data for a future paper. Note that this decision limited the number of participants in smaller-scale societies reported here, as roughly half of the participants studied in those societies heard synthesised songs rather than the naturalistic recordings.

1.2 Replication of confirmatory analyses using mixed-effects models

We replicated the main analyses of the industrialised society data using mixed-effects models with random intercepts for participant, song, and language. The results of these models conceptually replicated the simpler confirmatory analyses, but with slightly attenuated effect sizes.

Taking into account the variance accounted for by participant, stimulus and language, dance songs were rated significantly above the base rate on the "...for dancing" scale (β_{dance} = 0.49, SE = 0.08, $t(128.27)$ = 5.93, $p < .0001$), and lullabies were rated significantly below the base rate (β_{baby} = -0.46, SE = 0.08, $t(128.27)$ = -5.52, $p < .0001$). On the "...to soothe a baby" scale, lullabies were rated highest (β_{baby} = 0.52, SE = 0.07, $t(131.92)$ = 7.57, $p < .0001$), and dance songs lowest (β_{dance} = -0.29, SE = 0.07, $t(131.92)$ = -4.15, $p < .0001$); as in the song-level analyses, healing songs were rated below the base rate (β_{heal} = -0.17, SE = 0.07, $t(131.06)$ = -2.36, p = 0.02). On the "...to heal illness" scale, healing songs were rated higher than the base rate (β_{heal} = 0.12, SE = 0.05, $t(132.26)$ = 2.30, p = 0.02), whereas dance songs were rated below it (β_{dance} = -0.18, SE = 0.05, $t(133.01)$ = -3.54, $p < .001$). Last, as in the song-level analyses, love songs were not reliably identified as "...to express love for another person" ($p > 0.05$).

Figure S1 | Testing setup in smaller-scale societies. The photo depicts author M.S. testing a Mentawai participant in Indonesia. In each of the smaller-scale societies, participants sat across from the experimenter, listened on headphones only, and entered their responses on a button box. The experimenter was unaware of the song being played on each trial and the participant could not see the laptop's screen.

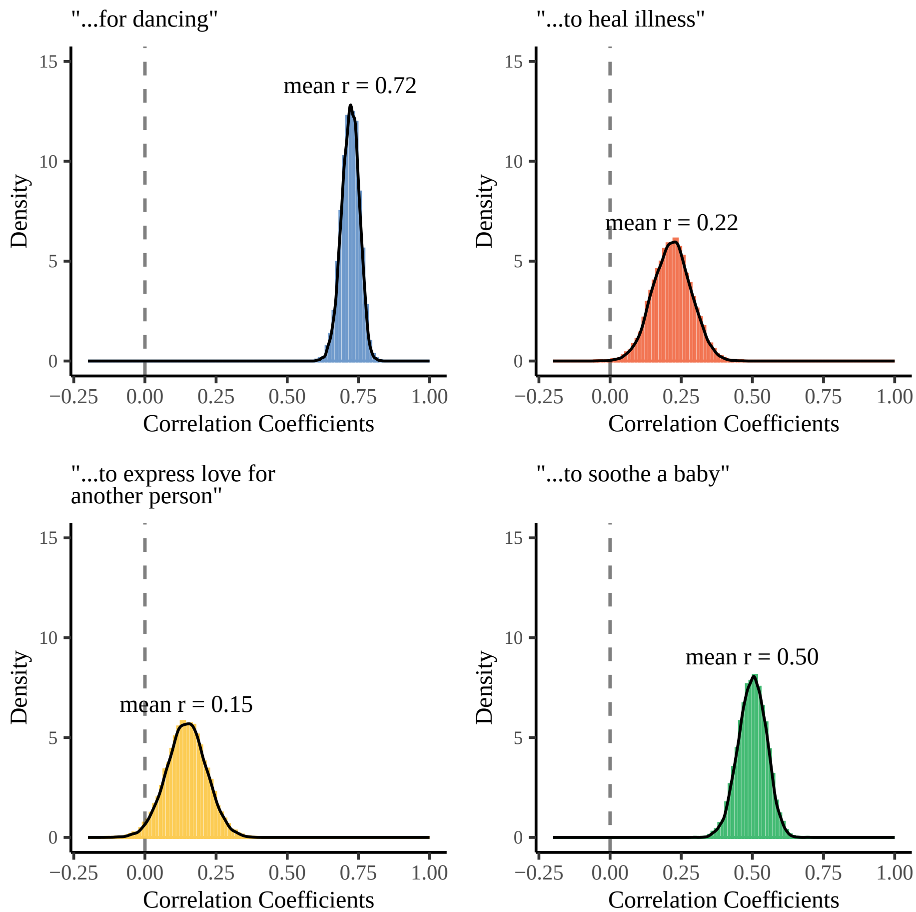

Figure S2 | Bootstrapped correlations between song-wise ratings from the industrialised societies and the smaller-scale societies. As an alternative to the simple correlations across cohort types, reported in the main text, we computed distributions of correlations via stratified bootstrapping. This approach helps to account for large differences in sample sizes between the cohorts and provides a principled estimate of the variability in each correlation coefficient. We sampled 30 observations per song from each cohort, generated new song-wise averages, and correlated these averages across both cohorts. This procedure was repeated 10,000 times. The plots show the four distributions correlations; in all four cases, the correlations were significantly larger than 0, but they varied in magnitude across behavioural contexts.

	Estimate	Std. Error	df	t value	p value
Dance songs					
Intercept	2.91	0.10	29.24	28.51	0.000
Shared Language	0.04	0.02	32320.46	1.99	0.046
Shared Sub-Region	0.17	0.04	31136.96	4.08	0.000
Interaction	-0.04	0.08	31516.69	-0.54	0.588
Lullabies					
Intercept	2.64	0.09	29.35	30.78	0.000
Shared Language	0.05	0.03	31529.05	1.72	0.086
Shared Sub-Region	0.10	0.04	31285.53	2.17	0.030
Interaction	0.15	0.09	31664.08	1.66	0.097
Healing Songs					
Intercept	2.63	0.06	27.65	45.19	0.000
Shared Language	0.02	0.03	27283.82	0.55	0.581
Shared Sub-Region	0.05	0.05	28487.64	1.00	0.319
Interaction	-0.03	0.10	28663.04	-0.31	0.757
Love Songs					
Intercept	2.55	0.05	30.10	48.06	0.000
Shared Language	0.05	0.02	28053.63	1.90	0.057
Shared Sub-Region	0.07	0.04	29923.48	1.84	0.066
Interaction	-0.03	0.08	30229.86	-0.42	0.674

Table S1 To test for a super-additive effect of linguistic and geographic proximity, we regressed the target behavioural context ratings (on their relevant dimension) onto two binary variables: language family (shared vs. different) and geographic subregion (shared vs. different), with random intercepts for participant and song. After including both language family and geographic subregion in the regression, sharing a language predicted higher ratings for dance songs only. Geographic proximity was associated with higher ratings on the appropriate dimensions for lullabies and dance songs. Super-additivity would be indicated by a significant interaction between the effect of linguistic and geographic proximity, such that the effect of sharing a geographic region depends on whether the listener is also more familiar with the language of the song. However, the interaction between the two variables was not significant for any of the four behavioural contexts.

References

1. E. S. Morton, On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist* **111**, 855–869 (1977).
2. K. Pisanski, G. A. Bryant, C. Cornec, A. Anikin, D. Reby, Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution* **34**, 303–321 (2022).
3. J. A. Endler, Some general comments on the evolution and design of animal communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **340**, 215–225 (1993).
4. W. T. Fitch, J. Neubauer, H. Herzel, Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour* **63**, 407–418 (2002).
5. K. Pisanski, J. Raine, D. Reby, Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society open science* **7**, 191642 (2020).
6. L. H. Arnal, A. Flinker, A. Kleinschmidt, A.-L. Giraud, D. Poeppel, Human screams occupy a privileged niche in the communication soundscape. *Current Biology* **25**, 2051–2056 (2015).
7. G. A. Bryant, H. C. Barrett, Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science* **18**, 746–751 (2007).
8. H. C. Barrett, G. Bryant, Vocal Emotion Recognition Across Disparate Cultures. *Journal of Cognition and Culture* **8**, 135–148 (2008).
9. P. Laukka, H. A. Elfenbein, Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review* **13**, 3–11 (2021).
10. J. Raine, K. Pisanski, R. Bond, J. Simner, D. Reby, Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS One* **14**, e0213034 (2019).
11. A. Sell, *et al.*, Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 3509–3518 (2010).
12. R. G. Kamiloğlu, K. E. Slocombe, D. B. Haun, D. A. Sauter, Human listeners' perception of behavioural context and core affect dimensions in chimpanzee vocalizations. *Proceedings of the Royal Society B* **287**, 20201148 (2020).
13. S. Lingle, T. Riede, Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *The American Naturalist* **184**, 510–522 (2014).
14. S. A. Mehr, *et al.*, Universality and diversity in human song. *Science* **366**, 957–970 (2019).
15. B. Nettl, *Theory and method in ethnomusicology* (Collier-Macmillan, 1964).
16. A. Lomax, *Folk song style and culture* (American Association for the Advancement of Science, 1968).
17. S. E. Trehub, A. M. Unyk, L. J. Trainor, Adults identify infant-directed music across cultures. *Infant Behavior and Development* **16**, 193–211 (1993).
18. S. E. Trehub, A. M. Unyk, L. J. Trainor, Maternal singing in cross-cultural perspective. *Infant Behavior and Development* **16**, 285–295 (1993).
19. S. A. Mehr, M. Singh, H. York, L. Glowacki, M. M. Krasnow, Form and function in human song. *Current Biology* **28**, 356–368 (2018).
20. C. B. Hilton, *et al.*, Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour* (2022) https:/doi.org/10.1101/2020.04.09.032995.
21. C. B. Hilton, L. Crowley-de Thierry, R. Yan, A. Martin, S. A. Mehr, Children infer the behavioral contexts of unfamiliar foreign songs. *Journal of Experimental Psychology: General* (2022).
22. M. Singh, S. A. Mehr, Universality, domain-specificity, and development of psychological responses to music. *Nature Reviews Psychology.*
23. E. H. Hagen, G. A. Bryant, Music and dance as a coalition signaling system. *Human Nature* **14**, 21–51 (2003).

24. E. H. Hagen, P. Hammerstein, 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**, 291–320 (2009).

25. S. A. Mehr, M. Krasnow, G. A. Bryant, E. H. Hagen, Origins of music in credible signaling. *Behavioral and Brain Sciences* (2021) https:/doi.org/10.31234/osf.io/nrqb3.

26. S. A. Mehr, E. S. Spelke, Shared musical knowledge in 11-month-old infants. *Developmental Science* **21** (2017).

27. P. J. Richerson, R. Boyd, *Not by genes alone: How culture transformed human evolution* (University of Chicago Press, 2008).

28. D. Sperber, L. A. Hirschfeld, The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences* **8**, 40–46 (2004).

29. D. E. Blasi, J. Henrich, E. Adamou, D. Kemmerer, A. Majid, Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences* (2022).

30. H. A. Elfenbein, N. Ambady, On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological bulletin* **128**, 203 (2002).

31. K. J. P. Woods, M. H. Siegel, J. Traer, J. H. McDermott, Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 1–9 (2017).

32. G. P. Murdock, *et al.*, *Outline of cultural materials* (Human Relations Area Files, Inc., 2008).

33. R. Naroll, The proposed HRAF probability sample. *Behavior Science Notes* **2**, 70–80 (1967).

34. R. Mace, *et al.*, The comparative method in anthropology [and comments and reply]. *Current anthropology* **35**, 549–564 (1994).

35. M. Dunn, S. J. Greenhill, S. C. Levinson, R. D. Gray, Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79–82 (2011).

36. K. R. Scherer, R. Banse, H. G. Wallbott, Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology* **32**, 76–92 (2001).

37. A. Wood, M. Rychlowska, P. M. Niedenthal, Heterogeneity of long-history migration predicts emotion recognition accuracy. *Emotion* **16**, 413 (2016).

38. H. Hammarström, R. Forkel, M. Haspelmath, *Glottolog 4.0* (Max Plank Institute for the Science of Human History, 2019).

39. J. Henrich, S. J. Heine, A. Norenzayan, The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61–83 (2010).

40. T. Yarkoni, The generalizability crisis. *Behavioral and Brain Sciences* **45** (2022).

41. T. Faragó, *et al.*, Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology letters* **10**, 20130926 (2014).

42. P. Filippi, *et al.*, 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** (2017).

43. A. S. Cowen, *et al.*, Sixteen facial expressions occur in similar contexts worldwide. *Nature* **589**, 251–257 (2021).

44. L.-L. Balkwill, W. F. Thompson, A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception* **17**, 43–64 (1999).

45. A. S. Cowen, X. Fang, D. Sauter, D. Keltner, 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* (2020) https:/doi.org/10.1073/pnas.1910704117.

46. T. Fritz, *et al.*, Universal recognition of three basic emotions in music. *Current Biology* **19**, 573–576 (2009).

47. B. Sievers, L. Polansky, M. Casey, T. Wheatley, Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences* **110**, 70–75 (2013).

48. D. M. Sidhu, P. M. Pexman, Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience* **33**, 25–31 (2018).

49. M. Imai, S. Kita, The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences* **369**, 20130298 (2014).

50. D. E. Blasi, S. Wichmann, H. Hammarström, P. F. Stadler, M. H. Christiansen, Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* **113**, 10818–10823 (2016).

51. A. Ćwiek, *et al.*, Novel vocalizations are understood across cultures. *Scientific reports* **11**, 1–12 (2021).