RESEARCH ARTICLE SUMMARY

PSYCHOLOGY OF MUSIC

Universality and diversity in human song

Samuel A. Mehr*, Manvir Singh*, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, S. Atwood, Christopher Lucas, Nori Jacoby, Alena A. Egner, Erin J. Hopkins, Rhea M. Howard, Joshua K. Hartshorne, Mariela V. Jennings, Jan Simson, Constance M. Bainbridge, Steven Pinker, Timothy J. O'Donnell, Max M. Krasnow, Luke Glowacki*

INTRODUCTION: Music is often assumed to be a human universal, emerging from an evolutionary adaptation specific to music and/or a by-product of adaptations for affect, language, motor control, and auditory perception. But universality has never actually been systematically demonstrated, and it is challenged by the vast diversity of music across cultures. Hypotheses of the evolutionary function of music are also untestable without comprehensive and representative data on its forms and behavioral contexts across societies.

RATIONALE: We conducted a natural history of song: a systematic analysis of the features of vocal music found worldwide. It consists of a corpus of ethnographic text on musical behavior from a representative sample of mostly small-scale societies, and a discography of audio recordings of the music itself. We then applied tools of computational social science, which minimize the influence of sampling error and other biases, to answer six questions. Does music appear universally? What kinds of behavior are associated with song, and how do they vary among societies? Are the musical features of a song indicative of its behavioral context (e.g., infant care)? Do the melodic and rhythmic patterns of songs vary systematically, like those patterns found in language? And how prevalent is tonality across musical idioms?

RESULTS: Analysis of the ethnography corpus shows that music appears in every society observed; that variation in song events is well characterized by three dimensions (formality, arousal, religiosity); that musical behavior varies more within societies than across them on these dimensions; and that music is regularly associated with behavioral contexts such as infant care, healing, dance, and love. Analysis of the discography corpus shows that identifiable acoustic features of songs (accent, tempo, pitch range, etc.) predict their primary behavioral context (love, healing, etc.); that musical forms vary along two dimensions (melodic and rhythmic complexity); that melodic and rhythmic bigrams fall into power-law distributions; and that tonality is widespread, perhaps universal.

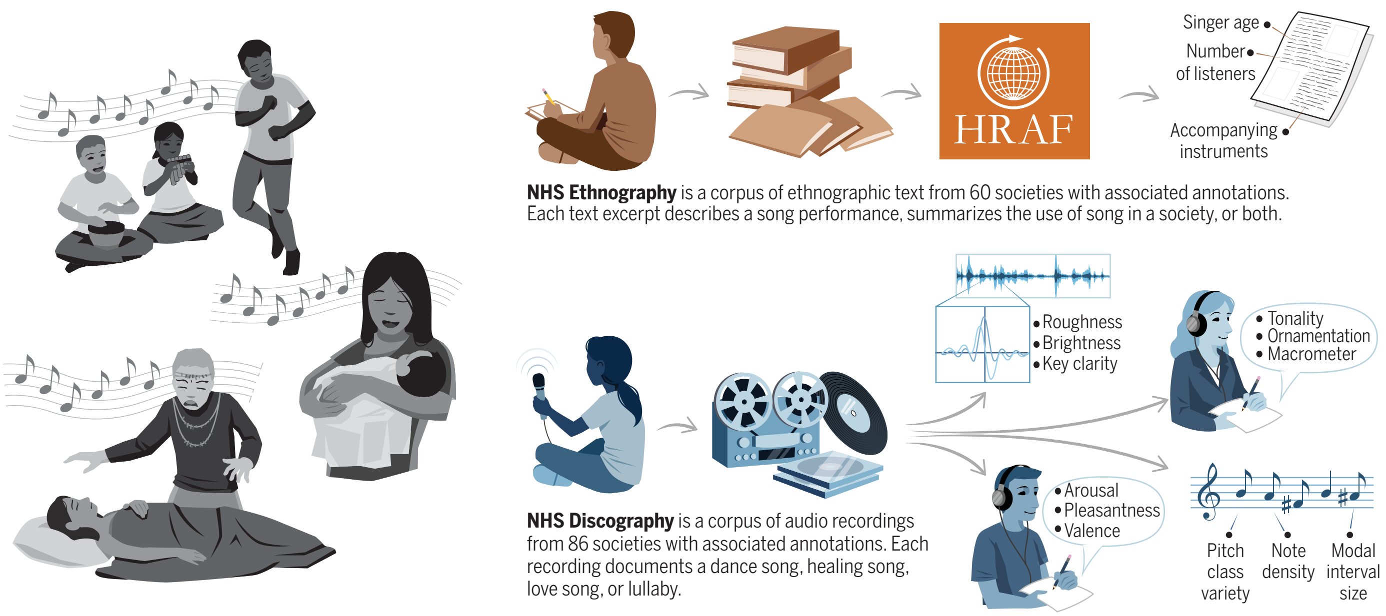
CONCLUSION: Music is in fact universal: It exists in every society (both with and without words), varies more within than between societies, regularly supports certain types of behavior, and has acoustic features that are systematically related to the goals and responses of singers and listeners. But music is not a fixed biological response with a single prototypical adaptive function: It is produced worldwide in diverse behavioral contexts that vary in formality, arousal, and religiosity. Music does appear to be tied to specific perceptual, cognitive, and affective faculties, including language (all societies put words to their songs), motor control (people in all societies dance), auditory analysis (all musical systems have signatures of tonality), and aesthetics (their melodies and rhythms are balanced between monotony and chaos). These analyses show how applying the tools of computational social science to rich bodies of humanistic data can reveal both universal features and patterns of variability in culture, addressing long-standing debates about each. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax0868

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sam@wjh.harvard.edu (S.A.M.); manvirsingh@fas.harvard.edu (M.S.); glowacki@psu.edu (L.G.)
Cite this article as S. A. Mehr *et al.*, *Science* **366**, eaax0868 (2019). DOI: 10.1126/science.aax0868

Studying world music systematically. We used primary ethnographic text and field recordings of song performances to build two richly annotated cross-cultural datasets: NHS Ethnography and NHS Discography. The original material in each dataset was annotated by humans (both amateur and expert) and by automated algorithms.



RESEARCH ARTICLE

PSYCHOLOGY OF MUSIC

Universality and diversity in human song

Samuel A. Mehr[1,2,3]*, Manvir Singh[4]*, Dean Knox[5], Daniel M. Ketter[6,7], Daniel Pickens-Jones[8], S. Atwood[2], Christopher Lucas[9], Nori Jacoby[10], Alena A. Egner[2], Erin J. Hopkins[2], Rhea M. Howard[2], Joshua K. Hartshorne[11], Mariela V. Jennings[11], Jan Simson[2,12], Constance M. Bainbridge[2], Steven Pinker[2], Timothy J. O'Donnell[13], Max M. Krasnow[2], Luke Glowacki[14]*

What is universal about music, and what varies? We built a corpus of ethnographic text on musical behavior from a representative sample of the world's societies, as well as a discography of audio recordings. The ethnographic corpus reveals that music (including songs with words) appears in every society observed; that music varies along three dimensions (formality, arousal, religiosity), more within societies than across them; and that music is associated with certain behavioral contexts such as infant care, healing, dance, and love. The discography—analyzed through machine summaries, amateur and expert listener ratings, and manual transcriptions—reveals that acoustic features of songs predict their primary behavioral context; that tonality is widespread, perhaps universal; that music varies in rhythmic and melodic complexity; and that elements of melodies and rhythms found worldwide follow power laws.

At least since Henry Wadsworth Longfellow declared in 1835 that "music is the universal language of mankind" (*1*), the conventional wisdom among many authors, scholars, and scientists is that music is a human universal, with profound similarities across societies (*2*). On this understanding, musicality is embedded in the biology of *Homo sapiens* (*3*), whether as one or more evolutionary adaptations for music (*4*, *5*), the by-products of adaptations for auditory perception, motor control, language, and affect (*6*–*9*), or some amalgam of these.

Music certainly is widespread (*10*–*12*), ancient (*13*), and appealing to almost everyone (*14*). Yet claims that it is universal or has universal features are commonly made without citation [e.g., (*15*–*17*)], and those with the greatest expertise on the topic are skeptical. With a few exceptions (*18*), most music scholars suggest that few if any universals exist in music (*19*–*23*). They point to variability in the interpretations of a given piece of music (*24*–*26*), the importance of natural and social environments in shaping music (*27*–*29*), the diverse forms of music that can share similar behavioral functions (*30*), and the methodological difficulty of comparing the music of different societies (*12*, *31*, *32*). Given these criticisms, along with a history of some scholars using comparative work to advance erroneous claims of cultural or racial superiority (*33*), the common view among music scholars today (*34*, *35*) is summarized by the ethnomusicologist George List: "The only universal aspect of music seems to be that most people make it. … I could provide pages of examples of the non-universality of music. This is hardly worth the trouble" (*36*).

Are there, in fact, meaningful universals in music? No one doubts that music varies across cultures, but diversity in behavior can shroud regularities emerging from common underlying psychological mechanisms. Beginning with Chomsky's hypothesis that the world's languages conform to an abstract Universal Grammar (*37*, *38*), many anthropologists, psychologists, and cognitive scientists have shown that behavioral patterns once considered arbitrary cultural products may exhibit deeper, abstract similarities across societies emerging from universal features of human nature. These include religion (*39*–*41*), mate preferences (*42*), kinship systems (*43*), social relationships (*44*, *45*), morality (*46*, *47*), violence and warfare (*48*–*50*), and political and economic beliefs (*51*, *52*).

Music may be another example, although it is perennially difficult to study. A recent analysis of the *Garland Encyclopedia of World Music* revealed that certain features—such as the use of words, chest voice, and an isochronous beat—appear in a majority of songs recorded within each of nine world regions (*53*). But the corpus was sampled opportunistically, which made generalizations to all of humanity impossible; the musical features were ambiguous, leading to poor interrater reliability; and the analysis studied only the forms of the societies' music, not the behavioral contexts in which it is performed, leaving open key questions about functions of music and their connection to its forms.

Music perception experiments have begun to address some of these issues. In one, internet users reliably discriminated dance songs, healing songs, and lullabies sampled from 86 mostly small-scale societies (*54*); in another, listeners from the Mafa of Cameroon rated "happy," "sad," and "fearful" examples of Western music somewhat similarly to Canadian listeners, despite having had limited exposure to Western music (*55*); in a third, Americans and Kreung listeners from a rural Cambodian village were asked to create music that sounded "angry," "happy," "peaceful," "sad," or "scared" and generated similar melodies to one another within these categories (*56*). These studies suggest that the form of music is systematically related to its affective and behavioral effects in similar ways across cultures. But they can only provide provisional clues about which aspects of music, if any, are universal, because the societies, genres, contexts, and judges are highly limited, and because they too contain little information about music's behavioral contexts across cultures.

A proper evaluation of claims of universality and variation requires a natural history of music: a systematic analysis of the features of musical behavior and musical forms across cultures, using scientific standards of objectivity, representativeness, quantification of variability, and controls for data integrity. We take up this challenge here. We focus on vocal music (hereafter, song) rather than instrumental music [see (*57*)] because it does not depend on technology, has well-defined physical correlates [i.e., pitched vocalizations (*19*)], and has been the primary focus of biological explanations for music (*4*, *5*).

Leveraging more than a century of research from anthropology and ethnomusicology, we built two corpora, which collectively we call the Natural History of Song (NHS). The NHS Ethnography is a corpus of descriptions of song performances, including their context, lyrics, people present, and other details, systematically assembled from the ethnographic record to representatively sample diversity across societies. The NHS Discography is a corpus of field recordings of performances of four kinds of song—dance, healing, love, and lullaby—from an approximately representative sample of human societies, mostly small-scale.

[1]Data Science Initiative, Harvard University, Cambridge, MA 02138, USA. [2]Department of Psychology, Harvard University, Cambridge, MA 02138, USA. [3]School of Psychology, Victoria University of Wellington, Wellington, New Zealand. [4]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [5]Department of Politics, Princeton University, Princeton, NJ 08544, USA. [6]Eastman School of Music, University of Rochester, Rochester, NY 14604, USA. [7]Department of Music, Missouri State University, Springfield, MO 65897, USA. [8]Unaffiliated scholar, Portland, OR 97212, USA. [9]Department of Political Science, Washington University, St. Louis, MO 63130, USA. [10]Computational Auditory Perception Group, Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany. [11]Department of Psychology, Boston College, Chestnut Hill, MA 02467, USA. [12]Department of Psychology, University of Konstanz, 78464 Konstanz, Germany. [13]Department of Linguistics, McGill University, Montreal, QC H3A 1A7, Canada. [14]Department of Anthropology, Pennsylvania State University, State College, PA 16802, USA.
***Corresponding author. Email: sam@wjh.harvard.edu (S.A.M.); manvirsingh@fas.harvard.edu (M.S.); glowacki@psu.edu (L.G.)**

We used the corpora to test five sets of hypotheses about universality and variability in musical behavior and musical forms:

1) We tested whether music is universal by examining the ethnographies of 315 societies, and then a geographically stratified pseudorandom sample of them.

2) We assessed how the behaviors associated with song differ among societies. We reduced the high-dimensional NHS Ethnography annotations to a small number of dimensions of variation while addressing challenges in the analysis of ethnographic data, such as selective nonreporting. This allowed us to assess how the variation in musical behavior across societies compares with the variation within a single society.

3) We tested which behaviors are universally or commonly associated with song. We cataloged 20 common but untested hypotheses about these associations, such as religious activity, dance, and infant care (*4*, *5*, *40*, *54*, *58*–*60*), and tested them after adjusting for sampling error and ethnographer bias, problems that have bedeviled prior tests.

4) We analyzed the musical features of songs themselves, as documented in the NHS Discography. We derived four representations of each song, including blind human ratings and machine summaries. We then applied machine classifiers to these representations to test whether the musical features of a song predict its association with particular behavioral contexts.

5) In exploratory analyses, we assessed the prevalence of tonality in the world's songs, found that variation in their annotations falls along a small number of dimensions, and plotted the statistical distributions of melodic and rhythmic patterns in them.

All data and materials are publicly available at http://osf.io/jmv3q. We also encourage readers to view and listen to the corpora interactively via the plots available at http://themusiclab.org/nhsplots.

Music appears in all measured human societies

Is music universal? We first addressed this question by examining the eHRAF World Cultures database (*61*, *62*), developed and maintained by the Human Relations Area Files organization. It includes high-quality ethnographic documents from 315 societies, subject-indexed by paragraph. We searched for text that was tagged as including music (instrumental or vocal) or that contained at least one keyword identifying vocal music (e.g., "singers").

Music was widespread: The eHRAF ethnographies describe music in 309 of the 315 societies. Moreover, the remaining six (the Turkmen, Dominican, Hazara, Pamir, Tajik, and Ghorbat peoples) do in fact have music, according to primary ethnographic documents available outside the database (*63*–*68*). Thus, music is present in 100% of a large sample of societies, consistent with the claims of writers and scholars since Longfellow (*1*, *4*, *5*, *10*, *12*, *53*, *54*, *58*–*60*, *69*–*73*). Given these data, and assuming that the sample of human societies is representative, the Bayesian 95% posterior credible interval for the population proportion of human societies that have music, with a uniform prior, is [0.994, 1].

To examine what about music is universal and how music varies worldwide, we built the NHS Ethnography (Fig. 1 and Text S1.1), a corpus of 4709 descriptions of song performances drawn from the Probability Sample File (*74*–*76*). This is a ~45-million-word subset of the 315-society database, comprising 60 traditionally living societies that were drawn pseudorandomly from each of Murdock's 60 cultural clusters (*62*), covering 30 distinct geographical

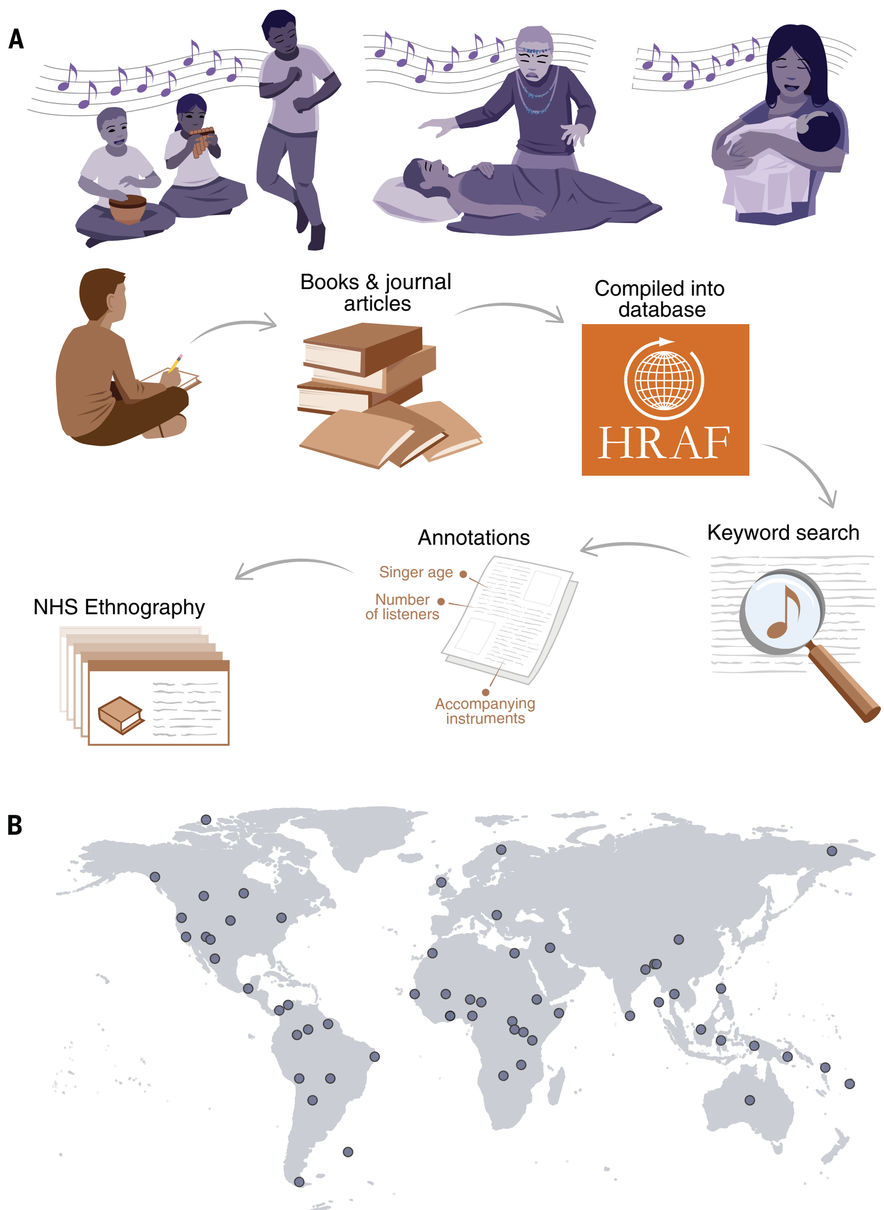


Fig. 1. Design of the NHS Ethnography. The illustration depicts the sequence from acts of singing to the ethnography corpus. (**A**) People produce songs in conjunction with other behavior, which scholars observe and describe in text. These ethnographies are published in books, reports, and journal articles and then compiled, translated, cataloged, and digitized by the Human Relations Area Files organization. (**B**) We conduct searches of the online eHRAF corpus for all descriptions of songs in the 60 societies of the Probability Sample File and annotate them with a variety of behavioral features. The raw text, annotations, and metadata together form the NHS Ethnography. Codebooks listing all available data are in tables S1 to S6; a listing of societies and locations from which texts were gathered is in table S12.

regions and selected to be historically mostly independent of one another. Because the corpus representatively samples from the world's societies, it has been used to test cross-cultural regularities in many domains (*46*, *77*–*83*), and these regularities may be generalized (with appropriate caution) to all societies.

The NHS Ethnography, it turns out, includes examples of songs in all 60 societies. Moreover, each society has songs with words, as opposed to just humming or nonsense syllables (which are reported in 22 societies). Because the societies were sampled independently of whether their people were known to produce music, in contrast to prior cross-cultural studies (*10*, *53*, *54*), the presence of music in each one—as recognized by the anthropologists who embedded themselves in the society and wrote their authoritative ethnographies—constitutes the clearest evidence supporting the claim that song is a human universal. Readers interested in the nature of the ethnographers' reports, which bear on what constitutes "music" in each society [see (*27*)], are encouraged to consult the interactive NHS Ethnography Explorer at http://themusiclab.org/nhsplots.

Musical behavior worldwide varies along three dimensions

How do we reconcile the discovery that song is universal with the research from ethnomusicology showing radical variability? We propose that the music of a society is not a fixed inventory of cultural behaviors, but rather the product of underlying psychological faculties that make certain kinds of sound feel appropriate to certain social and emotional circumstances. These include entraining the body to acoustic and motoric rhythms, analyzing harmonically complex sounds, segregating and grouping sounds into perceptual streams (*6*, *7*), parsing the prosody of speech, responding to emotional calls, and detecting ecologically salient sounds (*8*, *9*). These faculties may interact with others that specifically evolved for music (*4*, *5*). Musical idioms differ with respect to which acoustic features they use and which emotions they engage, but they all draw from a common suite of psychological responses to sound.

If so, what should be universal about music is not specific melodies or rhythms but clusters of correlated behaviors, such as slow soothing lullabies sung by a mother to a child or lively rhythmic songs sung in public by a group of dancers. We thus asked how musical behavior varies worldwide and how the variation within societies compares to the variation between them.

Reducing the dimensionality of variation in musical behavior

To determine whether the wide variation in the annotations of the behavioral context of songs in the database (Text S1.1) falls along a smaller number of dimensions capturing the principal ways that musical behavior varies worldwide, we used an extension of Bayesian principal components analysis (*84*), which, in addition to reducing dimensionality, handles missing data in a principled way and provides a credible interval for each observation's coordinates in the resulting space. Each observation is a "song event," namely, a description in the NHS Ethnography of a song performance, a characterization of how a society uses songs, or both.

We found that three latent dimensions is the optimum number, explaining 26.6% of variability in NHS Ethnography annotations. Figure 2 depicts the space and highlights examples from excerpts in the corpus; an interactive version is available at http://themusiclab.org/nhsplots. (See Text S2.1 for details of the model, including the dimension selection procedure, model diagnostics, a test of robustness, and tests of the potential influence of ethnographer characteristics on model results.) To interpret the space, we examined annotations that load highly on each dimension; to validate this interpretation, we searched for examples at extreme locations and examined their content. Loadings are presented in tables S13 to S15; a selection of extreme examples is given in table S16.

The first dimension (accounting for 15.5% of the variance, including error noise) captures variability in the Formality of a song: Excerpts high along this dimension describe ceremonial events involving adults, large audiences, and instruments; excerpts low on it describe informal events with small audiences and children. The second dimension (accounting for 6.2%) captures variability in Arousal: Excerpts high along this dimension describe lively events with many singers, large audiences, and dancing; excerpts low on it describe calmer events involving fewer people and less overt affect, such as people singing to themselves. The third dimension (4.9%) distinguishes Religious events from secular ones: Passages high along this dimension describe shamanic ceremonies, possession, and funerary songs; passages low on it describe communal events without spiritual content, such as community celebrations.

To validate whether this dimensional space captured behaviorally relevant differences among songs, we tested whether we could reliably recover clusters for four distinctive, easily identifiable, and regularly occurring song types: dance, lullaby, healing, and love (*54*). We searched the NHS Ethnography for keywords and human annotations that matched at least one of the four types (table S17).

Although each song type can appear throughout the space, clear structure is observable (Fig. 2): The excerpts falling into each type cluster together. On average, dance songs (1089 excerpts) occupy the high-Formality, high-Arousal, low-Religiosity region. Healing songs (289 excerpts) cluster in the high-Formality, high-Arousal, high-Religiosity region. Love songs (354 excerpts) cluster in the low-Formality, low-Arousal, low-Religiosity region. Lullabies (156 excerpts) are the sparsest category (although this was likely due to high missingness in variables associated with lullabies, such as one indicating the presence of infant-directed song; see Text S2.1.5) and are located mostly in the low-Formality and low-Arousal regions. An additional 2821 excerpts matched either more than one category or none of the four.

To specify the coherence of these clusters formally rather than just visually, we asked what proportion of song events are closer to the centroid of their own type's location than to any other type (Text S2.1.6). Overall, 64.7% of the songs were located closest to the centroid of their own type; under a null hypothesis that song type is unrelated to location, simulated by randomly shuffling the song labels, only 23.2% would do so ($P < 0.001$ according to a permutation test). This result was statistically significant for three of the four song types (dance, 66.2%; healing, 74.0%; love, 63.6%; $Ps < 0.001$) although not for lullabies (39.7%, $P = 0.92$). The matrix showing how many songs of each type were near each centroid is in table S18. Note that these analyses eliminated variables with high missingness; a validation model that analyzed the entire corpus yielded similar dimensional structure and clustering (figs. S1 and S2 and Text S2.1.5).

The range of musical behavior is similar across societies

We next examined whether this pattern of variation applies within all societies. Do all societies take advantage of the full spectrum of possibilities made available by the neural, cognitive, and cultural systems that underlie music? Alternatively, is there only a single, prototypical song type that is found in all societies, perhaps reflecting the evolutionary origin of music (love songs, say, if music evolved as a courtship display; or lullabies, if it evolved as an adaptation to infant care), with the other types haphazardly distributed or absent altogether, depending on whether the society extended the prototype through cultural evolution? As a third alternative, do societies fall into discrete typologies, such as a Dance Culture or a Lullaby Culture? As still another alternative, do they occupy sectors of the space, so that there are societies with only arousing songs or only religious songs, or societies whose songs are equally formal and vary only by arousal, or vice versa? The data in Fig. 2, which pool song events across societies, cannot answer such questions.

We estimated the variance of each society's scores on each dimension, aggregated across all ethnographies from that society. This revealed

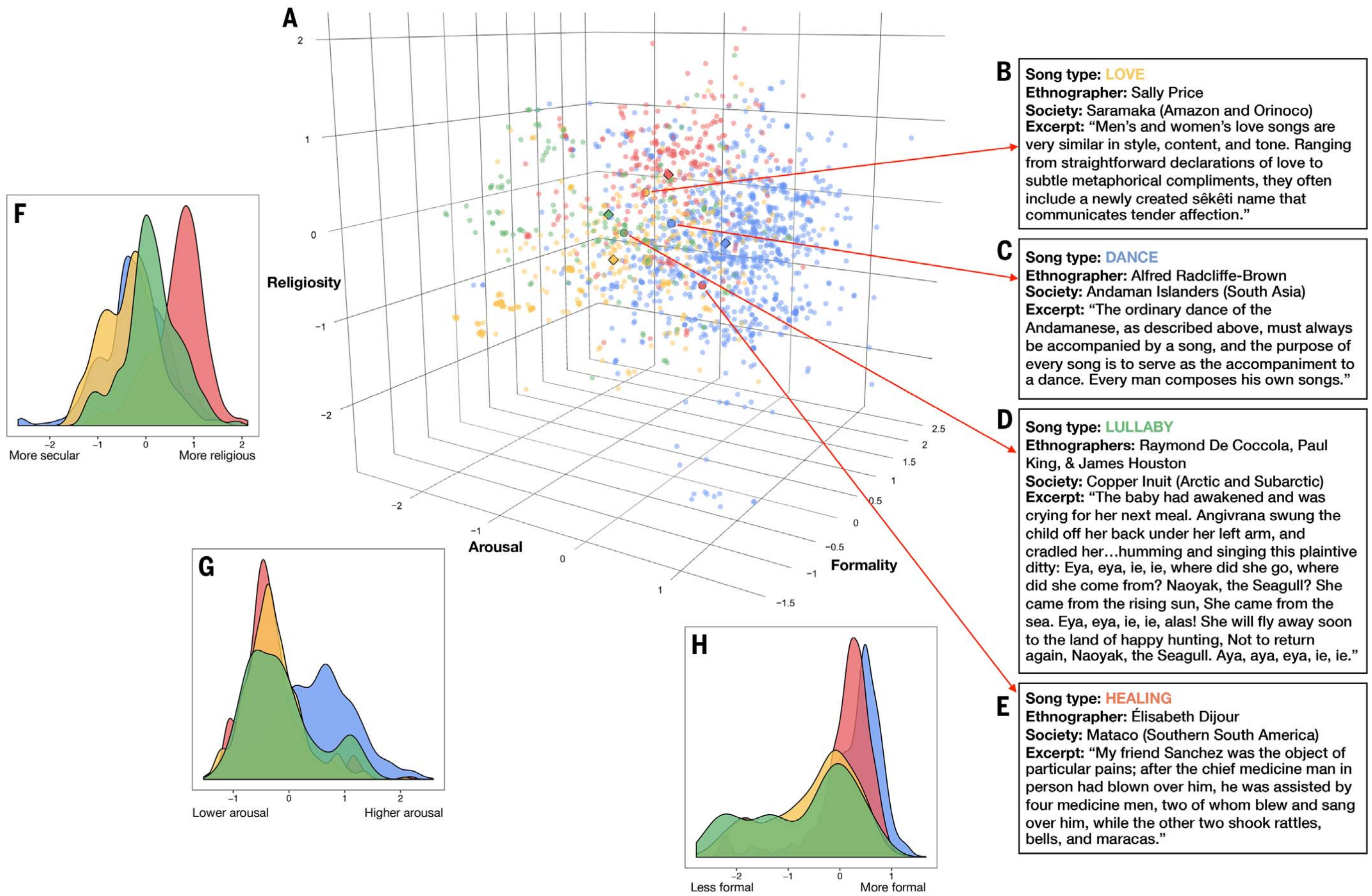


Fig. 2. Patterns of variation in the NHS Ethnography. (**A** to **E**) Projection of a subset of the NHS Ethnography onto three principal components. Each point represents the posterior mean location of an excerpt, with points colored by which of four types (identified by a broad search for matching keywords and annotations) it falls into: dance (blue), lullaby (green), healing (red), or love (yellow). The geometric centroids of each song type are represented by the diamonds. Excerpts that do not match any single search are not plotted but can be viewed in the interactive version of this figure at http://themusiclab.org/nhsplots, along with all text and metadata. Selected examples of each song type are presented here [highlighted circles and (B) to (E)]. (**F** to **H**) Density plots show the differences between song types on each dimension. Criteria for classifying song types from the raw text and annotations are shown in table S17.

that the distributions of each society's observed musical behaviors are remarkably similar (Fig. 3), such that a song with "average formality," "average arousal," or "average religiosity" could appear in any society we studied. This finding is supported by comparing the global average along each dimension to each society's mean and standard deviation, which summarizes how unusual the average song event would appear to members of that society. We found that in every society, a song event at the global mean would not appear out of place: The global mean always falls within the 95% confidence interval of every society's distribution (fig. S3). These results do not appear to be driven by any bias stemming from ethnographer characteristics such as sex or academic field (fig. S4 and Text S2.1.7), nor are they artifacts of a society being related to other societies in the sample by region, subregion, language family, subsistence type, or location in the Old versus New World (fig. S5 and Text S2.1.8).

We also applied a comparison that is common in studies of genetic diversity (*85*) and that has been performed in a recent cultural-phylogenetic study of music (*86*). It revealed that typical within-society variation is approximately six times the between-society variation. Specifically, the ratios of within- to between-society variances were 5.58 for Formality [95% Bayesian credible interval, (4.11, 6.95)]; 6.39 (4.72, 8.34) for Arousal; and 6.21 (4.47, 7.94) for Religiosity. Moreover, none of the 180 mean values for the 60 societies over the three dimensions deviated from the global mean by more than 1.96 times the standard deviation of the principal components scores within that society (fig. S3 and Text S2.1.9).

These findings demonstrate global regularities in musical behavior, but they also reveal that behaviors vary quantitatively across societies, consistent with the long-standing conclusions of ethnomusicologists. For instance, the Kanuri's musical behaviors are estimated to be less formal than those of any other society, whereas those of the Akan are estimated to be the most religious (in both cases, significantly different from the global mean on average). Some ethnomusicologists have attempted to explain such diversity, noting, for example, that more formal song performances tend to be found in more socially rigid societies (*10*).

Despite this variation, a song event of average formality would appear unremarkable in the Kanuri's distribution of songs, as would a song event of average religiosity in the Akan. Overall, we find that for each dimension, approximately one-third of all societies' means significantly differed from the global mean, and approximately half differed from the global mean on at least one dimension (Fig. 3). But despite variability in the societies' means on each dimension, their distributions overlap substantially with one another and with the global mean. Moreover, even the outliers in Fig. 3 appear to represent not genuine idiosyncrasy in some cultures but sampling error: The societies that differ more from the global mean on some dimension are those with sparser documentation in the ethnographic record (fig. S6 and Text S2.1.10). To ensure that these results are not artifacts of the

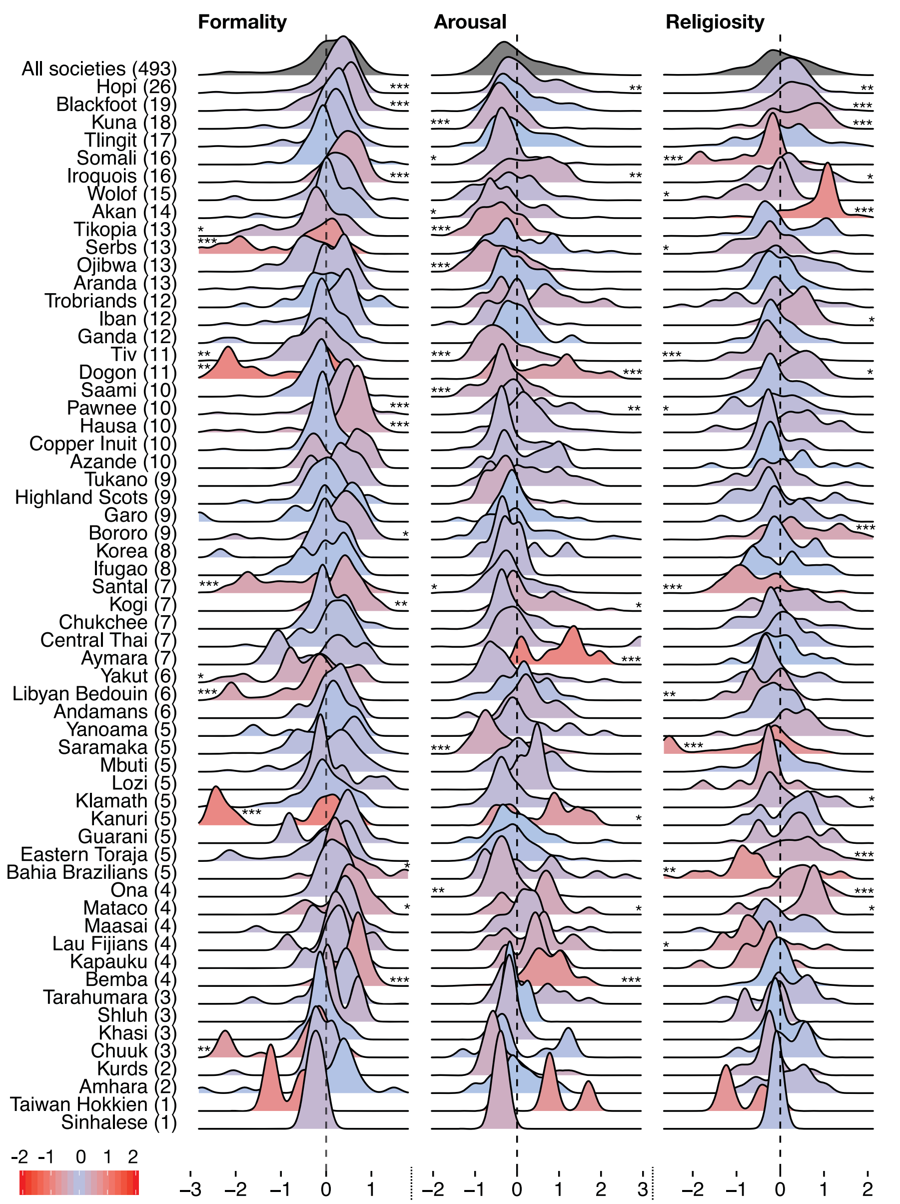


Fig. 3. Society-wise variation in musical behavior. Density plots for each society show the distributions of musical performances on each of the three principal components (Formality, Arousal, Religiosity). Distributions are based on posterior samples aggregated from corresponding ethnographic observations. Societies are ordered by the number of available documents in the NHS Ethnography (the number of documents per society is displayed in parentheses). Distributions are color-coded according to their mean distance from the global mean (in *z*-scores; redder distributions are farther from 0). Although some societies' means differ significantly from the global mean, the mean of each society's distribution is within 1.96 standard deviations of the global mean of 0. One society (Tzeltal) is not plotted because it has insufficient observations for a density plot. Asterisks denote society-level mean differences from the global mean. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

statistical techniques used, we applied them to a structurally analogous dataset whose latent dimensions are expected to vary across countries, namely climate features (for instance, temperature is related to elevation, which certainly is not universal); the results were entirely different from what we found when analyzing the NHS Ethnography (figs. S7 and S8 and Text S2.1.11).

The results suggest that societies' musical behaviors are largely similar to one another, such that the variability within a society exceeds the variability between them (all societies have more soothing songs, such as lullabies; more rousing songs, such as dance tunes; more stirring songs, such as prayers; and other recognizable kinds of musical performance), and that the appearance of uniqueness in the ethnographic record may reflect underreporting.

Associations between song and behavior, corrected for bias

Ethnographic descriptions of behavior are subject to several forms of selective nonreporting: Ethnographers may omit certain kinds of information because of their academic interests (e.g., the author focuses on farming and not shamanism), implicit or explicit biases (e.g., the author reports less information about the elderly), lack of knowledge (e.g., the author is unaware of food taboos), or inaccessibility (e.g., the author wants to report on infant care but is not granted access to infants). We cannot distinguish among these causes, but we can discern patterns of omission in the NHS Ethnography. For example, we found that when the singer's age is reported, the singer is likely to be young, but when the singer's age is not reported, cues that the singer is old are statistically present (such as the fact that a song is ceremonial). Such correlations—between the absence of certain values of one variable and the reporting of particular values of others—were aggregated into a model of missingness (Text S2.1.12) that forms part of the Bayesian principal components analysis reported above. This allowed us to assess variation in musical behavior worldwide, while accounting for reporting biases.

Next, to test hypotheses about the contexts with which music is strongly associated worldwide, in a similarly robust fashion, we compared the frequency with which a particular behavior appears in text describing song with the estimated frequency with which it appears across the board, in all the text written by that ethnographer about that society, which can be treated as the null distribution for that behavior. If a behavior is systematically associated with song, then its frequency in ethnographic descriptions of songs should exceed its frequency in that null distribution, which we estimated by randomly drawing the same number of passages from the same documents [see Text S2.2 for full model details].

We generated a list of 20 hypotheses about universal or widespread contexts for music (Table 1) from published work in anthropology, ethnomusicology, and cognitive science (*4*, *5*, *40*, *54*, *58*–*60*), together with a survey of nearly 1000 scholars that solicited opinions about which behaviors might be universally linked to music (Text S1.4.1). We then designed two sets of criteria for determining whether a given passage of ethnography represented a given behavior in this list. The first used human-annotated identifiers, capitalizing on the fact that every paragraph in the Probability Sample File comes tagged with one of more than 750 identifiers from the Outline of Cultural Materials (OCM),

Table 1. Cross-cultural associations between song and other behaviors. We tested 20 hypothesized associations between song and other behaviors by comparing the frequency of a behavior in song-related passages to that in comparably-sized samples of text from the same sources that are not about song. Behavior was identified with two methods: topic annotations from the Outline of Cultural Materials ("OCM identifiers") and automatic detection of related keywords ("WordNet seed words"; see table S19). Significance tests compared the frequencies in the passages in the full Probability Sample File containing song-related keywords ("Song freq.") with the frequencies in a simulated null distribution of passages randomly selected from the same documents ("Null freq."). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, using adjusted P values (88); 95% intervals for the null distribution are in parentheses.

Hypothesis	OCM identifier(s)	Song freq.	Null freq.	WordNet seed word(s)	Song freq.	Null freq.
Dance	DANCE	1499***	431 (397, 467)	dance	11,145***	3283 (3105, 3468)
Infancy	INFANT CARE	63*	44 (33, 57)	infant, baby, cradle, lullaby	688**	561 (491, 631)
Healing	MAGICAL AND MENTAL THERAPY; SHAMANS AND PSYCHOTHERAPISTS; MEDICAL THERAPY; MEDICAL CARE	1651***	1063 (1004, 1123)	heal, shaman, sick, cure	3983***	2466 (2317, 2619)
Religious activity	SHAMANS AND PSYCHOTHERAPISTS; RELIGIOUS EXPERIENCE; PRAYERS AND SACRIFICES; PURIFICATION AND ATONEMENT; ECSTATIC RELIGIOUS PRACTICES; REVELATION AND DIVINATION; RITUAL	3209***	2212 (2130, 2295)	religious, spiritual, ritual	8644***	5521 (5307, 5741)
Play	GAMES; CHILDHOOD ACTIVITIES	377***	277 (250, 304)	play, game, child, toy	4130***	2732 (2577, 2890)
Procession	SPECTACLES; NUPTIALS	371***	213 (188, 240)	wedding, parade, march, procession, funeral, coronation	2648***	1495 (1409, 1583)
Mourning	BURIAL PRACTICES AND FUNERALS; MOURNING; SPECIAL BURIAL PRACTICES AND FUNERALS	924***	517 (476, 557)	mourn, death, funeral	3784***	2511 (2373, 2655)
Ritual	RITUAL	187***	99 (81, 117)	ritual, ceremony	8520**	5138 (4941, 5343)
Entertainment	SPECTACLES	44***	20 (12, 29)	entertain, spectacle	744***	290 (256, 327)
Children	CHILDHOOD ACTIVITIES	178***	108 (90, 126)	child	4351***	3471 (3304, 3647)
Mood/emotions	DRIVES AND EMOTIONS	219***	138 (118, 159)	mood, emotion, emotive	796***	669 (607, 731)
Work	LABOR AND LEISURE	137***	60 (47, 75)	work, labor	3500**	3223 (3071, 3378)
Storytelling	VERBAL ARTS; LITERATURE	736***	537 (506, 567)	story, history, myth	2792***	2115 (1994, 2239)
Greeting visitors	VISITING AND HOSPITALITY	360***	172 (148, 196)	visit, greet, welcome	1611***	1084 (1008, 1162)
War	WARFARE	264	283 (253, 311)	war, battle, raid	3154***	2254 (2122, 2389)
Praise	STATUS, ROLE, AND PRESTIGE	385	355 (322, 388)	praise, admire, acclaim	481***	302 (267, 339)
Love	ARRANGING A MARRIAGE	158	140 (119, 162)	love, courtship	1625***	804 (734, 876)
Group bonding	SOCIAL RELATIONSHIPS AND GROUPS	141	163 (141, 187)	bond, cohesion	1582***	1424 (1344, 1508)
Marriage/weddings	NUPTIALS	327***	193 (169, 218)	marriage, wedding	2011	2256 (2108, 2410)
Art/creation	N/A	n/a	n/a	art, creation	905***	694 (630, 757)

such as MOURNING, INFANT CARE, or WARFARE.

The second set of criteria was needed because some hypotheses corresponded only loosely to the OCM identifiers (e.g., "love songs" is only a partial fit to ARRANGING A MARRIAGE and not an exact fit to any other identifier), and still others fit no identifier at all [e.g., "music perceived as art or as a creation" (59)]. So we designed a method that examined the text directly. Starting with a small set of seed words associated with each hypothesis (e.g., "religious," "spiritual," and "ritual" for the hypothesis that music is associated with religious activity), we used the WordNet lexical database (87) to automatically generate lists of conceptually related terms (e.g., "rite" and "sacred"). We manually filtered the

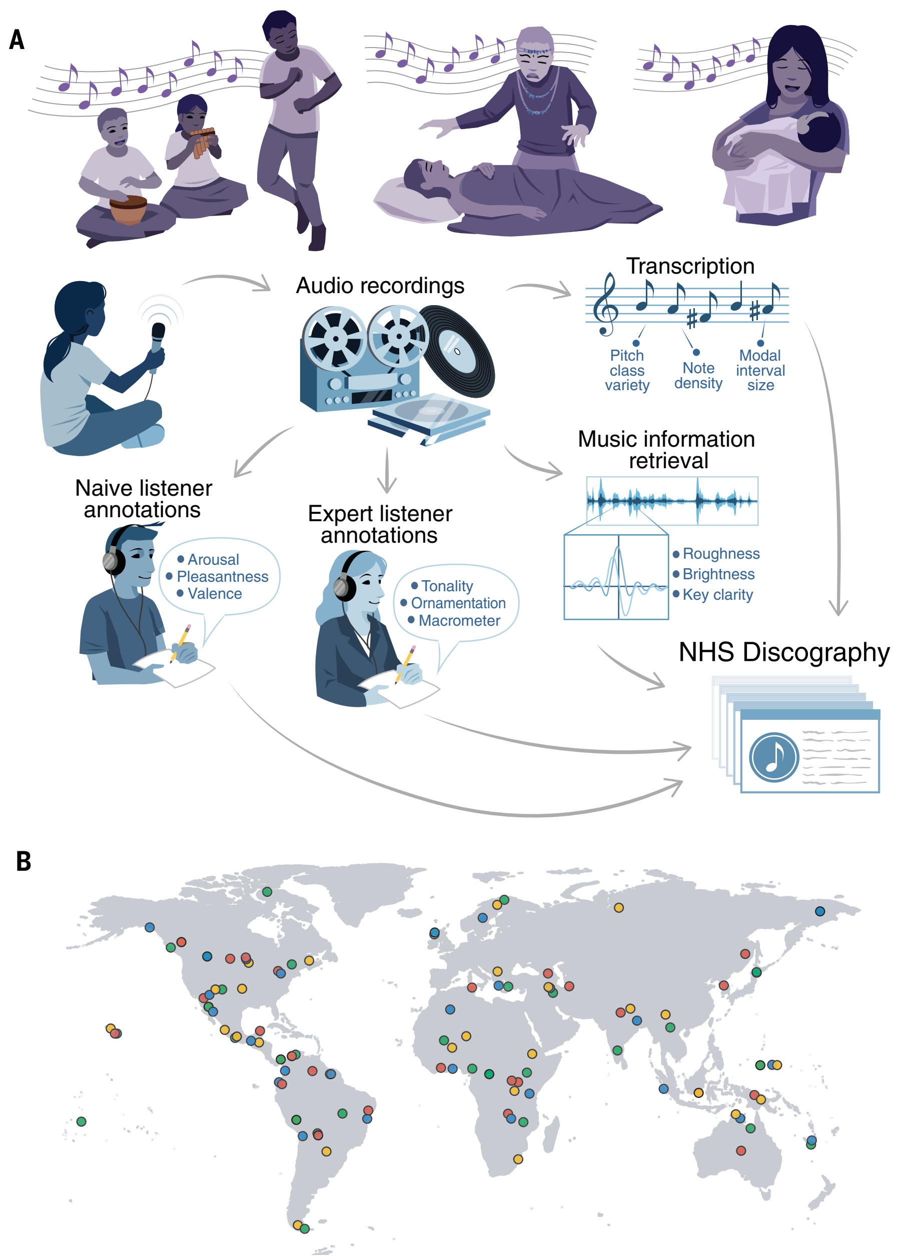


Fig. 4. Design of the NHS Discography. (**A**) Illustration depicting the sequence from acts of singing to the audio discography. People produce songs, which scholars record. We aggregate and analyze the recordings via four methods: automatic music information retrieval, annotations from expert listeners, annotations from naïve listeners, and staff notation transcriptions (from which annotations are automatically generated). The raw audio, four types of annotations, transcriptions, and metadata together form the NHS Discography. (**B**) Plot of the locations of the 86 societies represented, with points colored by the song type in each recording (blue, dance; red, healing; yellow, love; green, lullaby). Codebooks listing all available data are in tables S1 and S7 to S11; a listing of societies and locations from which recordings were gathered is in table S22.

lists to remove irrelevant words and homonyms and add relevant keywords that may have been missed, then conducted word stemming to fill out plurals and other grammatical variants (full lists are in table S19). Each method has limitations: Automated dictionary methods can erroneously flag a passage containing a word that is ambiguous, whereas the human-coded OCM identifiers may miss a relevant passage, misinterpret the original ethnography, or paint with too broad a brush, applying a tag to a whole paragraph or to several pages of text. Where the two methods converge, support for a hypothesis is particularly convincing.

After controlling for ethnographer bias via the method described above, and adjusting the *P* values for multiple hypotheses (*88*), we found support from both methods for 14 of the 20 hypothesized associations between music and a behavioral context, and support from one method for the remaining six (Table 1). To verify that these analyses specifically confirmed the hypotheses, as opposed to being an artifact of some other nonrandom patterning in this dataset, we reran them on a set of additional OCM identifiers matched in frequency to the ones used above [see Text S2.2.2 for a description of the selection procedure]. They covered a broad swath of topics, including DOMESTICATED ANIMALS, POLYGAMY, and LEGAL NORMS that were not hypothesized to be related to song (the full list is in table S20). We find that only one appeared more frequently in song-related paragraphs than in the simulated null distribution (CEREAL AGRICULTURE; see table S20 for full results). This contrasts sharply with the associations reported in Table 1, suggesting that they represent bona fide regularities in the behavioral contexts of music.

Universality of musical forms

We now turn to the NHS Discography to examine the musical content of songs in four behavioral contexts (dance, lullaby, healing, and love; Fig. 4A), selected because each appears in the NHS Ethnography, is widespread in traditional cultures (*59*), and exhibits shared features across societies (*54*). Using predetermined criteria based on liner notes and supporting ethnographic text (table S21), and seeking recordings of each type from each of the 30 geographic regions, we found 118 songs of the 120 possibilities (4 contexts × 30 regions) from 86 societies (Fig. 4B). This coverage underscores the universality of these four types; indeed, in the two possibilities we failed to find (healing songs from Scandinavia and from the British Isles), documentary evidence shows that both existed (*89*, *90*) despite our failure to find audio recordings of the practice. The recordings may be unavailable because healing songs were rare by the early 1900s, roughly when portable field recording became feasible.

The data describing each song comprised (i) machine summaries of the raw audio using automatic music information retrieval techniques, particularly the audio's spectral features (e.g., mean brightness and roughness, variability of spectral entropy) (Text S1.2.1); (ii) general impressions of musical features (e.g., whether its emotional valence was happy or sad) by untrained listeners recruited online from the United States and India (Text S1.2.2); (iii) ratings of additional music-theoretic features such as high-level rhythmic grouping structure [similar in concept to previous rating-scale approaches to analyzing world music (*10*, *53*)] from a group of 30 expert musicians including Ph.D. ethnomusicologists and music theorists (Text S1.2.3); and (iv) detailed manual

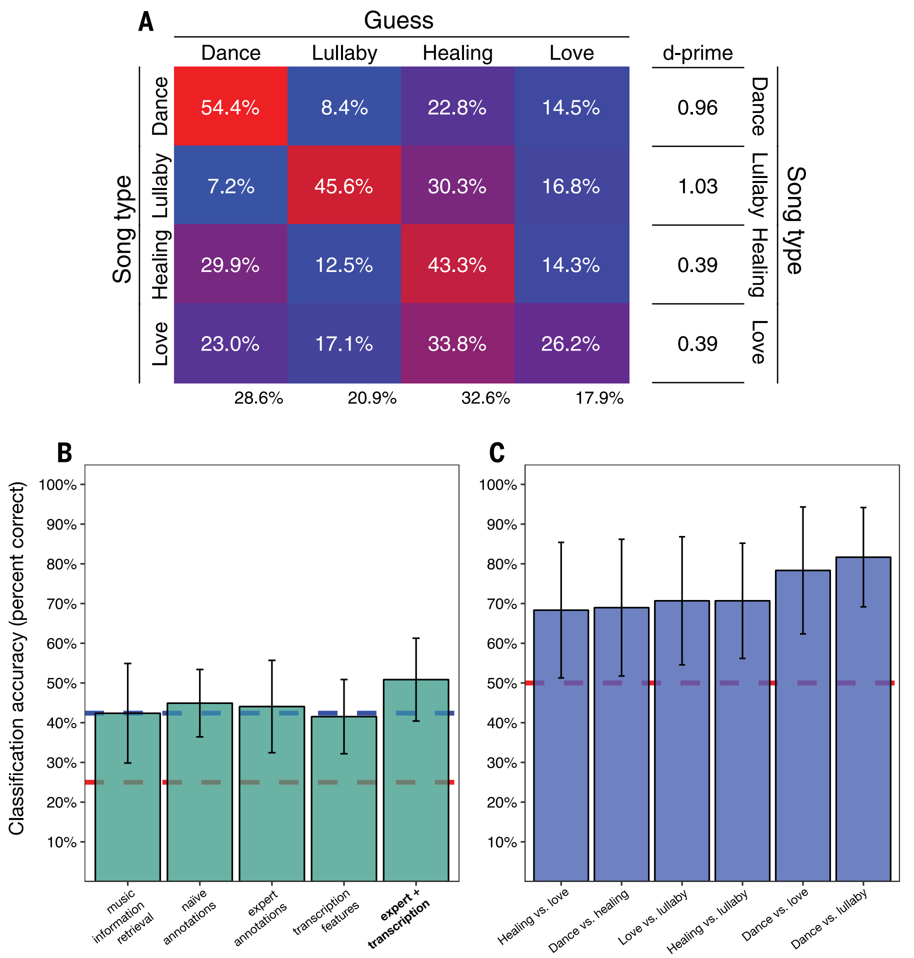


Fig. 5. Form and function in song. (**A**) In a massive online experiment (N = 29,357), listeners categorized dance songs, lullabies, healing songs, and love songs at rates higher than chance level of 25%, but their responses to love songs were by far the most ambiguous (the heat map shows average percent correct, color-coded from lowest magnitude, in blue, to highest magnitude, in red). Note that the marginals (below the heat map) are not evenly distributed across behavioral contexts: Listeners guessed "healing" most often and "love" least often despite the equal number of each in the materials. The *d*-prime scores estimate listeners' sensitivity to the song-type signal independent of this response bias. (**B**) Categorical classification of the behavioral contexts of songs, using each of the four representations in the NHS Discography, is substantially above the chance performance level of 25% (dotted red line) and is indistinguishable from the performance of human listeners, 42.4% (dotted blue line). The classifier that combines expert annotations with transcription features (the two representations that best ignore background sounds and other context) performs at 50.8% correct, above the level of human listeners. (**C**) Binary classifiers that use the expert annotation + transcription feature representations to distinguish pairs of behavioral contexts [e.g., dance from love songs, as opposed to the four-way classification in (B)] perform above the chance level of 50% (dotted red line). Error bars represent 95% confidence intervals from corrected resampled *t* tests (*94*).

transcriptions, also by expert musicians, of musical features (e.g., note density of sung pitches) (Text S1.2.4). To ensure that classifications were driven only by the content of the music, we excluded any variables that carried explicit or implicit information about the context (*54*), such as the number of singers audible on a recording and a coding of polyphony (which indicates the same thing implicitly). This exclusion could be complete only in the manual transcriptions, which are restricted to data on vocalizations; the music information retrieval and naïve listener data are practically inseparable from contextual information, and the expert listener ratings contain at least a small amount, because despite being told to ignore the context, the experts could still hear some of it, such as accompanying instruments. [See Text S2.3.1 for details about variable exclusion.]

Listeners accurately identify the behavioral contexts of songs

In a previous study, people listened to recordings from the NHS Discography and rated their confidence in each of six possible behavioral contexts (e.g., "used to soothe a baby"). On average, the listeners successfully inferred a song's behavioral context from its musical forms: The songs that were actually used to soothe a baby (i.e., lullabies) were rated highest as "used to soothe a baby"; dance songs were rated highly as "used for dancing," and so on (*54*).

We ran a massive conceptual replication (Text S1.4.2) where 29,357 visitors to the citizen-science website http://themusiclab.org listened to songs drawn at random from the NHS Discography and were asked to guess what kind of song they were listening to from among four alternatives (yielding 185,832 ratings, i.e., 118 songs rated about 1500 times each). Participants also reported their musical skill level and degree of familiarity with world music. Listeners guessed the behavioral contexts with a level of accuracy (42.4%) that is well above chance (25%), showing that the acoustic properties of a song performance reflect its behavioral context in ways that span human cultures.

The confusion matrix (Fig. 5A) shows that listeners identified dance songs most accurately (54.4%), followed by lullabies (45.6%), healing songs (43.3%), and love songs (26.2%), all significantly above chance (Ps < 0.001). Dance songs and lullabies were the least likely to be confused with each other, presumably because of their many contrasting features, such as tempo (a possibility we examine below; see Table 2). The column marginals suggest that the raters were biased toward identifying recordings as healing songs (32.6%, above their actual proportion of 23.7%) and away from identifying them as love songs (17.9%), possibly because healing songs are less familiar to Westernized listeners and they were overcompensating in identifying examples. As in previous research (*54*), love songs were least reliably identified, despite their ubiquity in Western popular music, possibly because they span a wide range of styles (for example, the vastly different Elvis Presley hit singles "Love Me Tender" and "Burning Love"). Nonetheless, *d*-prime scores (Fig. 5A), which capture the sensitivity to a signal independently of response bias, show that all behavioral contexts were identified at a rate higher than chance ($d' = 0$).

Are accurate identifications of the contexts of culturally unfamiliar songs restricted to listeners with musical training or exposure to world music? In a regression analysis, we found that participants' categorization accuracy was statistically related to their self-reported musical skill [$F(4,16245) = 2.57$, P = 0.036] and their familiarity with world music [$F(3,16167) = 36.9$, $P < 0.001$; statistics from linear probability models], but with small effect sizes: The largest difference was a 4.7–percentage point advantage for participants

Table 2. Features of songs that distinguish between behavioral contexts. The table reports the predictive influence of musical features in the NHS Discography in distinguishing song types across cultures, ordered by their overall influence across all behavioral contexts. The classifiers used the average rating for each feature across 30 annotators. The coefficients are from a penalized logistic regression with standardized features and are selected for inclusion using a LASSO for variable selection. For brevity, we only present the subset of features with notable influence on a pairwise comparison (coefficients greater than 0.1). Changes in the values of the coefficients produce changes in the predicted log-odds ratio, so the values in the table can be interpreted as in a logistic regression.

		Coefficient (pairwise comparison)					
Musical feature	**Definition**	**Dance (–) vs. Lullaby (+)**	**Dance (–) vs. Love (+)**	**Healing (–) vs. Lullaby (+)**	**Love (–) vs. Lullaby (+)**	**Dance (–) vs. Healing (+)**	**Healing (–) vs. Love (+)**
Accent	The differentiation of musical pulses, usually by volume or emphasis of articulation. A fluid, gentle song will have few accents and a correspondingly low value.	–0.64	–0.24	–0.85	–0.41	.	–0.34
Tempo	The rate of salient rhythmic pulses, measured in beats per minute; the perceived speed of the music. A fast song will have a high value.	–0.65	–0.51	.	.	–0.76	.
Quality of pitch collection	Major versus minor key. In Western music, a key usually has a "minor" quality if its third note is three semitones from the tonic. This variable was derived from annotators' qualitative categorization of the pitch collection, which we then dichotomized into Major (0) or Minor (1).	.	0.26	0.44	.	–0.37	0.35
Consistency of macrometer	Meter refers to salient repetitive patterns of accent within a stream of pulses. A micrometer refers to the low-level pattern of accents; a macrometer refers to repetitive patterns of micrometer groups. This variable refers to the consistency of the macrometer, in an ordinal scale, from "No macrometer" (1) to "Totally clear macrometer" (6). A song with a highly variable macrometer will have a low value.	–0.44	–0.49	.	.	–0.46	.
Number of common intervals	Variability in interval sizes, measured by the number of different melodic interval sizes that constitute more than 9% of the song's intervals. A song with a large number of different melodic interval sizes will have a high value.	.	0.58	.	.	.	0.62
Pitch range	The musical distance between the extremes of pitch in a melody, measured in semitones. A song that includes very high and very low pitches will have a high value.	.	.	.	–0.49	.	.

continued on next page

		Coefficient (pairwise comparison)					
Musical feature	**Definition**	**Dance (−) vs. Lullaby (+)**	**Dance (−) vs. Love (+)**	**Healing (−) vs. Lullaby (+)**	**Love (−) vs. Lullaby (+)**	**Dance (−) vs. Healing (+)**	**Healing (−) vs. Love (+)**
Stepwise motion	Stepwise motion refers to melodic strings of consecutive notes (1 or 2 semitones apart), without skips or leaps. This variable consists of the fraction of all intervals in a song that are 1 or 2 semitones in size. A song with many melodic leaps will have a low value.	.	.	.	.	0.61	−0.20
Tension/release	The degree to which the passage is perceived to build and release tension via changes in melodic contour, harmonic progression, rhythm, motivic development, accent, or instrumentation. If so, the song is annotated with a value of 1.	.	0.27	.	.	.	0.27
Average melodic interval size	The average of all interval sizes between successive melodic pitches, measured in semitones on a 12-tone equal temperament scale, rather than in absolute frequencies. A melody with many wide leaps between pitches will have a high value.	.	−0.46	.	.	.	.
Average note duration	The mean of all note durations; a song predominated by short notes will have a low value.	.	.	.	.	.	−0.49
Triple micrometer	A low-level pattern of accents that groups together pulses in threes.	.	.	.	.	−0.23	.
Predominance of most common pitch class	Variety versus monotony of the melody, measured by the ratio of the proportion of occurrences of the second most common pitch (collapsing across octaves) to the proportion of occurrences of the most common pitch; monotonous melodies will have low values.	.	.	.	.	−0.48	.
Rhythmic variation	Variety versus monotony of the rhythm, judged subjectively and dichotomously. Repetitive songs have a low value.	.	.	.	.	0.42	.
Tempo variation	Changes in tempo: A song that is perceived to speed up or slow down is annotated with a value of 1.	.	.	.	.	.	−0.27
Ornamentation	Complex melodic variation or “decoration” of a perceived underlying musical structure. A song perceived as having ornamentation is annotated with a value of 1.	.	0.25	.	.	.	.

continued on next page

		Coefficient (pairwise comparison)					
Musical feature	Definition	Dance (−) vs. Lullaby (+)	Dance (−) vs. Love (+)	Healing (−) vs. Lullaby (+)	Love (−) vs. Lullaby (+)	Dance (−) vs. Healing (+)	Healing (−) vs. Love (+)
Pitch class variation	A pitch class is the group of pitches that sound equivalent at different octaves, such as all the Cs, not just middle C. This variable, another indicator of melodic variety, counts the number of pitch classes that appear at least once in the song.	.	.	−0.25	.	.	.
Triple macrometer	If a melody arranges micrometer groups into larger phrases of three, like a waltz, it is annotated with a value of 1.	.	.	0.14	.	.	.
Predominance of most common interval	Variability among pitch intervals, measured as the fraction of all intervals that are the most common interval size. A song with little variability in interval sizes will have a high value.	.	.	.	.	0.12	.

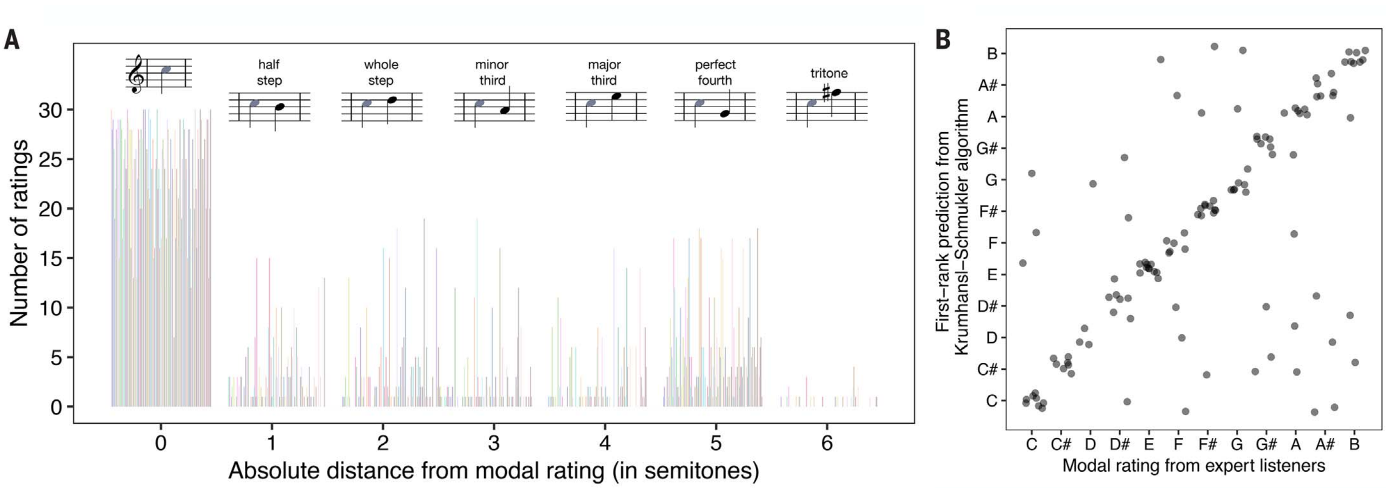


Fig. 6. Signatures of tonality in the NHS Discography. (**A**) Histograms representing 30 expert listeners' ratings of tonal centers in all 118 songs, each song corresponding to a different color, show two main findings: (i) Most songs' distributions are unimodal, such that most listeners agreed on a single tonal center (represented by the value 0). (ii) When listeners disagree, they are multimodal, with the most popular second mode (in absolute distance) five semitones away from the overall mode, a perfect fourth. The music notation is provided as a hypothetical example only, with C as a reference tonal center; note that the ratings of tonal centers could be at any pitch level. (**B**) The scatterplot shows the correspondence between modal ratings of expert listeners with the first-rank predictions from the Krumhansl-Schmuckler key-finding algorithm. Points are jittered to avoid overlap. Note that pitch classes are circular (i.e., C is one semitone away from C# and from B) but the plot is not; distances on the axes of (B) should be interpreted accordingly.

who reported that they were "somewhat familiar with traditional music" relative to those who reported that they had never heard it, and a 1.3–percentage point advantage for participants who reported that they have "a lot of skill" relative to "no skill at all." Moreover, when limiting the dataset to listeners with "no skill at all" or listeners who had "never heard traditional music," mean accuracy was almost identical to the overall cohort. These findings suggest that although musical experience enhances the ability to detect the behavioral contexts of songs from unfamiliar cultures, it is not necessary.

Quantitative representations of musical forms accurately predict behavioral contexts of song

If listeners can accurately identify the behavioral contexts of songs from unfamiliar cultures, there must be acoustic features that universally tend to be associated with these contexts. To identify them, we evaluated the relationship between a song's musical forms [measured in four ways; see Text S1.2.5 and (*12*, *31*, *32*, *91*–*93*) for discussion of how difficult it is to represent music quantitatively] and its behavioral context. We used a cross-validation procedure that determined whether the pattern of correlation between musical forms and context computed from a subset of the

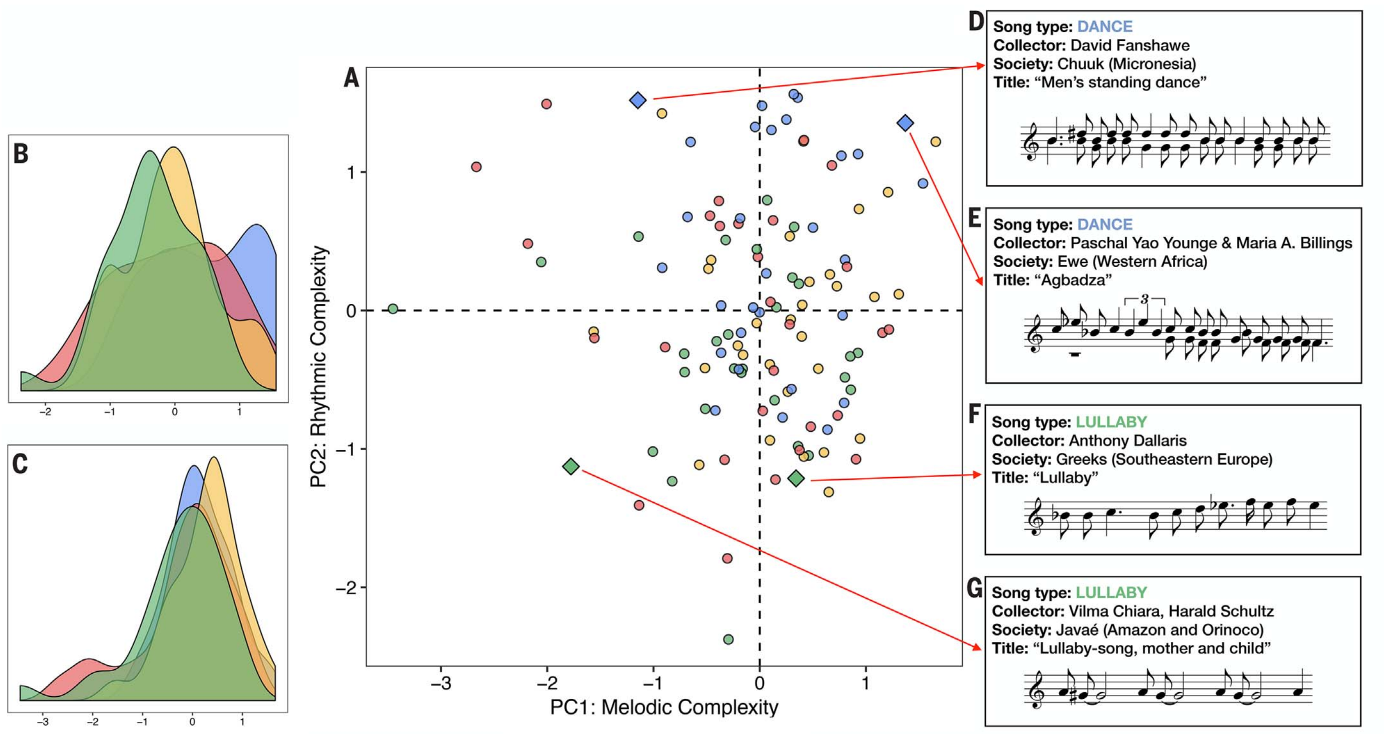


Fig. 7. Dimensions of musical variation in the NHS Discography. (**A**) A Bayesian principal components analysis reduction of expert annotations and transcription features (the representations least contaminated by contextual features) shows that these measurements fall along two dimensions that may be interpreted as rhythmic complexity and melodic complexity. (**B** and **C**) Histograms for each dimension show the differences—or lack thereof—between behavioral contexts. (**D** to **G**) Excerpts of transcriptions from songs at extremes from each of the four quadrants, to validate the dimension reduction visually. The two songs at the high–rhythmic complexity quadrants are dance songs (in blue); the two songs at the low–rhythmic complexity quadrants are lullabies (in green). Healing songs are depicted in red and love songs in yellow. Readers can listen to excerpts from all songs in the corpus at http://osf.io/jmv3q; an interactive version of this plot is available at http://themusiclab.org/nhsplots.

regions could be generalized to predict a song's context in the other regions (as opposed to being overfitted to arbitrary correlations within that subsample). Specifically, we trained a least absolute shrinkage and selection operator (LASSO) multinomial logistic classifier (*94*) on the behavioral context of all the songs in 29 of the 30 regions in the NHS Discography, and used it to predict the context of the unseen songs in the 30th. We ran this procedure 30 times, omitting a different region each time (table S23 and Text S2.3.2). We compared the accuracy of these predictions to two baselines: pure chance (25%) and the accuracy of listeners in the massive online experiment (see above) when guessing the behavioral context from among four alternatives (42.4%).

We found that with each of the four representations, the musical forms of a song can predict its behavioral context (Fig. 5B) at high rates, comparable to those of the human listeners in the online experiment. This finding was not attributable to information in the recordings other than the singing, which could be problematic if, for example, the presence of a musical instrument on a recording indicated that it was likelier to be a dance song than a lullaby (*54*), artificially improving classification. Representations with the least extraneous influence—the expert annotators and the summary features extracted from transcriptions—had no lower classification accuracy than the other representations. And a classifier run on combined expert + transcription data had the best performance of all, 50.8% [95% CI (40.4%, 61.3%), computed by corrected resampled t test (*95*)].

To ensure that this accuracy did not merely consist of patterns in one society predicting patterns in historically or geographically related ones, we repeated the analyses, cross-validating across groupings of societies, including superordinate world region (e.g., "Asia"), subsistence type (e.g., "hunter-gatherers"), and Old versus New World. In many cases, the classifier performed comparably to the main model (table S24), although low power in some cases (i.e., training on less than half the corpus) substantially reduced precision.

In sum, the acoustic form of vocal music predicts its behavioral contexts worldwide (*54*), at least in the contexts of dance, lullaby, healing, and love: All classifiers performed above chance and within 1.96 standard errors of the performance of human listeners.

Musical features that characterize the behavioral contexts of songs across societies

Showing that the musical features of songs predict their behavioral context provides no information about which musical features those are. To help identify them, we determined how well the combined expert + transcription data distinguished between specific pairs of behavioral contexts rather than among all four, using a simplified form of the classifiers described above, which not only distinguished the contexts but also identified the most reliable predictors of each contrast, without overfitting (*96*). This can reveal whether tempo, for example, helps distinguish dance songs from lullabies while failing to distinguish lullabies from love songs.

Performance once again significantly exceeded chance (in this case, 50%) for all six comparisons (Ps < 0.05; Fig. 5C). Table 2 lays out the musical features that drive these successful predictions and thereby characterize the four song types across cultures. Some are consistent with common sense; for instance, dance songs differ from lullabies in tempo, accent, and the consistency of their macrometer (i.e., the superordinate grouping of rhythmic notes). Other distinguishers are

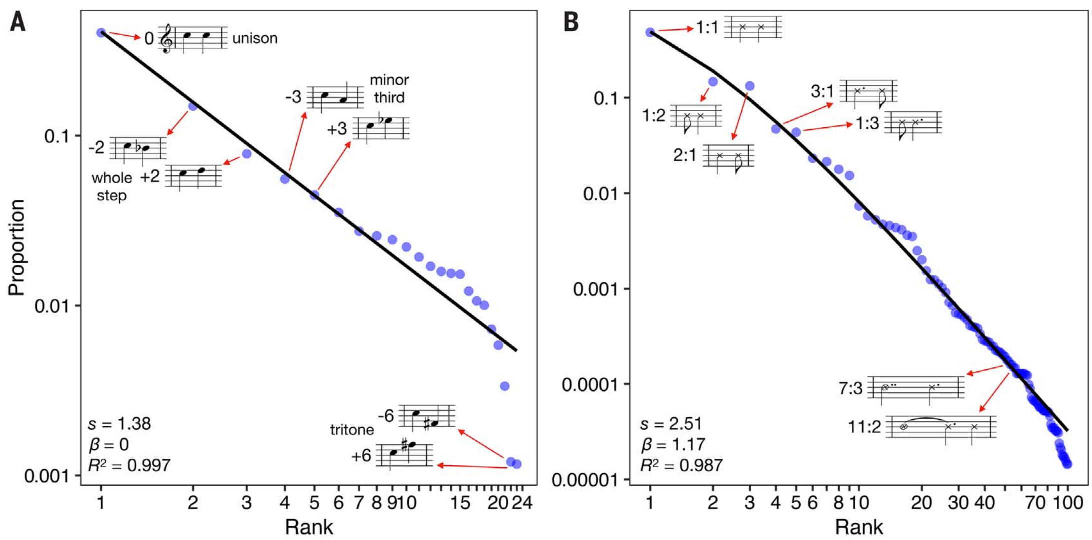


Fig. 8. The distributions of melodic and rhythmic patterns in the NHS Discography follow power laws. (**A** and **B**) We computed relative melodic (A) and rhythmic (B) bigrams and examined their distributions in the corpus. Both distributions followed a power law; the parameter estimates in the inset correspond to those from the generalized Zipf-Mandelbrot law, where *s* refers to the exponent of the power law and β refers to the Mandelbrot offset. Note that in both plots, the axes are on logarithmic scales. The full lists of bigrams are in tables S28 and S29.

subtler: The most common interval of a song occurs a smaller proportion of the time in a dance song than in a healing song, which suggests that dance songs are more melodically variable than healing songs (for explanations of musical terminology, see Table 2). Similarly, it is unsurprising that lullabies and love songs are more difficult to distinguish than lullabies and dance songs (*97*); nonetheless, they may be distinguished by two features: the strength of metrical accents and the size of the pitch range (both larger in love songs).

In sum, four common song categories, distinguished by their contexts and goals, tend to have distinctive musical qualities worldwide. These results suggest that universal features of human psychology bias people to produce and enjoy songs with certain kinds of rhythmic or melodic patterning that naturally go with certain moods, desires, and themes. These patterns do not consist of concrete acoustic features, such as a specific melody or rhythm, but rather of relational properties such as accent, meter, and interval structure.

Of course, classification accuracy that is twice the level of chance still falls well short of perfect prediction; hence, many aspects of music cannot be manifestations of universal psychological reactions. Although musical features can predict differences between songs from these four behavioral contexts, a given song may be sung in a particular context for other reasons, including its lyrics, its history, the style and instrumentation of its performance, its association with mythical or religious themes, and constraints of the culture's musical idiom. And although we have shown that Western listeners, who have been exposed to a vast range of musical styles and idioms, can distinguish the behavioral contexts of songs from non-Western societies, we do not know whether non-Western listeners can do the same. To reinforce the hypothesis of universal associations between musical form and context, similar methods should be tested with non-Western listeners.

Explorations of the structure of musical forms

The NHS Discography can be used to explore world music in many other ways. We present three exploratory analyses here, mindful of the limitation that they may apply only to the four genres the corpus includes.

Signatures of tonality appear in all societies studied

A basic feature of many styles of music is tonality, in which a melody is composed of a fixed set of discrete tones [perceived pitches as opposed to actual pitches, a distinction dating to Aristoxenus's *Elementa Harmonica* (*98*)], and some tones are psychologically dependent on others, with one tone felt to be central or stable (*99*–*101*). This tone (more accurately, perceived pitch class, embracing all the tones one or more octaves apart) is called the tonal center or tonic, and listeners characterize it as a reference point, point of stability, basis tone, "home," or tone that the melody "is built around" and where it "should end." For example, the tonal center of "Row Your Boat" is found in each of the "row"s, the last "merrily," and the song's last note, "dream."

Although tonality has been studied in a few non-Western societies (*102*, *103*), its cross-cultural distribution is unknown. Indeed, the ethnomusicologists who responded to our survey (Text S1.4.1) were split over whether the music of all societies should be expected to have a tonal center: 48% responded "probably not universal" or "definitely not universal." The issue is important because a tonal system is a likely prerequisite for analyzing music, in all its diversity, as the product of an abstract musical grammar (*73*). Tonality also motivates the hypothesis that melody is rooted in the brain's analysis of harmonically complex tones (*104*). In this theory, a melody can be considered a set of "serialized overtones," the harmonically related frequencies ordinarily superimposed in the rich tone produced by an elongated resonator such as the human vocal tract. In tonal melodies, the tonic corresponds to the fundamental frequency of the disassembled complex tone, and listeners tend to favor tones in the same pitch class as harmonics of the fundamental (*105*).

To explore tonality in the NHS Discography, we analyzed the expert listener annotations and the transcriptions (Text S2.4.1). Each of the 30 expert listeners was asked, for each song, whether or not they heard at least one tonal center, defined subjectively as above. The results were unambiguous: 97.8% of ratings were in the affirmative. More than two-thirds of songs were rated as "tonal" by all 30 expert listeners, and 113 of the 118 were rated as tonal by more than 90% of them. The song with the most ambiguous tonality (the Kwakwaka'wakw healing song) still had a majority of raters respond in the affirmative (60%).

If listeners heard a tonal center, they were asked to name its pitch class. Here too, listeners were highly consistent: Either there was widespread agreement on a single tonal center or the responses fell into two or three tonal centers (Fig. 6A; the distributions of tonality ratings for all 118 songs are in fig. S10). We used Hartigan's dip test (*106*) to measure the multimodality of the ratings. In the 73 songs that the test classified as unimodal, 85.3% of ratings were in agreement with the modal pitch class. In the remaining 45 songs, 81.7% of ratings were in agreement with the two most popular pitch classes, and 90.4% were in agreement with the three most popular. The expert listeners included six Ph.D. ethnomusicologists and six Ph.D. music theorists; when restricting the ratings to this group alone, the levels of consistency were comparable.

In songs where the ratings were multimodally distributed, the modal tones were often hierarchically related; for instance, ratings for the Ojibwa healing song were evenly split between B (pitch class 11) and E (pitch class 4), which are a perfect fourth (five semitones) apart. The most common intervals between the two modal tones were the perfect fourth (in 15 songs), a half-step (one semitone, in nine songs), a whole step (two semitones, in eight songs), a major third (four semitones, in seven songs), and a minor third (three semitones, in six songs).

We cannot know which features of a given recording our listeners were responding to in attributing a tonal center to it, nor whether their attributions depended on expertise that ordinary listeners lack. We thus sought converging, objective evidence for the prevalence of tonality in the world's music by submitting NHS Discography transcriptions to the Krumhansl-Schmuckler key-finding algorithm (*107*). This algorithm sums the durations of the tones in a piece of music and correlates this vector with each of a family of candidate vectors, one for each key, consisting of the relative centralities of those pitch classes in that key. The algorithm's first guess (i.e., the key corresponding to the most highly correlated vector) matched the expert listeners' ratings of the tonal center 85.6% of the time (measured via a weighted average of its hit rate for the most common expert rating when the ratings were unimodal and either of the two most common ratings when they were multimodal). When we relaxed the criterion for a match to the algorithm's first- and second-ranked guesses, it matched the listeners' ratings on 94.1% of songs; adding its third-ranked estimate resulted in matches 97.5% of the time, and adding the fourth resulted in matches with 98.3% [all Ps < 0.0001 above the chance level of 9.1%, using a permutation test (Text S2.4.1)]. These results provide convergent evidence for the presence of tonality in the NHS Discography songs (Fig. 6B).

These conclusions are limited in several ways. First, they are based on songs from only four behavioral contexts, omitting others such as mourning, storytelling, play, war, and celebration. Second, the transcriptions were created manually and could have been influenced by the musical ears and knowledge of the expert transcribers. (Current music information retrieval algorithms are not robust enough to transcribe melodies accurately, especially from noisy field recordings, but improved ones could address this issue.) The same limitation may apply to the ratings of our expert listeners. Finally, the findings do not show how the people from the societies in which NHS Discography songs were recorded hear the tonality in their own music. To test the universality of tonality perception, one would need to conduct field experiments in diverse populations.

Music varies along two dimensions of complexity

To examine patterns of variation among the songs in the NHS Discography, we applied the same kind of Bayesian principal components analysis used for the NHS Ethnography to the combination of expert annotations and transcription features (i.e., the representations that focus most on the singing, excluding context). The results yielded two dimensions, which together explain 23.9% of the variability in musical features. The first, which we call Melodic Complexity, accounts for 13.1% of the variance (including error noise); heavily loading variables included the number of common intervals, pitch range, and ornamentation (all positively) and the predominance of the most common pitch class, the predominance of the most common interval, and the distance between the most common intervals (all negatively; see table S25). The second, which we call Rhythmic Complexity, accounts for 10.8% of the variance; heavily loading variables included tempo, note density, syncopation, accent, and consistency of macrometer (all positively) and the average note duration and duration of melodic arcs (all negatively; see table S26). The interpretation of the dimensions is further supported in Fig. 7, which shows excerpts of transcriptions at the extremes of each dimension; an interactive version is at http://themusiclab.org/nhsplots.

In contrast to the NHS Ethnography, the principal components space for the NHS Discography does not distinguish the four behavioral contexts of songs in the corpus. We found that only 39.8% of songs matched their nearest centroid (overall P = 0.063 from a permutation test; dance: 56.7%, P = 0.12; healing: 7.14%, P > 0.99; love: 43.3%, P = 0.62; lullaby: 50.0%, P = 0.37; a confusion matrix is in table S27). Similarly, k-means clustering on the principal components space, asserting k = 4 (because there are four known clusters), failed to reliably capture any of the behavioral contexts. Finally, given the lack of predictive accuracy of songs' location in the two-dimensional space, we explored each dimension's predictive accuracy individually, using t tests of each context against the other three, adjusted for multiple comparisons (*88*). Melodic complexity did not predict context (dance, P = 0.79; healing, P = 0.96; love, P = 0.13; lullaby, P = 0.35). However, rhythmic complexity did distinguish dance songs (which were more rhythmically complex, P = 0.01) and lullabies (which were less rhythmically complex, P = 0.03) from other songs; it did not distinguish healing or love songs (Ps > 0.99). When we adjusted these analyses to account for across-region variability, the results were comparable (Text S2.4.2). Thus, although musical content systematically varies in two ways across cultures, this variation is mostly unrelated to the behavioral contexts of the songs, perhaps because complexity captures distinctions that are salient to music analysts but not strongly evocative of particular moods or themes among the singers and listeners themselves.

Melodic and rhythmic bigrams are distributed according to power laws

Many phenomena in the social and biological sciences are characterized by Zipf's law (*108*), in which the probability of an event is inversely proportional to its rank in frequency, an example of a power-law distribution (in the Zipfian case, the exponent is 1). Power-law distributions (as opposed to, say, the geometric distribution) have two key properties: A small number of highly frequent events account for the majority of observations, and there are a large number of individually improbable events whose probability falls off slowly in a thick tail (*109*).

In language, for example, a few words appear with very high frequency, such as pronouns, while a great many are rare, such as the names of species of trees, but any sample will nonetheless tend to contain several rare words (*110*). A similar pattern is found in the distribution of colors among paintings in a given period of art history (*111*). In music, Zipf's law has been observed in the melodic intervals of Bach, Chopin, Debussy, Mendelssohn, Mozart, and Schoenberg (*112*–*116*); in the loudness and pitch fluctuations in Scott Joplin piano rags (*117*); in the harmonies (*118*–*120*) and rhythms of classical music (*121*); and, as Zipf himself noted, in melodies composed by Mozart, Chopin, Irving Berlin, and Jerome Kern (*108*).

We tested whether the presence of power-law distributions is a property of music worldwide by tallying relative melodic bigrams (the number of semitones separating each pair of successive notes) and relative rhythmic bigrams (the ratio of the durations of each pair of successive notes) for all NHS Discography transcriptions (Text S2.4.3). The bigrams overlapped, with the second note of one bigram also serving as the first note of the next.

We found that both the melodic and rhythmic bigram distributions followed power laws

(Fig. 8), and this finding held worldwide: The fit between the observed bigrams and the best-fitting power function was high within each region (melodic bigrams: median R^2 = 0.97, range 0.92 to 0.99; rhythmic bigrams: median R^2 = 0.98, range 0.88 to 0.99). The most prevalent bigrams were the simplest. Among the melodic bigrams (Fig. 8A), three small intervals (unison, major second, and minor third) accounted for 73% of the bigrams; the tritone (six semitones) was the rarest, accounting for only 0.2%. The prevalence of these bigrams is significant: Using only unisons, major seconds, and minor thirds, one can construct any melody in a pentatonic scale, a scale found in many cultures (*122*). Among the rhythmic bigrams (Fig. 8B), three patterns with simple integer ratios (1:1, 2:1, and 3:1) accounted for 86% of observed bigrams, whereas a large and eclectic group of ratios (e.g., 7:3, 11:2) accounted for fewer than 1%. The distribution is thus consistent with earlier findings that rhythmic patterns with simple integer ratios appear to be universal (*123*). The full lists of bigrams, with their cumulative frequencies, are in tables S28 and S29.

These results suggest that power-law distributions in music are a human universal (at least in the four genres studied here), with songs dominated by small melodic intervals and simple rhythmic ratios and enriched with many rare but larger and more complex ones. Because the specification of a power law is sensitive to sampling error in the tail of the distribution (*124*), and because many generative processes can give rise to a power-law distribution (*125*), we cannot identify a single explanation. Among the possibilities are that control of the vocal tract is biased toward small jumps in pitch that minimize effort, that auditory analysis is biased toward tracking similar sounds that are likely to be emitted by a single sound-maker, that composers tend to add notes to a melody that are similar to ones already contained in it, and that human aesthetic reactions are engaged by stimuli that are power law–distributed, which makes them neither too monotonous nor too chaotic (*116*, *126*, *127*)—"inevitable and yet surprising," as the music of Bach has been described (*128*).

A new science of music

The challenge in understanding music has always been to reconcile its universality with its diversity. Even Longfellow, who declared music to be humanity's universal language, celebrated the many forms it could take: "The peasant of the North … sings the traditionary ballad to his children … the muleteer of Spain carols with the early lark … The vintager of Sicily has his evening hymn; the fisherman of Naples his boat-song; the gondolier of Venice his midnight serenade" (*1*). Conversely, even an ethnomusicologist skeptical of universals in music conceded that "most people make it" (*36*). Music is universal but clearly takes on different forms in different cultures. To go beyond these unexceptionable observations and understand exactly what is universal about music, while circumventing the biases inherent in opportunistic observations, we assembled databases that combine the empirical richness of the ethnographic and musicological record with the tools of computational social science.

The findings allow the following conclusions: Music exists in every society, varies more within than between societies, and has acoustic features that are systematically (albeit probabilistically) related to the behaviors of singers and listeners. At the same time, music is not a fixed biological response with a single, prototypical adaptive function such as mating, group bonding, or infant care: It varies substantially in melodic and rhythmic complexity and is produced worldwide in at least 14 behavioral contexts that vary in formality, arousal, and religiosity. But music does appear to be tied to identifiable perceptual, cognitive, and affective faculties, including language (all societies put words to their songs), motor control (people in all societies dance), auditory analysis (all musical systems have some signatures of tonality), and aesthetics (their melodies and rhythms are balanced between monotony and chaos).

Methods summary

To build the NHS Ethnography, we extracted descriptions of singing from the Probability Sample File by searching the database for text that was tagged with the topic MUSIC and that included at least one of 10 keywords that singled out vocal music (e.g., "singers," "song," "lullaby") (Text S1.1). This search yielded 4709 descriptions of singing (490,615 words) drawn from 493 documents (median 49 descriptions per society). We manually annotated each description with 66 variables to comprehensively capture the behaviors reported by ethnographers (e.g., age of the singer, duration of the song). We also attached metadata about each paragraph (e.g., document publication data; tagged nonmusical topics) using a matching algorithm that located the source paragraphs from which the description of the song was extracted. See Text S1.1 for full details on corpus construction, tables S1 to S6 for annotation types, and table S12 for a list of societies and locations.

Song events from all the societies were aggregated into a single dataset, without indicators of the society they came from. The range of possible missing values was filled in using a Markov chain Monte Carlo procedure that assumes that their absence reflects conditionally random omission with probabilities related to the features that the ethnographer did record, such as the age and sex of the singer or the size of the audience (Text S2.1). For the dimensionality reduction, we used an optimal singular value thresholding criterion (*129*) to determine the number of dimensions to analyze, which we then interpreted by three techniques: examining annotations that load highly on each dimension; searching for examples at extreme locations in the space and examining their content; and testing whether known song types formed distinct clusters in the latent space (e.g., dance songs versus healing songs; see Fig. 2).

To build the NHS Discography, and to ensure that the sample of recordings from each genre is representative of human societies, we located field recordings of dance songs, lullabies, healing songs, and love songs using a geographic stratification approach similar to that of the NHS Ethnography—namely, by drawing one recording representing each behavioral context from each of 30 regions. We chose songs according to predetermined criteria (table S21), studying recordings' liner notes and the supporting ethnographic text without listening to the recordings. When more than one suitable recording was available, we selected one at random. See Text S1.1 for details on corpus construction, tables S1 and S7 to S11 for annotation types, and table S22 for a list of societies and locations.

For analyses of the universality of musical forms, we studied each of the four representations individually (machine summaries, naïve listener ratings, expert listener ratings, and features extracted from manual transcriptions), along with a combination of the expert listener and manual transcription data, which excluded many "contextual" features of the audio recordings (e.g., the sound of an infant crying during a lullaby). For the explorations of the structure of musical forms, we studied the manual transcriptions of songs and also used the Bayesian principal components analysis technique (described above) on the combined expert + transcription data summarizing NHS Discography songs.

Both the NHS Ethnography and NHS Discography can be explored interactively at http://themusiclab.org/nhsplots.

REFERENCES AND NOTES

1. H. W. Longfellow, *Outre-mer: A Pilgrimage Beyond the Sea* (Harper, 1835).
2. L. Bernstein, *The Unanswered Question: Six Talks at Harvard* (Harvard Univ. Press, 2002).
3. H. Honing, C. ten Cate, I. Peretz, S. E. Trehub, Without it no music: Cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. B* **370**, 20140088 (2015). doi: 10.1098/rstb.2014.0088; pmid: 25646511
4. S. A. Mehr, M. M. Krasnow, Parent-offspring conflict and the evolution of infant-directed song. *Evol. Hum. Behav.* **38**, 674–684 (2017). doi: 10.1016/j.evolhumbehav.2016.12.005
5. E. H. Hagen, G. A. Bryant, Music and dance as a coalition signaling system. *Hum. Nat.* **14**, 21–51 (2003). doi: 10.1007/s12110-003-1015-z; pmid: 26189987
6. A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, 1990).

7. A. S. Bregman, S. Pinker, Auditory streaming and the building of timbre. *Can. J. Psychol.* **32**, 19–31 (1978). doi: 10.1037/h0081664; pmid: 728845
8. S. Pinker, *How the Mind Works* (Norton, 1997).
9. L. J. Trainor, The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philos. Trans. R. Soc. B* **370**, 20140089 (2015). doi: 10.1098/rstb.2014.0089; pmid: 25646512
10. A. Lomax, *Folk Song Style and Culture* (American Association for the Advancement of Science, 1968).
11. A. P. Merriam, *The Anthropology of Music* (Northwestern Univ. Press, 1964).
12. B. Nettl, *The Study of Ethnomusicology: Thirty-Three Discussions* (Univ. of Illinois Press, 2015).
13. N. J. Conard, M. Malina, S. C. Münzel, New flutes document the earliest musical tradition in southwestern Germany. *Nature* **460**, 737–740 (2009). doi: 10.1038/nature08169; pmid: 19553935
14. N. Martínez-Molina, E. Mas-Herrero, A. Rodríguez-Fornells, R. J. Zatorre, J. Marco-Pallarés, Neural correlates of specific musical anhedonia. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7337–E7345 (2016). doi: 10.1073/pnas.1611211113; pmid: 27799544
15. A. D. Patel, Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681 (2003). doi: 10.1038/nn1082; pmid: 12830158
16. D. Perani *et al*., Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4758–4763 (2010). doi: 10.1073/pnas.0909074107; pmid: 20176953
17. J. H. McDermott, A. F. Schultz, E. A. Undurraga, R. A. Godoy, Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* **535**, 547–550 (2016). doi: 10.1038/nature18635; pmid: 27409816
18. B. Nettl, in *The Origins of Music* (MIT Press, 2000), pp. 463–472.
19. J. Blacking, Can musical universals be heard? *World Music* **19**, 14–22 (1977).
20. F. Harrison, Universals in music: Towards a methodology of comparative research. *World Music* **19**, 30–36 (1977).
21. G. Herzog, Music's dialects: A non-universal language. *Indep. J. Columbia Univ.* **6**(10), 1–2 (1939).
22. M. Hood, *The Ethnomusicologist* (McGraw-Hill, 1971).
23. L. B. Meyer, Universalism and relativism in the study of ethnic music. *Ethnomusicology* **4**, 49–54 (1960). doi: 10.2307/924262
24. S. Feld, Sound structure as social structure. *Ethnomusicology* **28**, 383–409 (1984). doi: 10.2307/851232
25. M. Hood, in *Musicology*, F. L. Harrison, M. Hood, C. V. Palisca, Eds. (Prentice-Hall, 1963), pp. 217–239.
26. M. Roseman, The social structuring of sound: The Temiar of peninsular Malaysia. *Ethnomusicology* **28**, 411–445 (1984). doi: 10.2307/851233
27. S. Feld, *Sound and Sentiment: Birds, Weeping, Poetics, and Song in Kaluli Expression* (Duke Univ. Press, 2012).
28. N. Harkness, *Songs of Seoul: An Ethnography of Voice and Voicing in Christian South Korea* (Univ. of California Press, 2014).
29. T. Rose, Orality and technology: Rap music and Afro-American cultural resistance. *Pop. Music Soc.* **13**, 35–44 (1989). doi: 10.1080/03007768908591371
30. S. Feld, A. A. Fox, Music and language. *Annu. Rev. Anthropol.* **23**, 25–53 (1994). doi: 10.1146/annurev.an.23.100194.000325
31. T. Ellingson, in *Ethnomusicology*, H. Myers, Ed. (Norton, 1992), pp. 110–152.
32. T. F. Johnston, The cultural role of Tsonga beer-drink music. *Yearb. Int. Folk Music Counc.* **5**, 132–155 (1973). doi: 10.2307/767499
33. A. Rehding, The quest for the origins of music in Germany circa 1900. *J. Am. Musicol. Soc.* **53**, 345–385 (2000). doi: 10.2307/832011
34. A. K. Rasmussen, Response to "Form and function in human song". *Soc. Ethnomusicol. Newsl.* **52**, 7 (2018).
35. We conducted a survey of academics to solicit opinions about the universality of music. The overall pattern of results from music scholars was consistent with List's claim that music is characterized by very few universals. For instance, in response to the question "Do you think that music is mostly shaped by culture, or do you think that music is mostly shaped by a universal human nature?", the majority of music scholars responded in the "Music is mostly shaped by culture" half of the scale (ethnomusicologists, 71%; music theorists, 68%; other musical disciplines, 62%). See Text S1.4.1 for full details.
36. G. List, On the non-universality of musical perspectives. *Ethnomusicology* **15**, 399–402 (1971). doi: 10.2307/850640
37. N. A. Chomsky, *Language and Mind* (Harcourt Brace, 1968).
38. M. H. Christiansen, C. T. Collins, S. Edelman, *Language Universals* (Oxford Univ. Press, 2009).
39. P. Boyer, *Religion Explained: The Evolutionary Origins of Religious Thought* (Basic Books, 2007).
40. M. Singh, The cultural evolution of shamanism. *Behav. Brain Sci.* **41**, e66 (2018). doi: 10.1017/S0140525X17001893; pmid: 28679454
41. R. Sosis, C. Alcorta, Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evol. Anthropol. Issues News Rev.* **12**, 264–274 (2003). doi: 10.1002/evan.10120
42. D. M. Buss, Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* **12**, 1–14 (1989). doi: 10.1017/S0140525X00023992
43. B. Chapais, Complex kinship patterns as evolutionary constructions, and the origins of sociocultural universals. *Curr. Anthropol.* **55**, 751–783 (2014). doi: 10.1086/678972
44. A. P. Fiske, *Structures of Social Life: The Four Elementary Forms of Human Relations: Communal Sharing, Authority Ranking, Equality Matching, Market Pricing* (Free Press, 1991).
45. T. S. Rai, A. P. Fiske, Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychol. Rev.* **118**, 57–75 (2011). doi: 10.1037/a0021867; pmid: 21244187
46. O. S. Curry, D. A. Mullins, H. Whitehouse, Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Curr. Anthropol.* **60**, 47–69 (2019). doi: 10.1086/701478
47. J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Penguin, 2013).
48. R. W. Wrangham, L. Glowacki, Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: Evaluating the chimpanzee model. *Hum. Nat.* **23**, 5–29 (2012). doi: 10.1007/s12110-012-9132-1; pmid: 22388773
49. S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking, 2011).
50. A. P. Fiske, T. S. Rai, *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships* (Cambridge Univ. Press, 2015).
51. L. Aarøe, M. B. Petersen, K. Arceneaux, The behavioral immune system shapes political intuitions: Why and how individual differences in disgust sensitivity underlie opposition to immigration. *Am. Polit. Sci. Rev.* **111**, 277–294 (2017). doi: 10.1017/S0003055416000770
52. P. Boyer, M. B. Petersen, Folk-economic beliefs: An evolutionary cognitive model. *Behav. Brain Sci.* **41**, e158 (2018). doi: 10.1017/S0140525X17001960; pmid: 29022516
53. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8987–8992 (2015). doi: 10.1073/pnas.1414495112; pmid: 26124105
54. S. A. Mehr, M. Singh, H. York, L. Glowacki, M. M. Krasnow, Form and function in human song. *Curr. Biol.* **28**, 356–368.e5 (2018). doi: 10.1016/j.cub.2017.12.042; pmid: 29395919
55. T. Fritz *et al*., Universal recognition of three basic emotions in music. *Curr. Biol.* **19**, 573–576 (2009). doi: 10.1016/j.cub.2009.02.058; pmid: 19303300
56. 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). doi: 10.1073/pnas.1209023110; pmid: 23248314
57. W. T. Fitch, The biology and evolution of music: A comparative perspective. *Cognition* **100**, 173–215 (2006). doi: 10.1016/j.cognition.2005.11.009; pmid: 16412411
58. A. Lomax, Universals in song. *World Music* **19**, 117–129 (1977).
59. D. E. Brown, *Human Universals* (Temple Univ. Press, 1991).
60. S. Brown, J. Jordania, Universals in the world's musics. *Psychol. Music* **41**, 229–248 (2013). doi: 10.1177/0305735611425896
61. Human Relations Area Files Inc., eHRAF World Cultures Database; http://ehrafworldcultures.yale.edu/.
62. G. P. Murdock, C. S. Ford, A. E. Hudson, R. Kennedy, L. W. Simmons, J. W. M. Whiting, *Outline of Cultural Materials* (Human Relations Area Files Inc., New Haven, CT, 2008).
63. P. Austerlitz, *Merenge: Dominican Music and Dominican Identity* (Temple Univ. Press, 2007).
64. C. Irgens-Møller, *Music of the Hazara: An Investigation of the Field Recordings of Klaus Ferdinand 1954–1955* (Moesgård Museum, Denmark, 2007).
65. B. D. Koen, *Devotional Music and Healing in Badakhshan, Tajikistan: Preventive and Curative Practices* (UMI Dissertation Services, Ann Arbor, MI, 2005).
66. B. D. Koen, *Beyond the Roof of the World: Music, Prayer, and Healing in the Pamir Mountains* (Oxford Univ. Press, 2011).
67. A. Youssefzadeh, The situation of music in Iran since the revolution: The role of official organizations. *Br. J. Ethnomusicol.* **9**, 35–61 (2000). doi: 10.1080/09681220008567300
68. S. Zeranska-Kominek, The classification of repertoire in Turkmen traditional music. *Asian Music* **21**, 91–109 (1990). doi: 10.2307/834113
69. A. D. Patel, *Music, Language, and the Brain* (Oxford Univ. Press, 2008).
70. D. P. McAllester, Some thoughts on "universals" in world music. *Ethnomusicology* **15**, 379–380 (1971). doi: 10.2307/850637
71. A. P. Merriam, in *Cross-Cultural Perspectives on Music: Essays in Memory of Miczyslaw Kolinski*, R. Falck, T. Rice, M. Kolinski, Eds. (Univ. of Toronto Press, 1982), pp. 174–189.
72. D. L. Harwood, Universals in music: A perspective from cognitive psychology. *Ethnomusicology* **20**, 521–533 (1976). doi: 10.2307/851047
73. F. Lerdahl, R. Jackendoff, *A Generative Theory of Tonal Music* (MIT Press, 1983).
74. Human Relations Area Files Inc., The HRAF quality control sample universe. *Behav. Sci. Notes* **2**, 81–88 (1967). doi: 10.1177/106939716700200203
75. R. O. Lagacé, The HRAF probability sample: Retrospect and prospect. *Behav. Sci. Res.* **14**, 211–229 (1979). doi: 10.1177/106939717901400304
76. R. Naroll, The proposed HRAF probability sample. *Behav. Sci. Notes* **2**, 70–80 (1967). doi: 10.1177/106939716700200202
77. B. S. Hewlett, S. Winn, Allomaternal nursing in humans. *Curr. Anthropol.* **55**, 200–219 (2014). doi: 10.1086/675657; pmid: 24991682
78. Q. D. Atkinson, H. Whitehouse, The cultural morphospace of ritual form. *Evol. Hum. Behav.* **32**, 50–62 (2011). doi: 10.1016/j.evolhumbehav.2010.09.002
79. C. R. Ember, The relative decline in women's contribution to agriculture with intensification. *Am. Anthropol.* **85**, 285–304 (1983). doi: 10.1525/aa.1983.85.2.02a00020
80. D. M. T. Fessler, A. C. Pisor, C. D. Navarrete, Negatively-biased credulity and the cultural evolution of beliefs. *PLOS ONE* **9**, e95167 (2014). doi: 10.1371/journal.pone.0095167; pmid: 24736596
81. B. R. Huber, W. L. Breedlove, Evolutionary theory, kinship, and childbirth in cross-cultural perspective. *Cross-Cultural Res.* **41**, 196–219 (2007). doi: 10.1177/1069397106298261
82. D. Levinson, Physical punishment of children and wifebeating in cross-cultural perspective. *Child Abuse Negl.* **5**, 193–195 (1981). doi: 10.1016/0145-2134(81)90040-5
83. M. Singh, Magic, explanations, and evil: On the origins and design of witches and sorcerers. SocArXiv (2019). doi: 10.31235/osf.io/pbwc7
84. M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis. *J. R. Stat. Soc. B* **61**, 611–622 (1999). doi: 10.1111/1467-9868.00196
85. R. C. Lewontin, in *Evolutionary Biology*, T. Dobzhansky, M. K. Hecht, W. C. Steer, Eds. (Appleton-Century-Crofts, 1972), pp. 391–398.
86. T. Rzeszutek, P. E. Savage, S. Brown, The structure of cross-cultural musical diversity. *Proc. R. Soc. B* **279**, 1606–1612 (2012). doi: 10.1098/rspb.2011.1750; pmid: 22072606
87. Princeton University, WordNet: A lexical database for English (2010); http://wordnet.princeton.edu.
88. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
89. R. Karsten, *The Religion of the Samek: Ancient Beliefs and Cults of the Scandinavian and Finnish Lapps* (Brill, 1955).
90. M. J. Murphy, "A sprain." In Manuscript 1503 (serial #QBA7800250), pp. 237–238 (1957). Available from the Manuscript Collection of the National Folklore Collection, University College Dublin.
91. S. Feld, Linguistic models in ethnomusicology. *Ethnomusicology* **18**, 197–217 (1974). doi: 10.2307/850579
92. S. Arom, *African Polyphony and Polyrhythm: Musical Structure and Methodology* (Cambridge Univ. Press, 2004).

93. B. Nettl, *Theory and Method in Ethnomusicology* (Collier-Macmillan, 1964).
94. J. Friedman, T. Hastie, R. Tibshirani, Lasso and Elastic-Net Regularized Generalized Linear Models. Rpackage Version 2.0-5 (2016).
95. C. Nadeau, Y. Bengio, Inference for the generalization error. *Mach. Learn.* **52**, 239–281 (2003). doi: 10.1023/A:1024068626366
96. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996). doi: 10.1111/j.2517-6161.1996.tb02080.x
97. S. E. Trehub, A. M. Unyk, L. J. Trainor, Adults identify infant-directed music across cultures. *Infant Behav. Dev.* **16**, 193–211 (1993). doi: 10.1016/0163-6383(93)80017-3
98. A. Barker, *Greek Musical Writings: Harmonic and Acoustic Theory* (Cambridge Univ. Press, 2004).
99. C. L. Krumhansl, The Cognition of Tonality – as We Know it Today. *J. New Music Res.* **33**, 253–268 (2004). doi: 10.1080/0929821042000317831
100. J. H. McDermott, A. J. Oxenham, Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.* **18**, 452–463 (2008). doi: 10.1016/j.conb.2008.09.005; pmid: 18824100
101. R. Jackendoff, F. Lerdahl, The capacity for music: What is it, and what's special about it? *Cognition* **100**, 33–72 (2006). doi: 10.1016/j.cognition.2005.11.005; pmid: 16384553
102. M. A. Castellano, J. J. Bharucha, C. L. Krumhansl, Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* **113**, 394–412 (1984). doi: 10.1037/0096-3445.113.3.394; pmid: 6237169
103. C. L. Krumhansl *et al*., Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition* **76**, 13–58 (2000). doi: 10.1016/S0010-0277(00)00068-8; pmid: 10822042
104. H. von Helmholtz, *The Sensations of Tone as a Physiological Basis for the Theory of Music* (Longmans, 1885).
105. D. Cooke, *The Language of Music* (Oxford Univ. Press, 2001).
106. J. A. Hartigan, P. M. Hartigan, The Dip Test of Unimodality. *Ann. Stat.* **13**, 70–84 (1985). doi: 10.1214/aos/1176346577
107. C. L. Krumhansl, *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, 2001).
108. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley, 1949).
109. H. Baayen, *Word Frequency Distributions* (Kluwer Academic, 2001).
110. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014). doi: 10.3758/s13423-014-0585-6; pmid: 24664880
111. D. Kim, S.-W. Son, H. Jeong, Large-scale quantitative analysis of painting arts. *Sci. Rep.* **4**, 7370 (2014). doi: 10.1073/pnas.87.3.938; pmid: 11607061
112. K. J. Hsü, A. J. Hsü, Fractal geometry of music. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 938–941 (1990). doi: 10.1073/pnas.87.3.938; pmid: 11607061
113. D. H. Zanette, Zipf's law and the creation of musical context. *Music. Sci.* **10**, 3–18 (2006). doi: 10.1177/102986490601000101
114. H. J. Brothers, Intervallic scaling in the Bach cello suites. *Fractals* **17**, 537–545 (2009). doi: 10.1142/S0218348X09004521
115. L. Liu, J. Wei, H. Zhang, J. Xin, J. Huang, A statistical physics view of pitch fluctuations in the classical music from Bach to Chopin: Evidence for scaling. *PLOS ONE* **8**, e58710 (2013). doi: 10.1371/journal.pone.0058710; pmid: 23544047
116. B. Manaris *et al*., Zipf's law, music classification, and aesthetics. *Comput. Music J.* **29**, 55–69 (2005). doi: 10.1162/comj.2005.29.1.55
117. R. F. Voss, J. Clarke, '1/*f* noise' in music and speech. *Nature* **258**, 317–318 (1975). doi: 10.1038/258317a0
118. M. Rohrmeier, I. Cross, Statistical Properties of Tonal Harmony in Bach's Chorales. in *Proceedings of the 10th International Conference on Music Perception and Cognition* (2008), p. 9.
119. F. C. Moss, M. Neuwirth, D. Harasim, M. Rohrmeier, Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLOS ONE* **14**, e0217242 (2019). doi: 10.1371/journal.pone.0217242; pmid: 31170188
120. M. Beltrán del Río, G. Cocho, G. G. Naumis, Universality in the tail of musical note rank distribution. *Physica A* **387**, 5552–5560 (2008). doi: 10.1016/j.physa.2008.05.031
121. D. J. Levitin, P. Chordia, V. Menon, Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3716–3720 (2012). doi: 10.1073/pnas.1113828109; pmid: 22355125
122. T. Van Khe, Is the pentatonic universal? A few reflections on pentatonism. *World Music* **19**, 76–84 (1977).
123. N. Jacoby, J. H. McDermott, Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* **27**, 359–370 (2017). doi: 10.1016/j.cub.2016.12.031; pmid: 28065607
124. A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-Law Distributions in Empirical Data. *SIAM Rev.* **51**, 661–703 (2009). doi: 10.1137/070710111
125. M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2004). doi: 10.1080/15427951.2004.10129088
126. G. D. Birkhoff, *Aesthetic Measure* (Harvard Univ. Press, 2013).
127. B. Manaris, P. Roos, D. Krehbiel, T. Zalonis, J. R. Armstrong, in *Music Data Mining*, T. Li, M. Ogihara, G. Tzanetakis, Eds.(CRC Press, 2012), chapter 6.
128. M. R. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (Dover, 2009).
129. D. Donoho, M. Gavish, Minimax risk of matrix denoising by singular value thresholding. *Ann. Stat.* **42**, 2413–2440 (2014). doi: 10.1214/14-AOS1257

ACKNOWLEDGMENTS

We thank the hundreds of anthropologists and ethnomusicologists whose work forms the source material for all our analyses; the countless people whose music those scholars reported on; and the research assistants who contributed to the creation of the Natural History of Song corpora and to this research, here listed alphabetically: Z. Ahmad, P. Ammirante, R. Beaudoin, J. Bellissimo, A. Bergson, M. Bertolo, M. Bertuccelli, A. Bitran, S. Bourdaghs, J. Brown, L. Chen, C. Colletti, L. Crowe, K. Czachorowski, L. Dinetan, K. Emery, D. Fratina, E. Galm, S. Gomez, Y.-H. Hung, C. Jones, S. Joseph, J. Kangatharan, A. Keomurjian, H. J. Kim, S. Lakin, M. Laroussini, T. Lee, H. Lee-Rubin, C. Leff, K. Lopez, K. Luk, E. Lustig, V. Malawey, C. McMann, M. Montagnese, P. Moro, N. Okwelogu, T. Ozawa, C. Palfy, J. Palmer, A. Paz, L. Poeppel, A. Ratajska, E. Regan, A. Reid, R. Sagar, P. Savage, G. Shank, S. Sharp, E. Sierra, D. Tamaroff, I. Tan, C. Tripoli, K. Tutrone, A. Wang, M. Weigel, J. Weiner, R. Weissman, A. Xiao, F. Xing, K. Yong, H. York, and J. Youngers. We also thank C. Ember and M. Fischer for providing additional data from the Human Relations Area Files, and for their assistance using those data; S. Adams, P. Laurence, P. O'Brien, A. Wilson, the staff at the Archive of World Music at Loeb Music Library (Harvard University), and M. Graf and the staff at the Archives of Traditional Music (Indiana University) for assistance with locating and digitizing audio recordings; B. Hillers for assistance with information concerning traditional Gaelic music; D. Niles, S. Wadley, and H. Wild for contributing recordings from their personal collections; S. Collins for producing the NHS Ethnography validity annotations; M. Walter for assistance with digital processing of transcriptions; J. Hulbert and R. Clarida for assistance with copyright issues and materials sharing; V. Kuchinov for developing the interactive visualizations; S. Deviche for contributing illustrations; and the Dana Foundation, whose program "Arts and Cognition" led in part to the development of this research. Last, we thank A. Rehding, G. Bryant, E. Hagen, H. Gardner, E. Spelke, M. Tenzer, G. King, J. Nemirow, J. Kagan, and A. Martin for their feedback, ideas, and intellectual support of this work. **Funding:** Supported by the Harvard Data Science Initiative (S.A.M.); the National Institutes of Health Director's Early Independence Award DP5OD024566 (S.A.M.); the Harvard Graduate School of Education/Harvard University Presidential Scholarship (S.A.M.); the Harvard University Department of Psychology (S.A.M. and M.M.K.); a Harvard University Mind/Brain/Behavior Interfaculty Initiative Graduate Student Award (S.A.M. and M.S.); the National Science Foundation Graduate Research Fellowship Program (M.S.); the Microsoft Research postdoctoral fellowship program (D.K.); the Washington University Faculty of Arts and Sciences Dean's Office (C.L.); the Columbia University Center for Science and Society (N.J.); the Natural Sciences and Engineering Research Council of Canada (T.J.O.); Fonds de Recherche du Québec Société et Culture (T.J.O.); and ANR Labex IAST (L.G.). **Author contributions:** S.A.M., M.S., and L.G. created and direct the Natural History of Song project; they oversaw all aspects of this work, including the design and development of the corpora. S.P., M.M.K., and T.J.O. contributed to the conceptual foundation. D.K. designed and implemented all analyses, with support from S.A.M. and C.L. S.A.M., D.K., and M.S. designed the static figures and S.A.M. and D.K. created them. C.L. and S.A.M. designed the interactive figures and supervised their development. S.A.M. recruited and managed all staff, who collected, annotated, processed, and corrected data and metadata. S.A.M., D.M.K., and D.P.-J. transcribed the NHS Discography into music notation. S.A., A.A.E., E.J.H., and R.M.H. provided key support by contributing to annotations, background research, and project management. S.A.M., J.K.H., M.V.J., J.S., and C.M.B. designed and implemented the online experiment at http://themusiclab.org. N.J. assisted with web scraping, music information retrieval, and initial analyses. S.A.M., M.S., and L.G. designed the overall structure of the manuscript; S.A.M., M.S., and S.P. led the writing; and all authors edited it collaboratively. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All Natural History of Song data and materials are publicly archived at http://osf.io/jmv3q, with the exception of the full audio recordings in the NHS Discography, which are available via the Harvard Dataverse at https://doi.org/10.7910/DVN/SESAO1. All analysis scripts are available at http://github.com/themusiclab/nhs. Human Relations Area Files data and the eHRAF World Cultures database are available via licensing agreement at http://ehrafworldcultures.yale.edu; the document- and paragraph-wise word histograms from the Probability Sample File were provided by the Human Relations Area Files under a Data Use Agreement. The Global Summary of the Year corpus is maintained by the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and is publicly available at www.ncei.noaa.gov/data/gsoy/.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/eaax0868/suppl/DC1
Supplementary Text
Figs. S1 to S15
Tables S1 to S37
References (*130–147*)

1 March 2019; accepted 24 October 2019
10.1126/science.aax0868

Science

Universality and diversity in human song

Samuel A. Mehr, Manvir Singh, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, S. Atwood, Christopher Lucas, Nori Jacoby, Alena A. Egner, Erin J. Hopkins, Rhea M. Howard, Joshua K. Hartshorne, Mariela V. Jennings, Jan Simson, Constance M. Bainbridge, Steven Pinker, Timothy J. O'Donnell, Max M. Krasnow and Luke Glowacki

Science **366** (6468), eaax0868.
DOI: 10.1126/science.aax0868

Cross-cultural analysis of song

It is unclear whether there are universal patterns to music across cultures. Mehr *et al.* examined ethnographic data and observed music in every society sampled (see the Perspective by Fitch and Popescu). For songs specifically, three dimensions characterize more than 25% of the performances studied: formality of the performance, arousal level, and religiosity. There is more variation in musical behavior within societies than between societies, and societies show similar levels of within-society variation in musical behavior. At the same time, one-third of societies significantly differ from average for any given dimension, and half of all societies differ from average on at least one dimension, indicating variability across cultures.

Science, this issue p. eaax0868; see also p. 944

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.