PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Universal interpretations of vocal music

Lidya Yurdum[a,b,1], Manvir Singh[c], Luke Glowacki[d], Thomas Vardy[e], Quentin D. Atkinson[e], Courtney B. Hilton[e], Disa Sauter[b], Max M. Krasnow[f], and Samuel A. Mehr[a,e,1]

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received October 31, 2022; accepted June 21, 2023

Despite the variability of music across cultures, some types of human songs share acoustic characteristics. For example, dance songs tend to be loud and rhythmic, and lullabies tend to be quiet and melodious. Human perceptual sensitivity to the behavioral contexts of songs, based on these musical features, suggests that basic properties of music are mutually intelligible, independent of linguistic or cultural content. Whether these effects reflect universal interpretations of vocal music, however, is unclear because prior studies focus almost exclusively on English-speaking participants, a group that is not representative of humans. Here, we report shared intuitions concerning the behavioral contexts of unfamiliar songs produced in unfamiliar languages, in participants living in Internet-connected industrialized societies (n = 5,516 native speakers of 28 languages) or smaller-scale societies with limited access to global media (n = 116 native speakers of three non-English languages). Participants listened to songs randomly selected from a representative sample of human vocal music, originally used in four behavioral contexts, and rated the degree to which they believed the song was used for each context. Listeners in both industrialized and smaller-scale societies inferred the contexts of dance songs, lullabies, and healing songs, but not love songs. Within and across 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 behavioral contexts of three common forms of music are mutually intelligible cross-culturally and imply that musical diversity, shaped by cultural evolution, is nonetheless grounded in some universal perceptual phenomena.

music | cross-cultural | universality | cultural evolution | form and function

Significance

Music is thought to exist in every human culture, but it varies widely worldwide and appears in highly diverse contexts: from intricate ceremonies with coordinated group dances to informal, private lullabies. How do humans make sense of the music we hear? We test the hypothesis that universal acoustic properties of music make it intelligible across cultures. A globally diverse group of listeners, including native speakers of many non-English languages, heard songs recorded in many different societies. They reported their intuitions about the original behavioral context of each song. Their ratings were largely accurate, consistent with one another, and not explained by their linguistic or geographic proximity to the singer—showing that musical diversity is underlain by universal psychological phenomena.

Like many other animals, humans use vocalizations to convey their intentions and affective states (1, 2). Such vocalizations would be meaningless if members of one's own species, or members of other species, could not interpret them in a useful way. Indeed, many animal and human vocalizations are not arbitrary but instead display systematic relationships between their acoustic form and their behavioral 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 appropriate for the behavioral function of grabbing attention.

Such form–function relationships in human vocalizations 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 vocalizations even appear to be preserved across species: For instance, humans can infer the behavioral context and affect of chimpanzee vocalizations (12), and deer mothers are sensitive to the distress calls of a variety of mammals (13).

Systematic form–function relationships also occur in more complex vocalizations. Vocal music (hereafter, song) is a human universal characterized by rich variability within and across cultures (14–16). Some of the behavioral 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 behavioral contexts tend to have objective acoustic correlates; that is, they tend to display stereotyped musical features associated with their specific behavioral context. For example, dance songs tend to share clearly accented and predictable beat structures.

As with other types of vocalizations, form–function patterns in human songs may originate from our evolved psychology, perceptual biases, or unique social environment (23–26). These constraints on cultural–evolutionary processes result in musical behaviors 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 behavioral contexts of unfamiliar foreign music (14, 19), including young children, who have less musical experience relative to adults (21).

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: lidya.yurdum@yale.edu or sam@auckland.ac.nz.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218593120/-/DCSupplemental.

Published September 7, 2023.

While prior experiments have shown that people can infer the behavioral contexts of songs from different cultures using only acoustic features of the songs, 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)—an important problem affecting 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 behavior, and how much is a product of (Western) enculturation, education, and exposure to world music through globalized media.

Here, we test the prediction that the behavioral 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 industrialized 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 to 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 industrialized and smaller-scale societies would correctly infer the behavioral contexts of three types of unfamiliar songs (dance, lullaby, healing), reflecting sensitivity to acoustic and musical 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 behavioral 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).

Materials and Methods

Participants.

Industrialized 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 particip Dants 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, refs. 14 and 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 visualized in Fig. 1.

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 Appendix*, SI Text 1.1.

We aimed to maximize data quality with eight planned exclusion criteria: We excluded participants who i) performed poorly on a headphone detection task (32); 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 should 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 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. The study protocol was approved by the Harvard University Committee on the Use of Human Subjects (protocol IRB16-1080).

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

Materials. All data, protocols, code, and materials are publicly available at ref. (33) (see *Data, Materials, and Software Availability* for details).

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 (34, 35), over 75 languages, and a range of subsistence methods. The songs were originally used in four behavioral contexts: soothing a baby, dancing, expressing love, and healing the sick.

Three characteristics of the *Discography* help to minimize bias in categorizing the behavioral context of each song: i) Predetermined definitions of songs were used for categorization decisions (see table S21 in ref. 14); ii) in most cases (101 of 118 songs), behavioral contexts were determined by a consensus evaluation of substantive information found in searches of candidate recordings' liner notes and supporting ethnographic texts; and iii) the songs were categorized by researchers who had not yet listened to the songs, ensuring that their opinions concerning the sounds present on a given recording could not influence categorization decisions.

The excerpts were randomly selected 14-s segments of each song that contained singing (i.e., not instrumental-only sections), used in prior work (19). Readers can listen to all 118 song samples in the *Discography* and visually explore their acoustic and musical features at https://www.themusiclab.org/vocal-interpretations. The excerpts can be downloaded from ref. (36).

Procedure. For each trial of the listening task, participants first heard a 14-s 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 behavioral

*The Nyangatom communities had little exposure to TV, radio, and the Internet 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 5 y, 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.

Table 1. Linguistic and geographic information about the participants in the web-experiment

Language Family	Language	Total *n*	Subregion	Country	Country-wise *n*
Afro-Asiatic	Amharic	33	Eastern Africa	Ethiopia	33
	Arabic	534	Northern 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

Table 1. (Continued)

Language Family	Language	Total *n*	Subregion	Country	Country-wise *n*
	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

The "Language" column denotes the native language spoken by the participant (and the language they completed the experiment in); the "Total *n*" column denotes the number of participants recruited in that language; the "Language Family" column denotes the language family each language is part of, following the Glottolog system (31). Glottolog is a comprehensive catalog of the world's languages and their genealogy and can be accessed at https://glottolog.org. Within each language, participants were recruited from multiple countries, as noted in the "Country" column. For the cultural proximity analyses, participants were grouped into geographic subregions based on their reported location, following the typology used by the Human Relations Area Files. The "Country-wise *n*" column indicates the number of participants per language in each country.

contexts represented in the corpus, i.e., "for dancing," "to soothe a baby," "to heal illness," and "to express love for another person." The text was always presented in the participant's native language (see *Translations*, below).

We note that this procedure contrasts with that of our previous work in refs. 14 and 21, which include citizen-science listener experiments using forced-choice paradigms, and aligns with other studies from our lab using ordinal ratings of perceived behavioral contexts (19). Forced-choice paradigms have been criticized for biasing participants' responses toward the available options, resulting in false positives (37, 38). Here, we opted against a forced-choice paradigm to avoid leading listeners to artificially categorize songs into one and only one behavioral context, as songs can obviously be used for multiple, overlapping behavioral contexts in many societies. Using rating scales instead enabled us to identify one or more behavioral contexts that participants found appropriate for each song, along with those that they found inappropriate. We also asked participants to rate each song on two additional context dimensions, that were not represented by any songs in the corpus, as distractors ("to greet visitors" and "to praise a person's achievements").

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

In the industrialized 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 (*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 ruggedized laptop (*SI Appendix*, Fig. 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 labeled with a sequence of circles in ascending size, to help participants remember the direction of the scale. Participants were first familiarized with the box, identifying the three buttons corresponding to the possible responses. At the end of the experiment, participants were asked to reidentify 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 (by M.S. and L.G.), we simplified the task used in the smaller-scale societies by reducing the number of response options in each behavioral context dimension from 4 points to 3 points, and rephrased

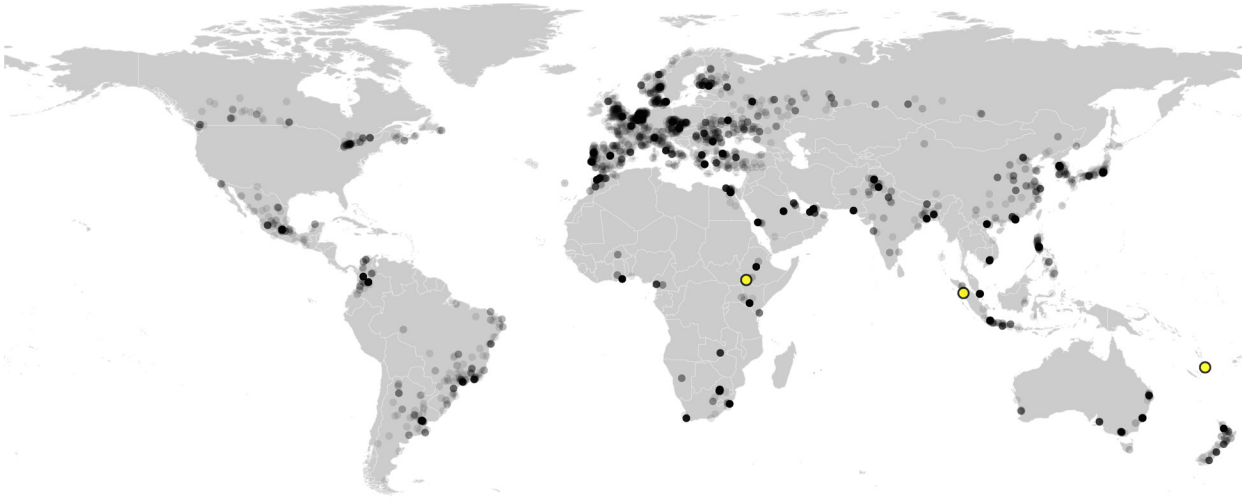

Fig. 1. Geographic distribution of participants. We recruited participants in industrialized societies and in three smaller-scale societies. The gray dots indicate the approximate locations of the participants in industrialized societies, as measured via IP geolocation. The yellow dots indicate the approximate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

Table 2. Information about the three smaller-scale societies

Region	Society	Language	Language Family	Subsistence type	Approx. Community Size	Distance to City (km)	Final *n*
Eastern Africa	Nyangatom	Nyangatom	Nilotic	Pastoralist	155	180	34
Southeast Asia	Mentawai Islanders	Mentawai	Austronesian	Horticulturalist	260	120	27
Melanesia	Tannese Ni-Vanuatu	Bislama	Indo-European Creole	Horticulturalist	6,000	224	55

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 ref. 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 organizations 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 standardized 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 reevaluate 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 a 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 backtranslated 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 behavioral context dimension. These scores reflected, on average, how likely the participants thought it was that each song was used in each of the six behavioral contexts. These song-wise averages were then *z*-scored.

Because each participant heard only a randomly selected subset of the corpus, the number of ratings averaged for each song in each cohort varied (industrialized societies: median = 1,094 ratings, range 917 to 1,183; smaller-scale societies: median = 18, range 8 to 28).

Three Forms of Song Are Mutually Intelligible. First, we asked whether listeners could accurately infer the behavioral contexts of the songs, using the same analysis strategy as in ref. 19, which included similar data types: We tested whether each behavioral 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 behavioral contexts. This approach measures whether songs originally used in a given behavioral context were perceived to be *more* appropriate for that context than the average song in the corpus. For an alternative analysis approach using mixed models in the industrialized societies, see *SI Appendix*, SI Text 1.2.

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

Response patterns across behavioral contexts were informative in both positive and negative directions. For example, the industrialized cohort rated dance songs 0.90 SDs above the base rate on the "…for dancing" dimension (β = 0.90, SE = 0.145, P < 0.0001) but rated lullabies 0.83 SDs *below* the base rate (β = −0.83, SE = 0.145, P < 0.0001). This suggests that listeners inferred that completely unfamiliar dance songs were suitable for dancing but also that lullabies were not. The reverse pattern was evident for the "…to soothe a baby" dimension, with lullabies rated 1.09 SDs above the base rate (β = 1.09, SE = 0.139, P < 0.0001) and dance songs well below the base rate (β = −0.62, SE = 0.139, P < 0.0001).

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 = 0.66, SE = 0.162, P < 0.0001), with lullabies below it (β = −0.68, SE = 0.162, P < 0.0001); and lullabies were rated 0.75 SD above the base rate of "…to soothe a baby" (β = 0.75, SE = 0.161, P < 0.0001). Moreover, both cohorts rated dance songs higher on the "…for dancing" dimension than each of the other three dimensions and likewise rated lullabies higher on the "…to soothe a baby" dimension than the other three dimensions (all Ps < 0.05).

Effects in healing songs were smaller in both cohorts but still indicated reliable inferences, with ratings on "…to heal illness" above the base rate in both industrialized societies (β = 0.49, SD = 0.18, P = 0.007) and smaller-scale societies (β = 0.47, SD = 0.18, P = 0.01). Healing songs scored higher on the "…to heal illness" dimension than the "…for dancing" dimension in both cohorts and also higher than the "…to soothe a baby" dimension in the smaller-scale cohort (all Ps < 0.05). Consistent with ref. 19, neither of the cohorts' ratings of love songs on "…to express love for another person" was higher than the base rate, suggesting an inability to accurately identify this behavioral context.[†] (Industrialized societies: β = 0.30, SD = 0.18, P = 0.1; Smaller-scale societies: β = 0.15, SD = 0.18, P = 0.41). The industrialized cohort did, however, rank love songs higher on the "…to express love for another person" dimension than "…to heal illness" (P = 0.02).

[†]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.

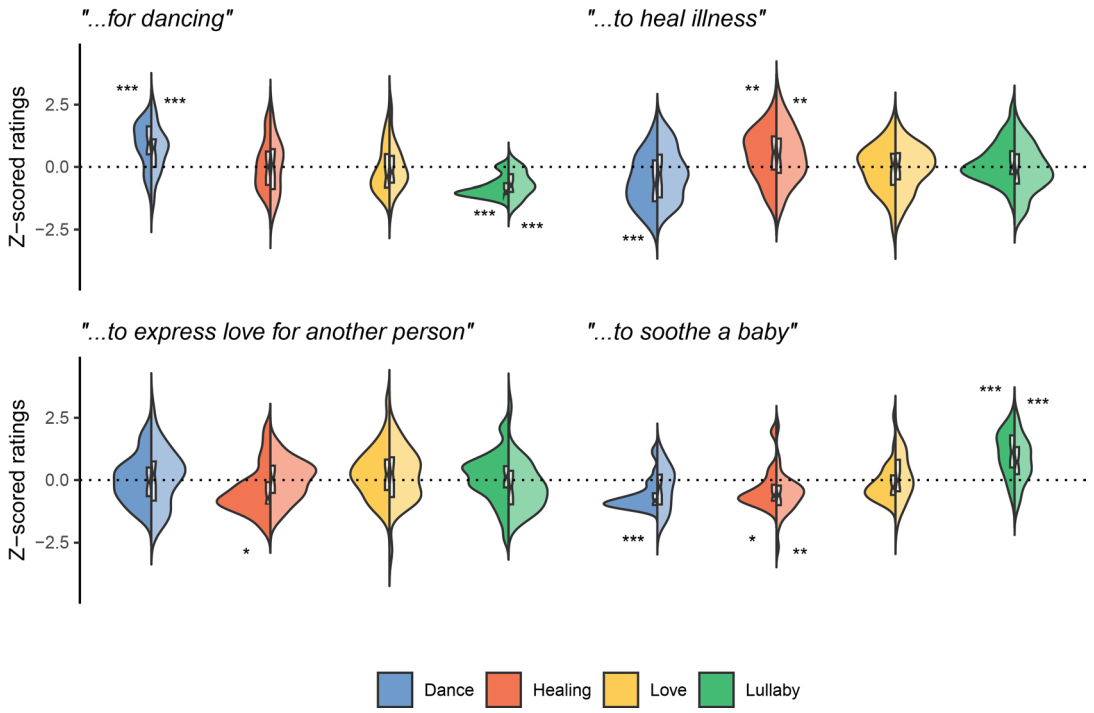

Fig. 2. The behavioral contexts of songs found worldwide are detectable by listeners recruited worldwide. Listeners heard a random selection of songs originally produced in one of four behavioral contexts: songs that were used "for dancing," "to heal illness," "to express love for another person," or "to soothe a baby." For each song, they were unaware of the culture or the behavioral context in which it was recorded. Each of the four plots visualizes the distributions of mean song-wise ratings for a particular behavioral context dimension (e.g., "for dancing"). The paired half-violins in each plot correspond to the four behavioral contexts, i.e., the actual behavioral contexts in which the songs originally appeared, denoted by color. The half-violins depict the distributions of mean song-wise ratings from each of the two cohorts of participants (i.e., from industrialized societies on the left, or smaller-scale societies on the right). 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 behavioral context. For dance songs, lullabies, and healing songs, the ratings of listeners in both types of societies accurately reflected the original behavioral context of the songs (e.g., dance songs, but not the other three behavioral 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 shaded area in the half-violins represent kernel density estimates; the vertical boxplots denote the median (horizontal line), 95% CI (notches), and interquartile range (edges of the boxes), all computed cohort- and song-wise within each plot. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

On an anonymous reviewer's suggestion, we also tested whether the four behavioral context dimensions that we asked listeners to rate the songs on have distinct latent underpinnings or whether they could be summarized into a smaller number of factors. We conducted a principal components analysis on listeners' ratings, separately for the industrialized and smaller-scale society cohorts. In both cohorts, dance songs and lullabies were clearly differentiated by the first component, which loaded positively on the "…for dancing" dimension and negatively on the "…to soothe a baby" dimension, suggesting that the "dancing" and "soothing a baby" dimensions were both tapping into a latent behavioral context that might be described as "high vs. low arousal contexts." This first component explained the majority of variance in responses in both cohorts. Indeed, dance songs and lullabies emerge as the most clearly differentiated pairing in all studies of the *Natural History of Song Discography*, distinguished by musical features such as melodic complexity, rhythmic complexity, tempo, arousal, and accent structure (14, 19, 21). Full statistical reporting of the principal components analysis is in *SI Appendix*, SI Text 1.3, and results are visualized in *SI Appendix*, Fig. S2.

In sum, these findings indicate that the behavioral contexts of dance songs, lullabies, and healing songs recorded worldwide are intelligible to listeners in both industrialized 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 industrialized cohort to listeners in the smaller-scale society cohort. Second, we measured the variation in listener responses across linguistic subgroups of the industrialized cohort.

Comparison of Listeners across Industrialized and Smaller-Scale Societies. As a general test of cross-cohort similarity, we computed Pearson correlations of the song-wise mean ratings on each dimension from each cohort. The four correlations were positive and statistically significant (Fig. 3*A*), but varied in magnitude, with the highest correlations in "…for dancing" ($r = 0.84$) and "…to soothe a baby" ($r = 0.59$). The correlations in the contexts of healing and expressing love were also statistically significant, but were lower; note that the data in these two dimensions in the smaller-scale societies are relatively noisy (see *SI Appendix*, SI Text 1.4 for an analysis of noise ceilings). Readers can also explore the results in Fig. 3*A* in an interactive audio version of the plot, at https://www.themusiclab.org/vocal-interpretations.

We then repeated this analysis with an alternate approach, using stratified bootstrapping to estimate the variability in each correlation, given the much larger heterogeneity of the industrialized cohort (*SI Appendix*, Fig. S3). The findings repeated, with modestly attenuated effect sizes. Such reduced effect sizes are to be expected given the increased sampling error due to sampling a smaller number of observations per song.

These correlations likely underestimate the true effect sizes, moreover, for two reasons. First, there were substantive task differences between the two cohorts: The smaller-scale society cohort

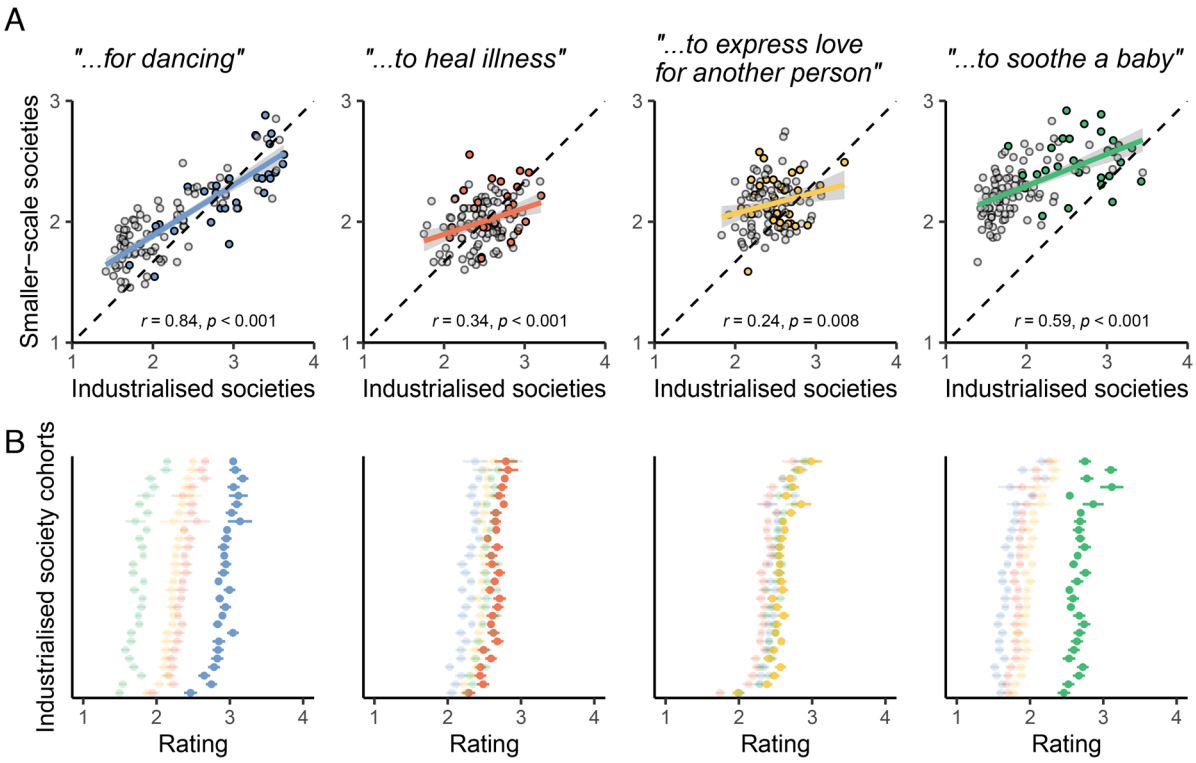

Fig. 3. Consistency of listeners' intuitions across cohorts and across languages. (*A*) The mean song-wise ratings of listeners in the industrialized and smaller-scale societies, across the full corpus of songs, correlated with one another, on each of the four dimensions of interest. In the scatterplots, each point denotes a song-wise mean plotted in terms of its rating by participants in the industrialized societies (*x* axis) and participants in the smaller-scale societies (*y* axis). The highlighted dots denote songs whose behavioral context corresponds with the dimension of that plot (e.g., the blue points in the left-most "… for dancing" plot 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 *Materials and Methods*. (*B*) Within each linguistic subgroup of the industrialized societies, the main effects repeated consistently. The forest plots show the mean ratings of songs originally used in each of the four behavioral contexts, on each of the dimensions (one per plot), within each of the 28 linguistic subgroups (i.e., each row of points summarizes 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 colors of the points correspond to the behavioral contexts, using the same scale as Fig. 2 (dance songs in blue, healing songs in red, love songs in yellow, and lullabies in green).

used a 3-point rating scale (with modified wording) instead of the industrialized cohort's 4-point scale, and the smaller-scale society listeners provided ratings on four "distractor" behavioral contexts that were not represented in the corpus, in contrast to the industrialized cohort who rated only two. Such task differences should, in principle, only reduce the measurable correlations across cohorts.

Second, in the smaller-scale societies, each song excerpt was rated by far fewer listeners than in the industrialized societies. This difference produced a significantly higher SEM for a given song in the smaller-scale society cohort, relative to the industrialized cohort (e.g., mean SE for lullabies: 0.20 in smaller-scale societies vs. 0.03 in industrialized societies). This limited the explainable variance in the smaller-scale society data and is likely to bias the cross-cohort correlations downward; we attempted to compensate for this bias with noise ceiling metrics (*SI Appendix*, SI Text 1.4 and ref. 39).

As a more conservative test of the differences between the intuitions of listeners in the two cohorts, we compared the *z*-scored ratings of the industrialized cohort for each behavioral 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 Fig. 2: 4 behavioral contexts × 4 dimensions). None of the 16 comparisons were statistically significant; the largest cohort-wise difference had $P = 0.09$, above the conventional alpha of 0.05 and well above a more conservative Bonferroni-adjusted alpha for 16 comparisons of 0.003.

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

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

First, we reran 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 (Fig. 3*B*). In 27 of the 28 linguistic subgroups, dance songs were rated significantly above the base rate of "…for dancing" (Ps < 0.001); 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 < 0.0001). All 28 linguistic subgroups rated lullabies above the base rate of "… to soothe a baby" (Ps < 0.0001).

As in the main effects, results in healing songs were somewhat weaker, with healing songs identified as most appropriate in the context of "…to heal illness" by 20 of the 28 subgroups (Ps < 0.05). Only 12 subgroups rated love songs significantly higher (Ps < 0.05) than the base rate of "…to express love for another person" across all songs.

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 industrialized and smaller-scale societies and within industrialized societies) were similar to one another.

Cultural Proximity Is Relatively Uninformative to Listeners. Having found 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 behavioral 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 behavioral context in a song should more closely match that song's actual behavioral context.

To operationalize 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 behaviors, such as linguistic features (40), vocalization styles (20), or camel-herding practices (41). Research on the universality of non-verbal expressions of emotion, for instance, has found that cross-cultural emotion recognition is higher when the judge's native language is closer to that of the poser (42).

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, 43). We used Glottolog (31) to classify local languages into language families and the Human Relations Area Files (http://ehrafworldcultures.yale.edu) World Subregion 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, for a participant recruited in 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 (not a Turkic 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 (both Greece and Turkey are in Southeastern Europe). Linguistic and geographic markers of proximity can overlap, but not necessarily.

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 ($\beta = 0.05$, SE = 0.022, $P = 0.03$), lullaby ($\beta = 0.06$, SE = 0.028, $P = 0.03$), and love songs ($\beta = 0.05$, SE = 0.024, $P = 0.04$), but not healing songs ($\beta = 0.02$, SE = 0.032, $P = 0.5$; Fig. 4).

These effects were very small, however: The largest, found for lullabies, showed that sharing a language family resulted in an estimated boost to the "…to soothe a baby" dimension 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 therefore minimal compared to the variance explained by the actual behavioral context and universal regularities in the songs' musical features.

Results were comparable for geographic proximity, with marginally larger effects for dance ($\beta = 0.16$, SE = 0.036, $P < 0.0001$), lullaby ($\beta = 0.14$, SE = 0.039, $P < 0.001$), and love songs ($\beta = 0.07$, SE = 0.035, $P = 0.04$), and no significant effect for healing songs ($\beta = 0.04$, SE = 0.040, $P = 0.27$). Here, the largest effect was found for sharing a geographical area when rating a dance song on the "…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 behavioral contexts in isolation, we regressed the listeners' ratings (from the dimension corresponding to that behavioral 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 behavioral 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 *SI Appendix*, Table S1).

Two proxies for cultural proximity therefore explained a small proportion of the variance in listener responses relative to the variance explained by the actual behavioral context of the song. This suggests that listeners were primarily relying on universal regularities in the songs' musical features to inform their inferences.

Such an interpretation is bolstered by previous work showing the consistency and distinctiveness with which musical features characterize dance songs, healing songs and lullabies worldwide, and how perceptual judgments reflect those features. For example, acoustic regularities underlying the songs used in particular behavioral contexts are robust enough to enable machine classification of behavioral contexts, on the basis of only musical features, at a high level of accuracy in held-out data (14). The acoustic regularities are also robust enough to enable reliable classification of songs by children (21), whose inferences are informed by similar musical features to adults' inferences. Further, subjective ratings of musical (e.g., perceived tempo) and contextual features (e.g., perceived number of singers) by nonmusicians differentiate the four song types. This leads listeners to make nonrandom errors on the basis of similar musical features that span behavioral contexts; for example, when a nonlullaby shares musical features with a prototypical lullaby, it is more likely to be rated highly as "…to soothe a baby" (19).

As a final exploratory analysis, we asked whether the musical features studied in these prior analyses similarly predicted listeners' ratings here and, if so, whether these musical features were in line

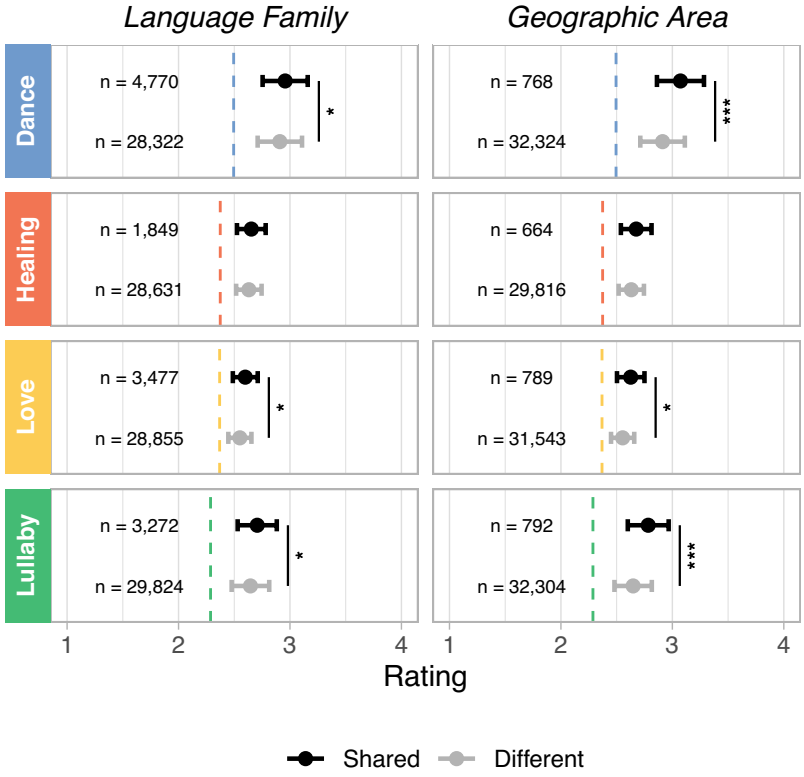

Fig. 4. Increased linguistic or geographic proximity between listeners and singers does not substantially improve performance. Because both songs and listeners came from global samples, in some cases, the culture of the listener is more related to the culture of the singer than others. This could, in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

with the features previously found to characterize each song type. Indeed, many of the features previously found to characterize dance songs, healing songs, and lullabies (such as tempo, accent structure, and a steady beat) also predicted listeners' ratings on the respective behavioral context dimensions. The full results are reported in *SI Appendix*, SI Text 1.5.

Discussion

In a global sample of people residing in both industrialized and smaller-scale societies and tested predominantly in non-English languages, we found that listeners' inferences about the behavioral contexts of unfamiliar, foreign songs are accurate, similar to one another, and relatively uninfluenced by cultural proximity. Moreover, many of the acoustic and musical features universally associated with the types of songs we studied (14) also predicted listeners' inferences about those very songs. Some basic aspects of musical interpretation therefore appear to be universal and grounded in globally shared perceptual principles.

These findings generalize prior findings reporting the ability of English-speaking participants recruited online to reliably infer the behavioral 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). The alternative use of many 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 industrialized 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, justifying claims about *human* psychology, as opposed to the psychology of a nonrepresentative subset of humans.

Moreover, the diversity and geographic breadth of stimuli used here help to ensure the generalizability of the results (44). That being said, while precautions were taken to avoid bias when constructing the stimulus set used here, we cannot fully rule out ethnographer bias in the selection and classification of songs therein. For example, it is possible that ethnographers tend to describe songs with functions that they themselves easily identify, or that they tend to record songs that resemble the songs of industrialized societies. This is a limitation of the study, as we cannot study what ethnographers did not record. This issue speaks to the essential nature of work preserving the cultural record of human history. Unfortunately, just as linguistic cultures can become endangered or lost (45), so too can musical cultures.

A second limitation of our study is that the use of rating scales may have biased participants toward evaluating each song within a cultural framework that they may not share and could have primed participants to interpret stimuli within given constraints (46, 47). As an alternative, the collection of free-response data (i.e., asking participants to generate their own list of behavioral contexts for a song, rather than choosing from prespecified options) would enable

participants to express the full range of culture-specific interpretations of song. Indeed, related research on cross-cultural emotion recognition has shown that respondents identify more emotions in expressions in free-response paradigms (48, 49); that the presence of an emotion word can influence whether that emotion is perceived in a face (50); and that diverging from a typical forced-choice paradigm can reduce the recognition of emotion categories (51). In the present study, we opted to use a rating scale paradigm to avoid issues associated with forced-choice paradigms that present behavioral contexts as mutually exclusive categories (see, e.g., ref. 52), but repeating the paradigm with a free-text approach would be a productive direction for future work.

What is the source of the cross-culturally robust musical inferences shown here? We consider the fact that effects were strongest for the contexts of dance and infant care to support theories that music evolved as a vocal signal in these specific contexts (23, 24, 26). Music appears to function as a credible signal in a similar fashion to the vocalizations produced and detected within and across many species (25).

The possibility of evolved 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., refs. 9 and 42), including across species (12, 53, 54), facial expressions (e.g., ref. 55), and nonreferential information in music (56–59). 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 robust form–function relations in speech worldwide (20, 60–63).

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, ref. 19, and small, but positive effects in a forced-choice task, ref. 14). Here, using a rating scale paradigm, we do not find a significant effect for love songs, suggesting that the effect in ref. 14 may have been a product of the forced-choice paradigm. These results further suggest that love songs are a fuzzy category of music when produced in an unfamiliar language. 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.

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' vocalizations, music operates in a fashion similar to other communicative domains. That culture does not appear to explain much variation at the level of language families or geographic subregions suggests that the patterns we see globally are likely cognitive universals rather than deeply inherited cultural traits. Nevertheless, significant cultural variability does exist among cultures that share the same language family or geographic subregion, and intuitively, culture must matter to some degree; at the extreme, we expect a Hadza tribesman to do a better job of categorizing Hadza songs than a non-Hadza. In other words, the proxies that we used here for cultural proximity may be too broad to capture shared cultural effects. It might be that relevant cultural information dissipates within a couple of centuries of independent evolution, such that being in the same language family (which are often thousands of years old) does not mean much. An interesting question is precisely how this effect of cultural knowledge tails off: Does it persist across cultures that separated centuries ago, or does it rely on idiosyncrasies that are quickly lost, such that one needs to be from the same culture as a song to see any marked improvement in categorization ability? 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. Cross-cultural experiments, perhaps relying on music with obscured or masked lyrics (because linguistic content is a strong cue to behavioral context in music), may further explore the roles that culture plays in shaping music perception.

Another interesting question raised by our findings is whether certain song types could be swapped across cultures and still be functionally effective. We believe that the answer to this question is "yes," with an important caveat. The present research does not address many of the culture-specific effects of music that are arguably the most interesting: the simplistic "What's this song for"-style paradigm used here with four basic behavioral contexts cannot, of course, capture the myriad creative and functional ways that music is used around the world. It would strain credulity, for instance, to expect that naive listeners could identify the "navigation songs" of some indigenous Australian societies (64).

However, certain more basic types of songs occur universally across human societies, these songs share characteristic musical features, and their features allow the songs to be mutually intelligible across cultural boundaries. In at least one case, lullabies, songs are demonstrably effective in a culture-independent fashion: When Western babies listened to the lullabies from the same corpus studied in this paper, they showed behavioral and physiological signs of relaxation (65). Similarly, we suspect that readers of this manuscript might be moved to dance by the dance songs studied here. How music does and does not transcend languages and cultures is a promising topic for future work.

Data, Materials, and Software Availability. A reproducible R Markdown manuscript is available at ref. (33), with all associated data and materials. The same repository includes code for running the listener task in Qualtrics (for the industrialized societies) and E-Prime (for the smaller-scale societies), including translations of all experiments. The excerpted audio corpus (the *Natural History of Song Discography*) is available at ref. (36).

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 NIH 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 the members of The Music Lab for feedback on the paper.

Author affiliations: [a]Child Study Center, Yale University, New Haven, CT 06520; [b]Department of Psychology, University of Amsterdam, Amsterdam 1018WT, Netherlands; [c]Department of Anthropology, University of California, Davis, Davis CA 95616; [d]Department of Anthropology, Boston University, Boston, MA 02215; [e]School of Psychology, University of Auckland, Auckland 1010, New Zealand; and [f]Division of Continuing Education, Harvard University, Cambridge, MA 02138

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.

1. E. S. Morton, On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* **111**, 855–869 (1977).
2. K. Pisanski, G. A. Bryant, C. Cornec, A. Anikin, D. Reby, Form follows function in human nonverbal vocalisations. *Ethol. Ecol. Evol.* **34**, 303–321 (2022).
3. J. A. Endler, Some general comments on the evolution and design of animal communication systems. *Philos. Trans. R. Soc. B Biol. Sci.* **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. *Anim. Behav.* **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. *R Soc. Open Sci.* **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. *Curr. Biol.* **25**, 2051–2056 (2015).
7. G. A. Bryant, H. C. Barrett, Recognizing intentions in infant-directed speech: Evidence for universals. *Psychol. Sci.* **18**, 746–751 (2007).
8. H. C. Barrett, G. Bryant, Vocal emotion recognition across disparate cultures. *J. Cogn. 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 Rev.* **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. *Proc. R. Soc. London B Biol. Sci.* **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. *Proc. R. Soc. B* **287**, 20201148 (2020).
13. S. Lingle, T. Riede, Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *Am. Nat.* **184**, 510–522 (2014).
14. S. A. Mehr *et al.*, Universality and diversity in human song. *Science* **366**, 957–970 (2019).
15. B. Nettl, On the question of universals. *The world of music* **19**, 2–7 (1977).
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 Behav. Dev.* **16**, 193–211 (1993).
18. S. E. Trehub, A. M. Unyk, L. J. Trainor, Maternal singing in cross-cultural perspective. *Infant Behav. Dev.* **16**, 285–295 (1993).
19. S. A. Mehr, M. Singh, H. York, L. Glowacki, M. M. Krasnow, Form and function in human song. *Curr. Biol.* **28**, 356–368 (2018).
20. C. B. Hilton *et al.*, Acoustic regularities in infant-directed speech and song across cultures. *Nat. Hum. Behav.* **6**, 1545–1556 (2022), 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. *J. Exp. Psychol. Gen.* (2023), 10.1037/xge0001289.
22. M. Singh, S. A. Mehr, Universality, domain-specificity and development of psychological responses to music. *Nat. Rev. Psychol.* **2**, 333–346 (2023).
23. E. H. Hagen, G. A. Bryant, Music and dance as a coalition signaling system. *Hum. Nat.* **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. M. Krasnow, G. A. Bryant, E. H. Hagen, Origins of music in credible signaling. *Behav. Brain Sci.* **44**, e60 (2021).
26. S. A. Mehr, E. S. Spelke, Shared musical knowledge in 11-month-old infants. *Dev. Sci.* **21** e21542 (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 Cogn. Sci.* **8**, 40–46 (2004).
29. D. E. Blasi, J. Henrich, E. Adamou, D. Kemmerer, A. Majid, Over-reliance on English hinders cognitive science. *Trends Cogn. Sci.* **26**, 1153–1170 (2022).
30. H. A. Elfenbein, N. Ambady, On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.* **128**, 203 (2002).
31. H. Hammarström, R. Forkel, M. Haspelmath, *Glottolog 4.0* (Max Plank Institute for the Science of Human History, 2019).
32. K. J. P. Woods, M. H. Siegel, J. Traer, J. H. McDermott, Headphone screening to facilitate web-based auditory experiments. *Atten. Percept. Psychophys.* **79**, 2064–2072 (2017).
33. L.Y. Yurdum, S.A. Mehr, Universal interpretations of vocal music. GitHub. https://github.com/themusiclab/universal-music. Accessed 14 August 2023.
34. G. P. Murdock *et al.*, *Outline of Cultural Materials* (Human Relations Area Files Inc., 2008).
35. R. Naroll, The proposed HRAF probability sample. *Behav. Sci. Notes* **2**, 70–80 (1967).
36. S.A. Mehr, Cross-cultural music corpus: The Natural History of Song Discography (randomized 14s excerpts). Zenodo. https://zenodo.org/record/7265514. Accessed 14 August 2023.
37. N. Betz, K. Hoemann, L. F. Barrett, Words are a context for mental inference. *Emotion* **19**, 1463–1477 (2019).
38. M. G. Frank, J. Stennett, The forced-choice paradigm and the perception of facial expressions of emotion. *J. Pers. Soc. Psychol.* **80**, 75 (2001).
39. A. S. Cowen, D. Keltner, Universal facial expressions uncovered in art of the ancient Americas: A computational approach. *Sci. Adv.* **6**, eabb1005 (2020).
40. 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).
41. R. Mace *et al.*, The comparative method in anthropology [and comments and reply]. *Curr. Anthropol.* **35**, 549–564 (1994).
42. K. R. Scherer, R. Banse, H. G. Wallbott, Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross Cul. Psychol.* **32**, 76–92 (2001).
43. A. Wood, M. Rychlowska, P. M. Niedenthal, Heterogeneity of long-history migration predicts emotion recognition accuracy. *Emotion* **16**, 413 (2016).
44. T. Yarkoni, The generalizability crisis. *Behav. Brain Sci.* **45**, e1 (2022).
45. H. Skirgård *et al.*, Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Sci. Adv.* **9**, eadg6175 (2023).
46. J. A. Russell, Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* **115**, 102 (1994).
47. J. A. Russell, Forced-choice response format in the study of facial expression. *Motiv. Emot.* **17**, 41–51 (1993).
48. J. Haidt, D. Keltner, Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cogn. Emot.* **13**, 225–266 (1999).
49. M. Gendron, D. Roberson, J. M. van der Vyver, L. F. Barrett, Cultural relativity in perceiving emotion from vocalizations. *Psychol. Sci.* **25**, 911–920 (2014).
50. M. Gendron, K. A. Lindquist, L. Barsalou, L. F. Barrett, Emotion words shape emotion percepts. *Emotion* **12**, 314 (2012).
51. C. Crivelli, S. Jarillo, J. A. Russell, J.-M. Fernández-Dols, Reading emotions from faces in two indigenous societies. *J. Exp. Psychol. General* **145**, 830 (2016).
52. M. Gendron, C. Crivelli, L. F. Barrett, Universality reconsidered: Diversity in making meaning of facial expressions. *Curr. Dir. Psychol. Sci.* **27**, 211–219 (2018).
53. T. Faragó *et al.*, Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biol. Lett.* **10**, 20130926 (2014).
54. P. Filippi *et al.*, Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proc. R Soc. B: Biol. Sci.* **284**, 20170990 (2017).
55. A. S. Cowen *et al.*, Sixteen facial expressions occur in similar contexts worldwide. *Nature* **589**, 251–257 (2021).
56. 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).
57. 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. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1924–1934 (2020). 10.1073/pnas.1910704117.
58. T. Fritz *et al.*, Universal recognition of three basic emotions in music. *Curr. Biol.* **19**, 573–576 (2009).
59. B. Sievers, L. Polansky, M. Casey, T. Wheatley, Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 70–75 (2013).
60. D. M. Sidhu, P. M. Pexman, Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Lang. Cogn. Neurosci.* **33**, 25–31 (2018).
61. M. Imai, S. Kita, The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. R Soc. B: Biol. Sci.* **369**, 20130298 (2014).
62. D. E. Blasi, S. Wichmann, H. Hammarström, P. F. Stadler, M. H. Christiansen, Sound–meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10818–10823 (2016).
63. A. Ćwiek *et al.*, Novel vocalizations are understood across cultures. *Sci. Rep.* **11**, 1–12 (2021).
64. R. P. Norris, B. Y. Harney, Songlines and navigation in Wardaman and other Australian aboriginal cultures. *J. Astron. Hist. Herit.* **17**, 15 (2014).
65. C. M. Bainbridge *et al.*, Infants relax in response to unfamiliar foreign lullabies. *Nat. Hum. Behav.* **5**, 256–264 (2021), 10.1038/s41562-020-00963-z.