doi:10.1038/nature18635

# Indifference to dissonance in native Amazonians reveals cultural variation in music perception

Josh H. McDermott[1], Alan F. Schultz[2], Eduardo A. Undurraga[3,4] & Ricardo A. Godoy[3]

**Music is present in every culture, but the degree to which it is shaped by biology remains debated. One widely discussed phenomenon is that some combinations of notes are perceived by Westerners as pleasant, or consonant, whereas others are perceived as unpleasant, or dissonant[1]. The contrast between consonance and dissonance is central to Western music[2,3], and its origins have fascinated scholars since the ancient Greeks[4–10]. Aesthetic responses to consonance are commonly assumed by scientists to have biological roots[11–14], and thus to be universally present in humans[15,16]. Ethnomusicologists[17] and composers[8], in contrast, have argued that consonance is a creation of Western musical culture[6]. The issue has remained unresolved, partly because little is known about the extent of cross-cultural variation in consonance preferences[18]. Here we report experiments with the Tsimane'—a native Amazonian society with minimal exposure to Western culture—and comparison populations in Bolivia and the United States that varied in exposure to Western music. Participants rated the pleasantness of sounds. Despite exhibiting Western-like discrimination abilities and Western-like aesthetic responses to familiar sounds and acoustic roughness, the Tsimane' rated consonant and dissonant chords and vocal harmonies as equally pleasant. By contrast, Bolivian city- and town-dwellers exhibited significant preferences for consonance, albeit to a lesser degree than US residents. The results indicate that consonance preferences can be absent in cultures sufficiently isolated from Western music, and are thus unlikely to reflect innate biases or exposure to harmonic natural sounds. The observed variation in preferences is presumably determined by exposure to musical harmony, suggesting that culture has a dominant role in shaping aesthetic responses to music.**

We conducted two studies to measure consonance preferences in populations with varying exposure to Western music. In Study 1, we measured preferences for sounds in residents of the United States and compared them to three populations in Bolivia: (1) residents of the capital city (La Paz); (2) residents of a rural town (San Borja); and (3) members of a native society of horticulturalist-foragers (the Tsimane') in a remote village (Santa Maria) in the Amazon rainforest. City- and town-dwellers were fluent in Spanish and generally had televisions and radios. By contrast, the Tsimane' were mostly monolingual in their own language, lacked televisions, and had limited access to music via radio[19]. The Tsimane' village lacked electricity and tap water, was inaccessible by road, and could be reached only by canoe (Fig. 1a). Their contact with Western culture was mainly limited to occasional trips to nearby towns. To compare differences between cultures to intra-cultural variation within Westerners, we tested two groups from the United States—one with at least two years of experience playing a musical instrument, and one with at most one year of experience.

The Tsimane' were of particular interest because harmony, polyphony, and group performances are by all accounts absent from their music. This conclusion was suggested by previous recordings and documentation[20] and was substantiated by interviews with Tsimane' musicians (see Methods for a description of Tsimane' music, and Supplementary Audio 1 and 2 for examples of Tsimane' songs). The apparent absence of harmony raised the question of whether they would exhibit aesthetic responses to consonance and dissonance despite not having prior exposure to them.

Participants were presented with sounds over headphones (Fig. 1b) and rated the pleasantness of sounds on a four-point scale. We observed pronounced differences between groups in the ratings of consonant and dissonant chords (Fig. 2a, b). US residents showed strong consonance preferences, as expected from previous studies[21], with musicians showing stronger preferences than non-musicians (synthetic chords: $F(46,1) = 7.44$, $P = 0.009$; sung chords: $F(46,1) = 12.97$, $P = 0.001$). Moreover, Bolivian city- and town-dwellers displayed more modest but statistically significant preferences for consonance ($P < 0.01$ in all cases, $t$-tests). By contrast, the Tsimane' rated consonant and dissonant chords as equally pleasant. This finding held for chords composed of synthetic tones as well as chords composed of recorded notes sung by a vocalist, producing an interaction in both cases between stimulus class and participant group (synthetic: $F(124,4) = 48.43$, $P < 10^{-23}$; sung: $F(126,4) = 26.79$, $P < 10^{-15}$). Tsimane' ratings varied somewhat across chords (synthetic: $F(9,270) = 1.96$, $P = 0.04$; sung: $F(9,288) = 1.93$, $P = 0.05$), with some tendency for larger pitch intervals to be preferred over smaller ones (Extended Data Fig. 1), but without consistent differences between consonant and dissonant chords.

Similar results were obtained when participants rated single tones whose frequencies were manipulated to be harmonic (multiples of a common fundamental frequency) or inharmonic, a distinction believed to differentiate consonant from dissonant chords[4,12,13,21,22] (Fig. 2c). US residents showed a preference for harmonic over inharmonic tones, but the effect was stronger in musicians than in non-musicians ($F(46,2) = 16.12$, $P = 0.0002$), was attenuated in Bolivian city-dwellers, and was undetectable in town-dwellers and the Tsimane', again producing an interaction between stimulus and participant group ($F(157,4) = 22.49$, $P < 10^{-13}$).

To rule out the possibility that participant groups might have varied in their understanding of task instructions, or in their ability to assign pleasantness ratings to sounds, we asked them to rate recordings of emotional vocalizations[22] (Fig. 2d). In this case all groups showed a preference for laughter over gasps ($F(157,1) = 200.48$, $P < 1 \times 10^{-29}$), with no interaction with group ($F(157,4) = 1.09$, $P = 0.36$). Thus, for familiar sounds expected to have a positive or negative association, the Tsimane' exhibited a preference in the expected direction, comparable to that of the other groups.

To test whether the Tsimane' would generally fail to exhibit preferences for unfamiliar sounds, we obtained ratings for synthetic tones that varied in roughness, another quality known to affect aesthetic responses in Westerners[5,7,21–25] (Fig. 2e). Roughness was manipulated by presenting pairs of frequencies to the same or different ears. In

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Anthropology, Baylor University, Waco, Texas 76798, USA. [3]Heller School for Social Policy and Management, Brandeis University, Waltham, Massachusetts 02453, USA. [4]Center for Intercultural and Indigenous Research, Pontificia Universidad Católica de Chile, Santiago, Región Metropolitana 7820436, Chile.

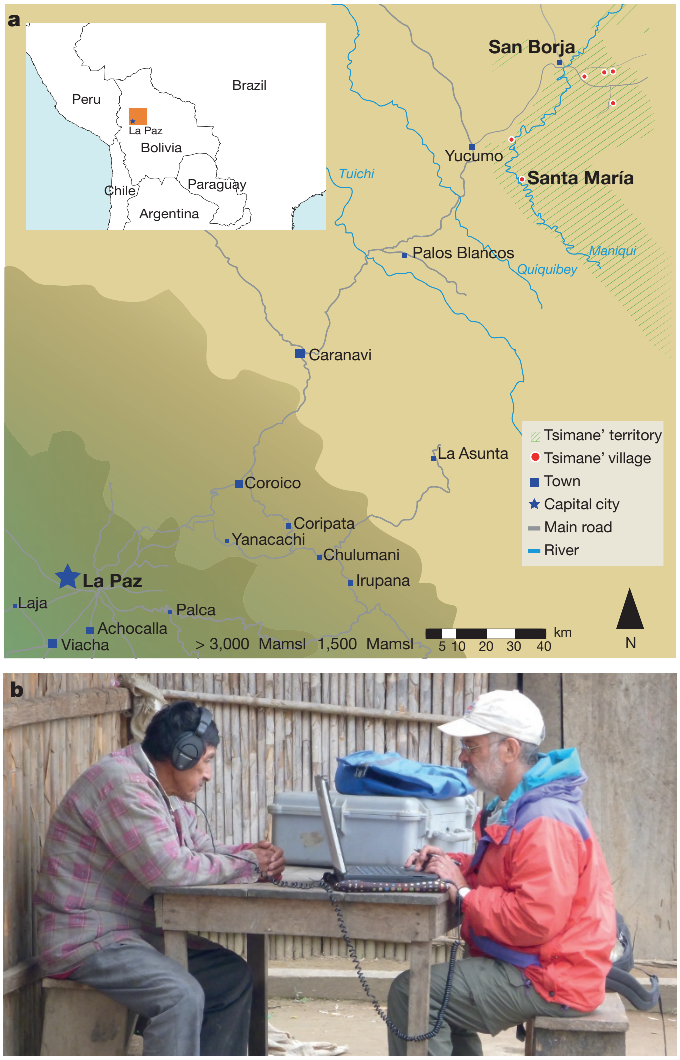



**Figure 1 | Location and setup of experiments. a,** Map of the region from which Bolivian participants were drawn. Participants resided in the Bolivian capital city La Paz, the rural town of San Borja, the Tsimane' village of Santa Maria (Study 1), or Tsimane' villages around San Borja (Study 2; not labelled with names to minimize clutter). Town symbols are approximately proportional in size to town population. Colour of territory denotes elevation. **b,** Sounds were presented over closed headphones via laptop (charged with a gasoline generator when needed). For all but the discrimination experiment (Fig. 4), participants provided a pleasantness rating (with a four-point scale) following each sound.

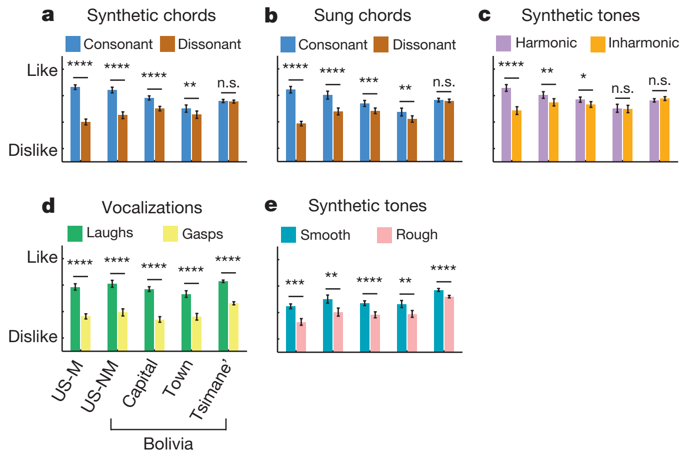



**Figure 2 | Results of Study 1. a–e**, Average pleasantness ratings of sounds from five experiments across five populations: 23 US musicians (US-M), 25 US non-musicians (US-NM), 24 Bolivian city-dwellers (Capital), 26 Bolivian town-dwellers, and 64 Tsimane'. Each experiment featured sounds from two classes expected to differ in pleasantness for US listeners. Chord notes were either synthetic (resembling a piano), or recorded from a trained singer. Chords were conventionally consonant (major third, perfect fourth, perfect fifth, and major triad) or conventionally dissonant (minor second, major second, tritone, major seventh, and augmented triad). Vocalizations were recordings of human laughs and gasps. Synthetic tones varied in harmonicity or roughness. Asterisks denote statistical significance: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; n.s., not significant (two-tailed $t$-tests, uncorrected for multiple comparisons because they were conducted post-hoc, following analysis of variances (ANOVAs) to test for main effects and interactions). Data are mean and s.e.m.

contrast to the results for chords, and for harmonic and inharmonic tones, all groups exhibited a significant preference for smooth over rough tones ($F(157,1) = 85.51$, $P < 10^{-15}$), with no interaction with group ($F(157,4) = 2.04$, $P = 0.09$), even though for US participants this preference was the smallest of those measured.

We replicated and extended these results in Study 2, testing a different group of Tsimane' listeners and a comparison group of musically experienced listeners in the United States. We first tested whether the absence of a consonance preference would extend to harmony in realistic musical material. We recorded Tsimane' vocalists singing Tsimane' song phrases several times, and then pitch-shifted and superimposed the phrases to create harmonies in conventionally consonant or dissonant intervals (see Supplementary Audio 3–10). We obtained pleasantness ratings for these harmonies from US and Tsimane' listeners as in Study 1. For replication purposes, we also conducted analogous experiments with sung and synthetic two-note intervals, and with sung triads.

The results with chords replicated those of Study 1 (Fig. 3a–c): in every case, US listeners preferred consonance to dissonance, while Tsimane' listeners did not, producing interactions between stimulus and participant group (synthetic intervals: $F(72,1) = 42.4$, $P < 10^{-8}$; sung intervals: $F(95,1) = 39.4$, $P < 10^{-7}$; sung triads: $F(95,1) = 17.9$, $P < 10^{-4}$). Ratings of individual chords by Tsimane' listeners again varied across both synthetic and sung intervals (Extended Data Fig. 2), but were primarily explained by interval size, with higher ratings for larger intervals, unlike US listeners (significant correlations between rating and interval size for Tsimane; synthetic: $r = 0.94$, $P < 10^{-5}$; sung: $r = 0.81$, $P = 0.001$; but not for US listeners; synthetic: $r = 0.07$, $P = 0.82$; sung: $r = 0.23$, $P = 0.48$).

Notably, similar results were obtained with harmonies generated from Tsimane' songs (Fig. 3d). Even though the music was foreign to US participants, they reliably judged consonant renditions as more pleasant than dissonant ($t(46) = 6.2$, $P < 10^{-6}$), whereas the Tsimane' did not ($t(49) = 1.2$, $P = 0.22$; stimulus × group interaction: $F(95,1) = 30.2$, $P < 10^{-6}$). Moreover, Tsimane' listeners reliably preferred some of the song excerpts used to generate harmonies over others ($\chi^2(25) = 49.01$, $P = 0.003$; Extended Data Fig. 3). The materials thus elicited consistent aesthetic responses in the Tsimane', but these were not driven by consonance and dissonance. As in Study 1, when presented with recorded vocalizations (Fig. 3e), both Tsimane' and US listeners showed preferences for laughter over gasps ($F(95,1) = 129.4$, $P < 10^{-18}$; no interaction with participant group: $F(95,1) = 1.8$, $P = 0.18$), indicating that Tsimane' listeners could readily perform the task. These results again suggest that the preference for consonance is absent in the Tsimane'.

To explore the effects of harmonicity and roughness found in Study 1, we measured pleasantness ratings for pairs of pure tones (single frequencies) separated by intervals from the chromatic scale (0–8 semitones)[21] (Fig. 3f). This range includes some consonant intervals, for which the tone frequencies approximate harmonics of a common fundamental (and are thus related by simple integer ratios), and some dissonant intervals, for which the tone frequencies are inharmonic. Headphones were used to present the two tones to both ears (diotic presentation) or to separate ears (dichotic presentation), as in the smooth/rough tone experiment from Study 1. Diotic presentation

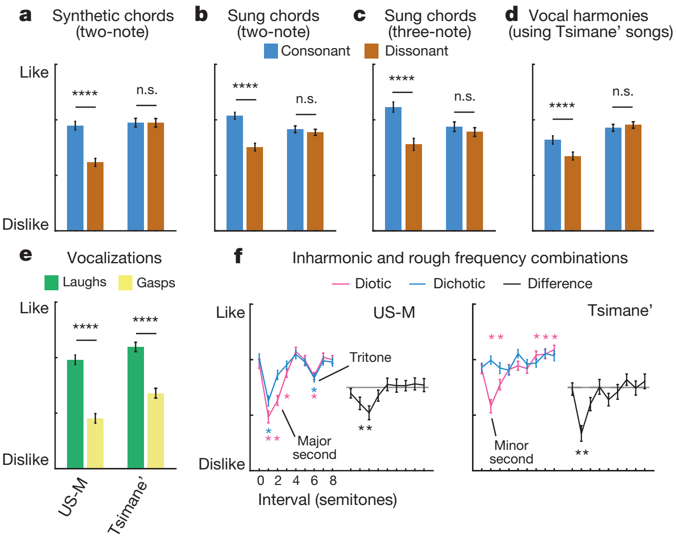



**Figure 3 | Results of preference experiments from Study 2. a–d**, Average pleasantness ratings of consonant and dissonant musical stimuli by 47 US musicians (US-M) and 50 Tsimane'. Two-note chords were intervals from the chromatic scale. Triads were major or augmented. Vocal harmonies (intervals from the chromatic scale) were generated from recordings of Tsimane' vocalists. **e**, Average pleasantness ratings of human laughs and gasps. Asterisks denote statistical significance: $****P < 0.0001$ (two-tailed *t*-tests) (**a**–**e**). **f**, Average pleasantness ratings of concurrent pure tones (single frequencies)[21] presented diotically (both tones to both ears) or dichotically (each tone to a different ear). Blue and pink asterisks indicate ratings significantly different from that of the unison ($P < 0.05$; Wilcoxon signed rank tests, two-tailed, uncorrected for multiple comparisons). Black curves plot the difference between ratings of dichotic and diotic stimuli. Black asterisks denote dichotic–diotic differences significantly different from zero ($P < 0.05$; Wilcoxon signed rank tests, two-tailed, uncorrected). Data are mean and s.e.m.

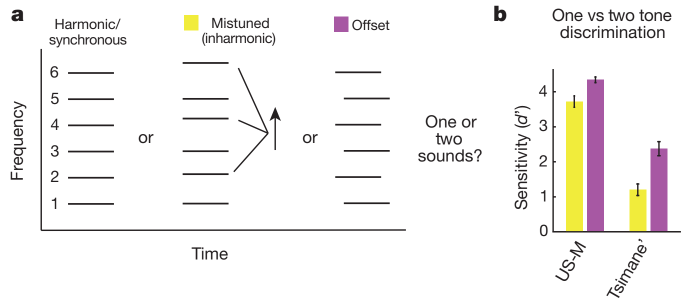



**Figure 4 | Results of discrimination experiment from Study 2. a**, Participants (47 US musicians and 49 Tsimane') heard six harmonics and judged whether they heard one or two sounds. On half of the trials the even harmonics were mistuned by 12% or presented 150 ms earlier in time (harmonics were 750 ms in duration). **b**, Ability to discriminate one vs. two sounds by harmonicity or onset differences. Chance performance corresponds to $d' = 0$. Data are mean and s.e.m.

produces roughness (via beating) when the tones are close in frequency, whereas dichotic presentation minimizes beating while preserving the musical interval (ratio) between frequencies. The stimuli thus dissociated roughness and harmonicity.

Among US residents, pleasantness ratings dipped at dissonant intervals (the minor second, major second, and tritone) for both diotic and dichotic presentation. The variation in pleasantness for dichotic intervals ($\chi^2(7) = 62.7$, $P < 10^{-10}$) implicates factors other than roughness in listeners' aesthetic judgments, potentially that of harmonic frequency relations. However, for small intervals (<3 semitones), diotic ratings were nonetheless lower than dichotic ratings, evident in the difference curve, which varied significantly with interval ($\chi^2(7) = 20.6$, $P = 0.004$; significant differences from zero only for 1 and 2 semitones, $P < 0.05$). This effect is the signature of an aversion to roughness, which is audible only for small intervals (for which frequencies interact via cochlear filtering)[5,7,21].

Tsimane' ratings also suggested an aversion to roughness: diotic intervals were again rated lower than dichotic only for small intervals, producing significant variation in the diotic-dichotic difference ($\chi^2(7) = 29.25$, $P = 0.0001$; significant differences again only for 1 and 2 semitones, $P < 0.05$). By contrast, the influence of harmonicity was absent. Ratings of dichotic intervals did not vary with interval ($\chi^2(7) = 9.52$, $P = 0.22$), and the dip at the tritone was not present for either presentation mode. These results provide further evidence that an aesthetic aversion to roughness is present in the Tsimane', but that the aesthetic response to consonance (and simple integer frequency ratios) is not. This dissociation substantiates the conclusion that the Western notion of dissonance is distinct from acoustic roughness, but closely related to inharmonicity[21,22].

The Tsimane' indifference to consonance and harmonicity raises the question of whether they can hear the underlying stimulus distinction. To test their ability to discern harmonic from inharmonic frequency relations, we conducted a discrimination experiment in which listeners judged whether a stimulus contained one or two sounds (Fig. 4a). Listeners heard the first six harmonics of a synthetic tone. On half of the trials, the even harmonics were mistuned or presented earlier in time. Both manipulations used an established sound segregation cue to cause the even and odd harmonics to be heard as distinct sounds[26,27]. In particular, the mistuning disrupted harmonicity, causing the even harmonics to have a different fundamental frequency than the odd harmonics, producing the perception of a dissonant interval in Western listeners.

US residents reliably identified the intended number of tones in both conditions, as expected (Fig. 4b, left: $t(46) = 22.9$, $P < 10^{-26}$; $t(46) = 54.4$, $P < 10^{-42}$). Tsimane' listeners were worse overall at the task, but were well above chance in both cases (Fig. 4b, right: $t(48) = 7.3$, $P < 10^{-8}$; $t(48) = 11.8$, $P < 10^{-15}$), indicating sensitivity to harmonicity. Analysis of subsets of Tsimane' participants revealed that even those who performed best on this task (mean sensitivity index ($d'$) of 2.0 for the mistuned condition) remained indifferent to dissonance, dissociating perceptual sensitivity and preference (Extended Data Fig. 4).

Taken together, the results suggest that the Tsimane' are able to make pleasantness judgments about sounds. Moreover, for familiar sounds associated with a valence (laughs and gasps), and for synthetic sounds varying in roughness, their judgments resemble those of Westerners. In addition, they are able to hear the acoustic distinctions associated with consonance and dissonance, discriminating harmonic from inharmonic frequencies. The absence of a measurable Tsimane' preference for consonance over dissonance thus appears to specifically reflect differences in their aesthetic response to this contrast.

The cross-cultural variation we observed suggests that consonance preferences are unlikely to be innate, and that they are not driven by exposure to harmonic natural sounds such as vocalizations[14]. Instead, consonance preferences seem to depend on exposure to particular types of music, presumably those that feature consonant harmony. Incidental exposure to Western music could explain the modest preferences evident in Bolivian city- and town-dwellers, along with those previously reported in human infants[28–30]. The lack of harmony in Tsimane' music leaves open the question of what preferences might be observed in individuals exposed to conventionally dissonant intervals, as are prevalent in some non-Western music[17]. But Western music is presumably not unique in fostering preferences for some tone combinations over others.

The belief that consonance is biologically determined has been fuelled by observations that consonance and dissonance differ in acoustic properties such as harmonicity and roughness[4,5,7,9,12,21,22], and as such are not arbitrary categories. However, we found preferences for harmonic tones, for simple integer frequency ratios, and for consonant chords/harmonies to all co-vary with presumptive exposure to

Western culture. Thus, rather than being an inevitable consequence of auditory system biology, it seems that the preferences exhibited by Western listeners for harmonic frequencies arise from exposure to Western music because the musical structures prevalent therein tend to produce harmonic frequencies[21,22]. And although an aversion to roughness is apparently present cross-culturally, it seems to be unrelated to consonance and dissonance, perhaps because musical sounds are in practice not very rough[21,22]. The role of biology in consonance could therefore be limited to constraining what is discriminable, potentially rendering some aesthetic contrasts more readily acquired than others.

The roles of culture and biology in consonance and dissonance have remained unresolved in part because cross-cultural experimental data are rare. Studies of culturally isolated populations are increasingly difficult to conduct due to the diffusion of Western culture around the world, but our results underscore their importance for revealing the diversity of human musical behaviour.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 November 2014; accepted 10 June 2016.**
**Published online 13 July 2016.**

1. Parncutt, R. & Hair, G. Consonance and dissonance in theory and psychology: Disentangling dissonant dichotomies. *J. Interdiscipl. Music Stud*. **5**, 119–166 (2011).
2. Huron, D. Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Percept.* **11,** 289–305 (1994).
3. Bigand, E., Parncutt, R. & Lerdahl, F. Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* **58,** 124–141 (1996).
4. Stumpf, C. *Tonpsychologie* (Verlag S. Hirzel, 1890).
5. von Helmholtz, H. *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik* (F. Vieweg und Sohn, 1863).
6. Lundin, R. W. Toward a cultural theory of consonance. *J. Psychol.* **23,** 45–49 (1947).
7. Plomp, R. & Levelt, W. J. M. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38,** 548–560 (1965).
8. Cazden, N. The definition of consonance and dissonance. *Int. Rev. Aesthet. Soc.* **11,** 123–168 (1980).
9. Sethares, W. A. *Tuning, Timbre, Spectrum, Scale* (Springer, 1999).
10. Tenney, J. *A History of 'Consonance' and 'Dissonance'* (Excelsior Music Publishing Company, 1988).
11. Fishman, Y. I. *et al.* Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *J. Neurophysiol.* **86,** 2761–2788 (2001).
12. Tramo, M. J., Cariani, P. A., Delgutte, B. & Braida, L. D. Neurobiological foundations for the theory of harmony in western tonal music. *Ann. NY Acad. Sci.* **930,** 92–116 (2001).
13. Bidelman, G. M. & Heinz, M. G. Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.* **130,** 1488–1502 (2011).
14. Bowling, D. L. & Purves, D. A biological rationale for musical consonance. *Proc. Natl Acad. Sci. USA* **112,** 11155–11160 (2015).
15. Butler, J. W. & Daston, P. G. Musical consonance as musical preference: a cross-cultural study. *J. Gen. Psychol.* **79,** 129–142 (1968).
16. Fritz, T. *et al.* Universal recognition of three basic emotions in music. *Curr. Biol.* **19,** 573–576 (2009).
17. Brown, S. & Jordania, J. Universals in the world's musics. *Psychol. Music* **41,** 229–248 (2013).
18. Maher, T. F. "Need for resolution" ratings for harmonic musical intervals: A comparison between Indians and Canadians. *J. Cross Cult. Psychol.* **7,** 259–276 (1976).
19. Godoy, R. *et al.* Moving beyond a snapshot to understand changes in the well-being of native Amazonians. *Curr. Anthropol.* **50,** 563–573 (2009).
20. Riester, J. *Canción y Producción en la Vida de un Pueblo Indígena: los Chimane del Oriente Boliviano* (Los Amigos del Libro, 1978).
21. McDermott, J. H., Lehr, A. J. & Oxenham, A. J. Individual differences reveal the basis of consonance. *Curr. Biol.* **20,** 1035–1041 (2010).
22. Cousineau, M., McDermott, J. H. & Peretz, I. The basis of musical consonance as revealed by congenital amusia. *Proc. Natl Acad. Sci. USA* **109,** 19858–19863 (2012).
23. Terhardt, E. On the perception of periodic sound fluctuations (roughness). *Acustica* **30,** 201–213 (1974).
24. Kumar, S., Forster, H. M., Bailey, P. & Griffiths, T. D. Mapping unpleasantness of sounds to their auditory representation. *J. Acoust. Soc. Am.* **124,** 3810–3817 (2008).
25. Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L. & Poeppel, D. Human screams occupy a privileged niche in the communication soundscape. *Curr. Biol.* **25,** 2051–2056 (2015).
26. Moore, B. C., Glasberg, B. R. & Peters, R. W. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* **80,** 479–483 (1986).
27. Darwin, C. J. Perceiving vowels in the presence of another sound: constraints on formant perception. *J. Acoust. Soc. Am.* **76,** 1636–1647 (1984).
28. Zentner, M. R. & Kagan, J. Perception of music by infants. *Nature* **383,** 29 (1996).
29. Trainor, L. J., Tsang, C. D. & Cheung, V. H. W. Preference for sensory consonance in 2- and 4-month-old infants. *Music Percept.* **20,** 187–194 (2002).
30. Plantinga, J. & Trehub, S. E. Revisiting the innate preference for consonance. *J. Exp. Psychol. Hum. Percept. Perform.* **40,** 40–49 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors gratefully acknowledge the support of the National Science Foundation to R.A.G., a McDonnell Scholar Award to J.H.M., the TAPS Bolivia Study Team (particularly T. Huanca), C. García for assistance with Fig. 1a, E. Gibson for logistical help, N. Jacoby and M. Salinas for assistance recording and interviewing Tsimane' musicians, S. Popham and L. Chen for data collection, and D. Boebinger and K. Woods for comments on the manuscript.

**Author Contributions** J.H.M., A.F.S., E.A.U. and R.A.G. designed the experiments, collected the data, and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.M. (jhm@mit.edu).

**Reviewer Information** *Nature* thanks S. Trehub and the anonymous reviewer(s) for their contribution to the peer review of this work.

# METHODS

**Participants (Study 1).** *US musicians.* The first group of US residents ($n = 23$, 19 female; mean age = 20.2 years, s.d. = 2.4 years) was selected to have experience playing a musical instrument (mean number of years spent playing an instrument = 7.7 years, s.d. = 4.9 years, range = 2–18 years). None were professional musicians, but all reported having taken music lessons or classes for much of the time that they were musically active (mean number of years during which lessons were taken = 6.8 years, s.d. = 4.3 years). All were residents of New York City. Two participants had immigrated to the United States as young children (at the ages of 2 and 4 years old); the rest were born and raised in the United States.

*US non-musicians.* The second group of US residents ($n = 25$, 12 female; mean age = 37.9 years, s.d. = 15.1 years) was selected to have minimal experience playing or studying music (mean number of years spent playing an instrument or singing = 0.2 years, s.d. = 0.3 years, range = 0–1 years). 17 out of 25 (68%) had never played an instrument or sung in an organized setting, and most had never taken music lessons or classes (mean number of years during which lessons were taken = 0.1 years, s.d. = 0.3 years). All were residents of the Boston metropolitan area and had been born and raised in the United States. No attempt was made to achieve demographic diversity among US participants because numerous similar experiments had previously been conducted in a variety of populations from Western societies[21,22,31–33], and the effects are known to generalize across age, gender, and geographical region of North America. However, the non-musician US sample was similar in demographic composition and musical experience to each of the Bolivian populations tested.

*La Paz (capital).* Participants ($n = 24$, 14 female; mean age = 29.9 years, s.d. = 9.3 years) had been born and raised in La Paz, the capital of Bolivia. Only three reported any experience playing music (across the group: mean number of years spent playing an instrument = 0.3 years, s.d. = 0.8 years), and only five reported experience singing (across the group: mean number of years spent singing = 0.5 years, s.d. = 1.1 years).

*San Borja (town).* Participants ($n = 26$, 15 female; mean age = 31.6 years, s.d. = 12.9 years) had resided for most of their life in San Borja, a small town in the department of Beni in the Amazon basin. San Borja could be reached by car during dry months of the year from other towns in Bolivia, but was accessible only by plane during much of the rainy season owing to land slides and flooding. Most of the townspeople tested (17 out of 26) reported having experience singing, but only 8 out of 26 had experience playing a musical instrument.

*Tsimane'.* Participants ($n = 64$, 31 female; mean age = 31.5 years, s.d. = 10.2 years) lived in a small rainforest village (Santa Maria) without electricity or tap water. Santa Maria could only be reached by canoeing up the Maniqui River (an indirect tributary of the Amazon). Age data are approximate, as many Tsimane' do not know their precise age in years[34]. Results of Fig. 2 were similar (all statistical comparisons produced the same result) if we limited the analysis to participants between the ages of 16 and 35, who were more likely to know their exact age ($n = 41$, mean age of 25.0 years, s.d. = 5.8 years). Music performance among the Tsimane' was largely limited to a small number of expert practitioners who would perform traditional songs at social gatherings; most Tsimane' participants did not perform music themselves.

Sample sizes for the United States were chosen based on previous knowledge of effect sizes and variability[21]. Sample sizes for Bolivia were chosen to approximately match the US groups (capital and town) or to be as large as possible given practical constraints (Tsimane'). See Extended Data Table 1 for a summary of the participant group demographics for Study 1.

**Chord stimuli and analysis (Study 1).** Chord stimuli were taken from a previous study[21]. Pleasantness ratings were obtained for 10 chord types: the minor second, major second, tritone, major seventh, and augmented triad (conventionally dissonant in Western music), and the major third, perfect fourth, perfect fifth, and major triad (conventionally consonant in Western music), along with the unison (conventionally neutral in Western music). The consonant and dissonant chords had been found previously to yield high and low pleasantness ratings, respectively, in Western listeners[21]. All chords were composed of intervals from the equal-tempered scale.

Synthetic chords were composed of synthetic notes: complex tones containing the first 10 harmonics (in sine phase). Harmonic amplitudes were attenuated at 14 dB per octave to mimic naturally occurring musical note spectra. Synthetic chords were given temporal envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) that was truncated at 2 s.

Sung chords were composed of notes (using the vowel /u/; 'ooh') produced and recorded by a professional singer (A. Lehr). They were high-pass filtered with a cutoff frequency of 150 Hz (fourth order Butterworth filter) to eliminate handling noise from the microphone. The temporal envelope of the onset of the sung notes was not altered apart from applying a half-Hanning window (10 ms). A linear ramp was applied over the last half of each note to fade them smoothly down to zero, such that the duration was the same as the synthetic notes. Example stimuli (taken from ref. 21) are available online: http://mcdermottlab.mit.edu/consonance_examples/index.html.

Each chord was presented four times, each time with a different root note, drawn from C#4, D#4, F4, and G4 for the synthetic chords and G#3, A#4, B#4, and D4 for the sung chords, where C4 denotes middle C (the range was lower for the sung chords to accommodate the range of the singer). There were thus a total of 40 chords presented for each of the synthetic or sung chord experiments. Sung and synthetic chords were presented in separate blocks. Tsimane' participants completed either the sung ($n = 33$) or the synthetic ($n = 31$) chord block due to time constraints. All other participants completed both blocks.

Ratings were averaged across consonant chords and dissonant chords to yield a single average rating for each aesthetic class for each participant. Apart from the analysis in Extended Data Fig. 1, all statistics were performed on these averaged ratings. The ratings of the unison were omitted from the main analysis to simplify data presentation (to two stimulus conditions per experiment).

**Harmonic/inharmonic synthetic tone stimuli (Study 1).** The harmonic and inharmonic tone stimuli were taken from a previous study[21]. The harmonic tones contained a subset of the frequencies of a normal harmonic tone. There were three types of harmonic tone: one containing harmonics 1, 2, 4 and 8, one containing harmonics 1, 2, 3, 5 and 9, and one containing just the first harmonic. There were also three types of inharmonic tone. Two of them were modifications of the second harmonic stimulus. In the first type of inharmonic tone (jittered), the harmonic frequencies were perturbed up (even components) or down (odd components) by 0.5 semitones. In the second type of inharmonic tone (shifted), all the harmonic frequencies were increased by 30 Hz. To avoid differences in beating between the harmonic and inharmonic tones, alternate frequency components were presented to opposite ears (the wide spacing between components additionally minimized beating). The third type of inharmonic tone consisted of two pure tones separated by 1.5 semitones, presented dichotically to minimize beating. The resulting stimulus was inharmonic, while being roughly matched in frequency to the single-frequency harmonic stimulus, but without introducing salient differences in beating.

Inharmonicity detection thresholds for a single mistuned frequency component in a harmonic complex have been previously estimated to be about 1% of the harmonic's frequency[26]. Our inharmonic perturbations were well above this: ~3% for the jittered condition, between ~1.5–10%, depending on the harmonic, for the shifted condition, and 9% for the dichotic pure tone stimulus. Each of the perturbations therefore produced audible inharmonicity, which to a typical Western listener is perceived as unpleasant[21].

All harmonic and inharmonic stimuli had components whose amplitudes decreased by 14 dB per octave to resemble naturally occurring sounds. All tones were given temporal envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) truncated at 2 s.

Each type of tone was presented four times, each time with a different base frequency, drawn without replacement from the 4-semitone range above middle C.

**Smooth/rough synthetic tone stimuli and analysis (Study 1).** The smooth and rough tone stimuli were taken from a previous study[21]. Tones that were perceptually rough or smooth were generated by presenting pairs of single frequencies to either the same or different ears (diotic and dichotic presentation, respectively). Diotic presentation of two similar but non-identical frequencies is known to produce the 'rough' sensation of beats, typically considered unpleasant by Western listeners[5,7,21,25]. In contrast, dichotic presentation of two such frequencies greatly attenuates perceived beats[35], but leaves the spectrum (and its pitch) unchanged relative to the diotic version[36]. The frequencies composing each stimulus were separated by either 0.75 or 1.5 semitones (1.5 for the low- and mid-frequency ranges, and 0.75 for the high-frequency range, to produce beat frequencies with prominent roughness), such that considerable beating was heard when presented diotically[5,7,21]. These stimuli, like the harmonic/inharmonic synthetic tones, were given envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) truncated at 2 s.

We generated smooth/rough tones in three frequency ranges. Each type of tone was presented four times, each time with a different base frequency, drawn without replacement from the 4-semitone range above either middle C (low-frequency), one octave above middle C (medium-frequency), or two octaves above middle C (high-frequency). The low-frequency dichotic tone also served as an inharmonic stimulus that could be compared to the single frequency harmonic stimulus.

The harmonic/inharmonic and smooth/rough tones were presented in random order in a single block. There were four exemplars of each of the eleven stimulus types, for a total of 44 trials. To analyse the harmonic/inharmonic tone data, the ratings of the three harmonic stimuli were averaged together, as were the ratings of the three inharmonic stimuli (one of which was the low-frequency dichotic frequency pair stimulus). All statistics were performed on these two average ratings. To analyse the smooth/rough tone data, the ratings of the three dichotic (smooth) stimulus types were averaged together, as were the three diotic (rough) stimulus types. All statistics were performed on these two average ratings.

**Vocalization stimuli (Study 1).** Vocalizations were a subset of the Montreal Affective Vocalization Set[37], which consists of vocalizations produced by actors in a laboratory setting. Five vocalizations were selected from each of the categories of laughter, gasps, and crying, which had previously been found to be rated as pleasant (laughter) and unpleasant (gasps and cries) by North American listeners[22]. The stimuli were presented in random order in a single block of 15 trials. The crying sounds gave similar results to the gasps (low ratings by all groups) and their ratings were omitted from the analysis to simplify the data presentation (to two stimulus conditions per experiment).

**Stimulus presentation (Study 1).** All stimuli were normalized to have the same rms level. Stimuli were played out by a laptop computer over closed headphones (Sennheiser HD 280Pro). The volume was set to a level that was determined in pilot experiments to be comfortable (approximately 70 dB SPL), and that was fixed across participants.

**Experimental protocol (Study 1).** Sound stimuli for a given experiment were presented sequentially in random order. After each stimulus was presented, participants gave their rating verbally (participants could respond "like it a lot", "like it a little", "dislike it a little", or "dislike it a lot"), and the experimenter recorded the response on a data entry sheet. To avoid biasing participants' responses or data entry, experimenters were always blind to the stimulus being presented, and stimuli were audible only to the participant.

The experiments were completed in a single sitting along with several other experiments that are not reported here. Experiment order was counterbalanced across the participants in each group. Experiments were conducted in June–July 2011 (Tsimane'), January–February 2012 (San Borja and US musicians), July 2012 (La Paz), and March–April 2014 (US non-musicians).

Both studies were approved by the Tsimane' Council (the governing body of the Tsimane' in the Maniqui basin, where the studies took place), as well as the Committee on the Use of Humans as Experimental Subjects at MIT and the Committee for Protection of Human Subjects at Brandeis. Experiments were conducted with the informed consent of the participants.

**Participants (Study 2).** *US musicians.* A group of musically experienced US residents was tested to provide baseline preference measures. They were selected to have experience playing a musical instrument ($n=47$, 32 female; mean age $=25.9$ years, s.d. $=7.5$ years; mean number of years spent playing an instrument $=8.4$ years, s.d. $=6.5$ years, range $=1$–$25$ years). None of the participants were professional musicians; we label them here as 'musicians' primarily to distinguish them from the non-musicians of Study 1 who were selected to have almost no experience playing music or singing. All were residents of the Boston metropolitan area and had been born and raised in the United States. All participants completed all experiments in the study.

*Tsimane'.* Participants ($n=50$, 27 female; mean age $=27.4$ years, s.d. $=9.8$ years; as in Study 1, ages are approximate) lived in five small villages surrounding San Borja (Arenales, Manguito, Pachual, Limoncito, and Tacuaral). Unlike the village of Santa Maria from Study 1, the villages of Study 2 could be reached by automobile during the dry season and had begun to be equipped with electricity at the time of testing (some families were connected to the electrical grid). The choice of villages was dictated by accessibility given the local weather conditions during the study (roads were often muddy and impassable). We tested a larger group of participants and only analysed data from the subset who did not own a radio, to minimize the chance that they had significant exposure to Western music. Only 27 of these 50 participants completed the synthetic interval experiment (it was substituted for another experiment half-way through our fieldwork). One of the 50 participants did not complete the discrimination experiment. The size of the Tsimane' group was as large as possible given the practical constraints of the study (the US group was chosen to approximately match it in size). See Extended Data Table 2 for a summary of the participant group demographics for Study 2.

**Chord stimuli (Study 2).** Chord stimuli were generated as in Study 1. The synthetic and sung two-note chord experiments presented intervals from 0 to 11 semitones at each of 4 different root-pitches, yielding 48 trials in total. The sung triad experiment presented the major, minor, diminished, and augmented triads each at 4 different root-note pitches, yielding 16 trials in total (in the main analysis only the major and augmented triads were analysed, as they produced the largest contrast in US listeners). The root notes for the synthetic chords were the first four notes above middle C. For both sung chord experiments, the root notes were the first four notes above G#3 to accommodate the vocal range of the singer. The notes in the sung chord experiments were sung vowels ("aah" for two of the exemplars and "ooh" for the other two, randomly assigned to the root notes). Within experiments, trials were randomly ordered.

**Vocal harmony stimuli (Study 2).** Because the Tsimane' generally do not sing in groups and were unable to sing or play in coordination upon request, it was necessary to generate harmonies from solo recordings. We recorded two Tsimane' vocalists singing the same melodic phrase several times in a row. The phrases were chosen by the vocalists to be Tsimane' songs they knew well and tended to sing in social settings. We selected sets of phrases by each singer with very similar timing (by visually comparing the waveforms), yielding 26 phrases with two repetitions each. We then pitch-shifted one of the phrases in a set and added it to the other (unshifted) version of the phrase. Pitch shifting was performed using STRAIGHT[38]. The experiment presented harmonies with intervals ranging from 1 to 11 semitones, as well as single phrases and the unison (two different repetitions of the same phrase at the same pitch). There were 4 exemplars of each interval, each generated from a different melodic phrase, yielding 52 trials in total. Each of the 26 sung phrases was used twice in the experiment. We generated 15 random assignments of phrases to conditions with the constraint that the two occurrences of a phrase not be assigned to the same condition. One of these assignments was chosen at random for each participant. Although the song excerpts were not selected to be widely recognizable (they were simply those that the singers were comfortable singing), 19 out of the 50 participants were able to name at least one song that they recognized during the experiment.

**Pure tone interval stimuli (Study 2).** Pure tone intervals were generated by the same procedure used for the smooth/rough synthetic tone stimuli from Study 1, the only difference being that the tones in a pair were separated by integer numbers of semitones. The experiment presented intervals from 0 to 8 semitones (in diotic and dichotic versions), each at 2 different root-note frequencies (523 and 740 Hz). The diotic and dichotic 0-semitone conditions were identical.

**Discrimination experiment stimuli/task (Study 2).** Listeners heard harmonics 1–6 of a fundamental frequency (F0). On one-quarter of trials the even-numbered harmonics were mistuned by +12%, rendering the tone inharmonic and causing the even and odd harmonics to segregate and be heard (by typical Western listeners) as two separate tones. On another quarter of trials the even harmonics onset 150 ms earlier than the odd harmonics, again causing the even and odd harmonics to perceptually segregate. On the remaining 50% of trials the even harmonics were unperturbed and the complex tone tended to be heard as a single sound. The task was to judge whether a trial contained one or two sounds. After each trial the experimenter told the participant if they had answered correctly (the correct answer was considered to be "one" when the harmonics were unperturbed, and "two" when mistuned or offset). The tone F0 took 9 different values (0, 4 and 8 semitones above 200, 300 and 400 Hz), yielding 36 trials in total. Harmonics were 750 ms in duration and were given flat amplitude envelopes that began and ended with half-Hanning windows (10 ms). Stimuli were presented at 75 dB SPL.

The experiment was preceded by a practice experiment that was identical in all respects except that only 5 F0 s were used (200, 282.8, 300, 400 and 424.3 Hz), yielding 20 trials in total.

**Stimulus presentation (Study 2).** Same as in Study 1.

**Experimental protocol (Study 2).** Pleasantness ratings were obtained as in Study 1, except that the experimenter entered responses directly into the computer.

The experiments were completed in a single sitting. Experiment order was generated randomly for each participant, with the exception that the discrimination experiments were conducted first or last with equal probability (because the instructions were different than for the pleasantness rating experiments). Experiments were conducted in July 2015 (Tsimane') and July–October 2015 (US musicians).

**Analysis.** For preference experiments, the ratings for each participant were averaged across the exemplars for a condition (for example, laughs or gasps for the vocalization experiment, or major triad for one of the chord experiments), yielding one mean rating per condition per participant per experiment. In the chord, tone, and vocal harmony experiments, most analyses were done after averaging these mean ratings across groups of conditions, for example, the consonant chords/harmonies, or the harmonic tones, again yielding one mean rating for each participant for each group of conditions in each experiment. The results graphs plot these mean ratings averaged across participants, and the error bars plot the standard error of these mean ratings across participants.

For the discrimination experiment, we obtained one measure of sensitivity ($d'$) for each condition for each participant from their hit and false alarm rates. The results graphs plot the mean and standard error of this sensitivity measure across subjects.

**Statistics.** Repeated-measures ANOVAs were used to test for main effects of stimulus condition and for interactions between the effect of stimulus condition and participant group. Mauchly's test was used to test for violations of the sphericity assumption. When Mauchly's test was significant, the Greenhouse–Geisser correction was used. Pair-wise comparisons were made with two-tailed $t$-tests, and were not corrected for multiple comparisons because they were preceded by

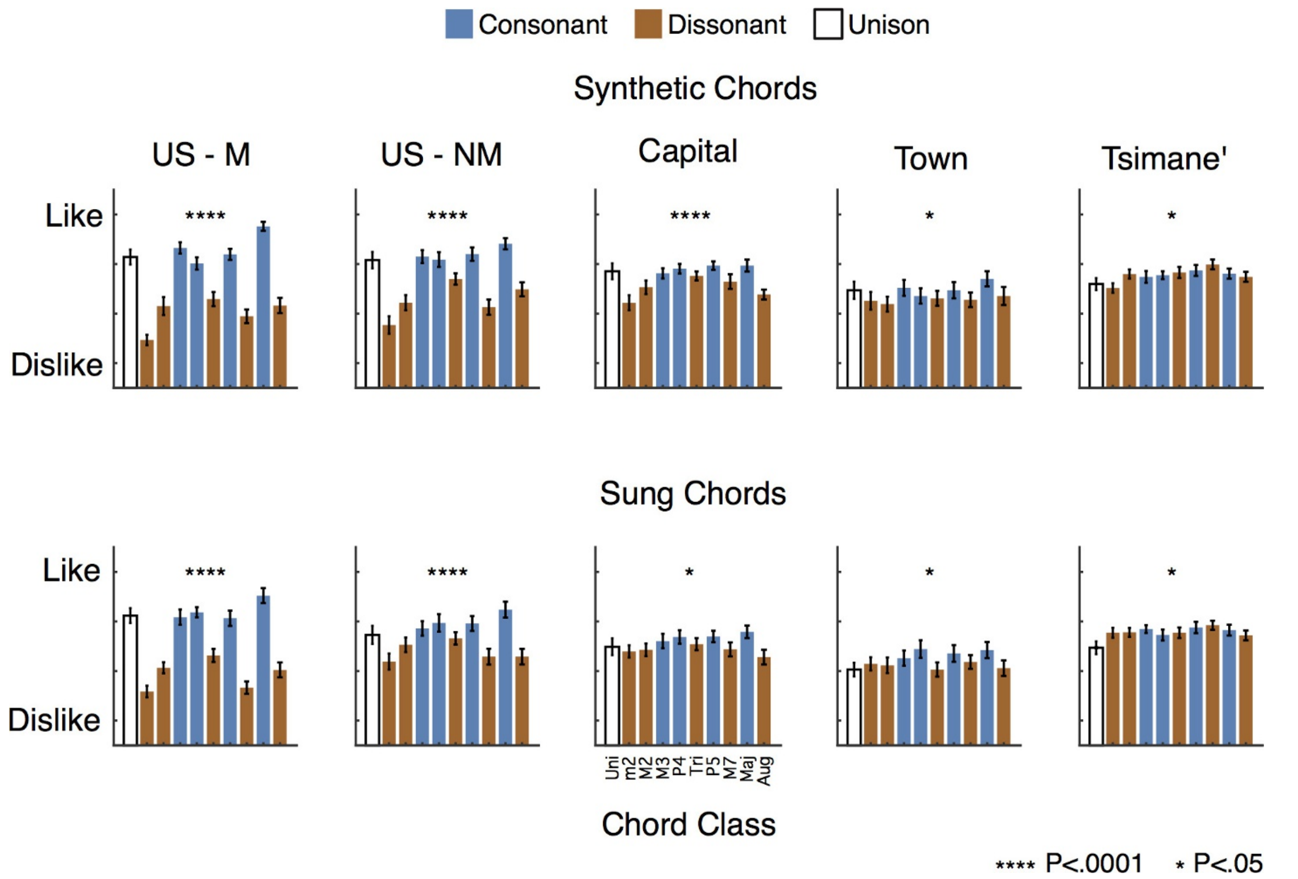
ANOVAs that revealed significant main effects or interactions. Pearson correlations were used to determine the relationship between chord ratings and interval size (Extended Data Fig. 2) and to test the reliability of ratings of song excerpts and vocal harmonies (Extended Data Fig. 3). Data distributions were assumed to be normal, and were evaluated as such by eye. The exceptions to the above were in the analysis of individual melodic phrase ratings in the vocal harmony experiment (Extended Data Fig. 3) and of the pure tone interval experiment from Study 2 (Fig. 3f), in which only two ratings per condition were collected per subject due to the large number of conditions (producing non-normal data distributions), necessitating non-parametric tests. In these cases, Friedman's non-parametric test of differences among repeated measures was used to test for main effects, and the Wilcoxon signed rank test was used for pairwise comparisons.

**Tsimane' music.** Tsimane' music consists of songs that have apparently been passed down for generations. At present these songs are sung primarily at social gatherings by particular individuals who know the songs, typically when adults gather to consume "chicha", a traditional homemade alcoholic beverage. The primary documentation of Tsimane' music comes from the work of J. Riester and G. Rocckel (an anthropologist and ethnomusicologist, respectively), who compiled 140 traditional songs and analysed their subject matter and musical content[20]. We have also compiled Tsimane' songs as part of a broader longitudinal study[39]. Known Tsimane' songs tend to reflect their everyday lives, concerning production activities (such as hunting), division of labour, love, social organization, and religion. Musically, many songs are based on anhemitonic pentatonic scales. Others are constructed from three notes spanning a perfect fourth (typically divided into intervals of a whole tone and minor third). Examples of recorded songs are provided in Supplementary Information (Supplementary Audio 1 and 2). Instrumental renditions of songs are also common, typically on wind or string instruments. Drums exist but do not seem to be used to accompany other instruments or singing, at least not any more. There is one report of Tsimane' performing traditional dances in the mid-twentieth century[40], but to the best of our knowledge, traditional dancing is no longer practiced.

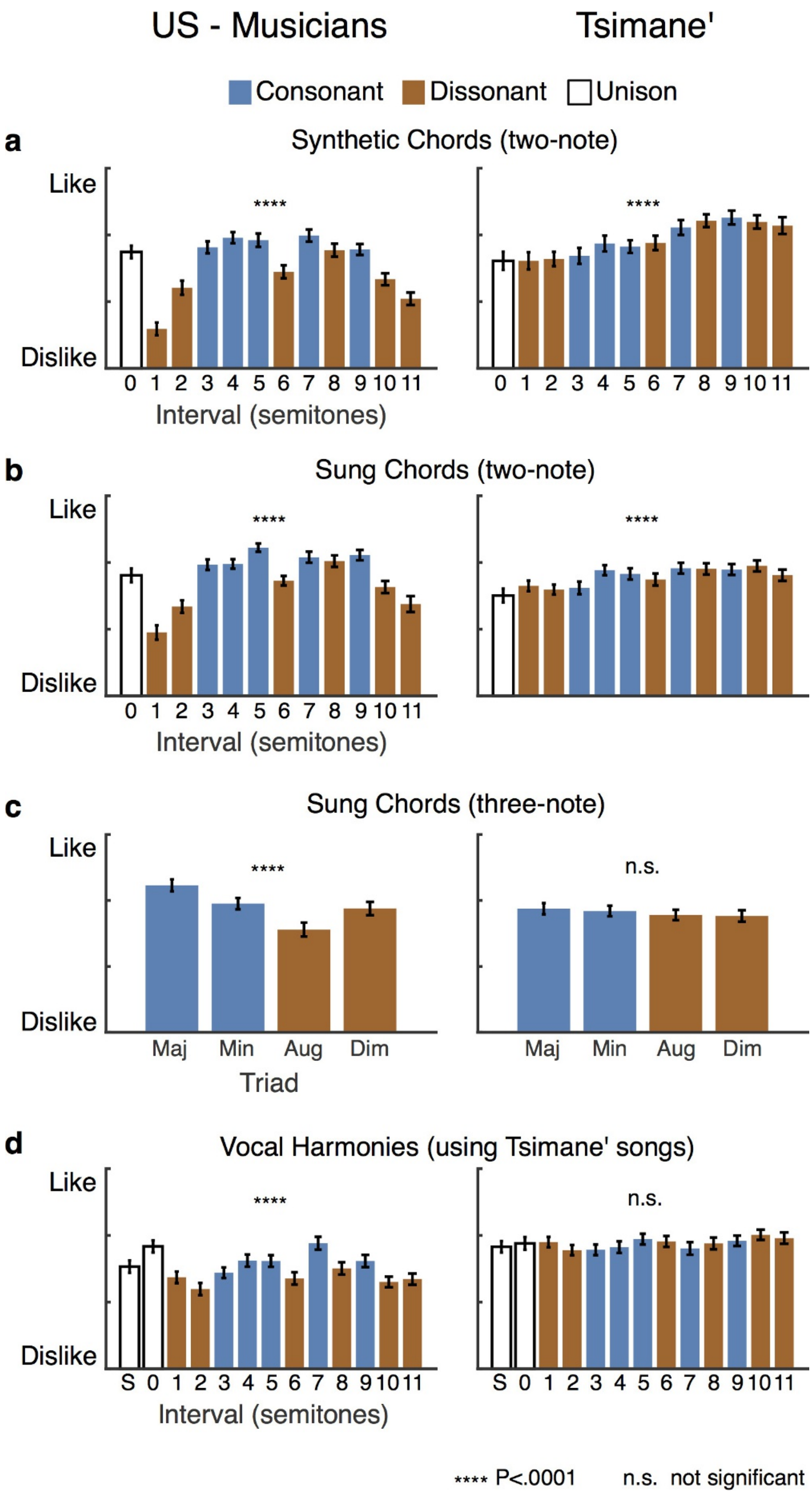
For our purposes, the most interesting feature of Tsimane' music is that group performances appear to be absent. We substantiated this over several days of interviews with 10 Tsimane' musicians. They consistently reported that musical performances often occur in social settings, but that only one person performs at a time. Multiple musicians might perform in succession, but not concurrently; instrumental accompaniment to singing is rare. We were initially sceptical of this claim, and so brought pairs of musicians together and asked them to perform together. They were usually reluctant to do so (despite being eager to perform solo songs for each other), and on the few occasions when we could elicit concurrent performances, they were unable to coordinate. Our experience suggests that group musical performance in Tsimane' culture is rare at best.

The origins of the Tsimane' music tradition are unknown, but they clearly have had some contact with the music traditions of other cultures for at least the past century, including exposure to Western musical instruments, primarily through friars and missionaries as well as through Bolivian highlanders. It is also clear that the Tsimane' musical tradition is slowly fading from their culture, in part due to the influence of missionaries. For instance, shamans were once a staple guardian of Tsimane' songs, but were banned by Protestant missionaries who became influential in parts of the Tsimane' territory[41]. The past two decades have also seen the diffusion of highland Andean music and Christian hymns into the Tsimane' territory through the radio station of Protestant missionaries, nearby commercial radio stations, and highland migrants moving into the lowlands. As part of a longitudinal study with the Tsimane we found that of the 2,549 adults interviewed annually during 2002–2010 in 13 villages along the Maniqui River (varying in proximity to market towns and roads), 30% reported owning at least one battery-operated radio. These figures may overstate the impact of radios because some of the radios lacked batteries, or were broken. It is also unclear to what extent the Tsimane' use their radios to listen to music, as they are also a primary source of local news. More recently, the Bolivian government has started to expand the electricity grid to the countryside surrounding the town of San Borja, and as a result, some villages have begun to acquire electricity. Our impression is that the effect of these outside influences has been to reduce the prevalence of traditional Tsimane' music in their culture. Most Tsimane' that we interviewed had knowledge of the traditional songs, but this might well diminish in the future.

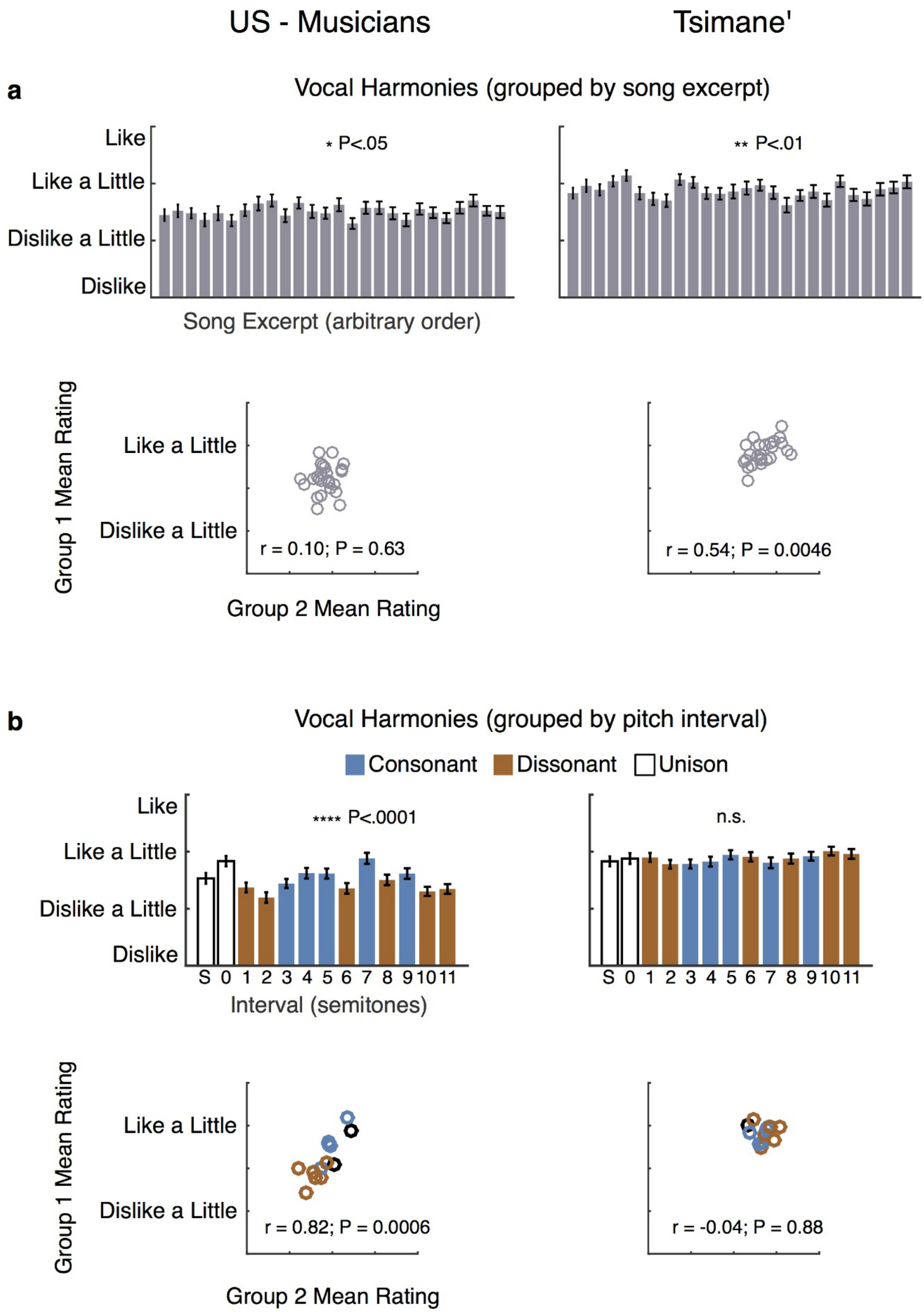

31. Malmberg, C. F. The perception of consonance and dissonance. *Psychol. Monogr.* **25,** 93–133 (1918).
32. Hutchinson, W. & Knopoff, L. The acoustical component of western consonance. *Interface* **7,** 1–29 (1978).
33. Roberts, L. Consonance judgments of musical chords by musicians and untrained listeners. *Acustica* **62,** 163–171 (1986).
34. Gurven, M., Kaplan, H. & Supa, A. Z. Mortality experience of Tsimane Amerindians of Bolivia: regional variation and temporal trends. *Am. J. Hum. Biol.* **19,** 376–398 (2007).
35. Rutschmann, J. & Rubinstein, L. Binaural beats and binaural amplitude-modulated tones: successive comparison of loudness fluctuations. *J. Acoust. Soc. Am.* **38,** 759–768 (1965).
36. Bernstein, J. G. & Oxenham, A. J. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* **113,** 3323–3334 (2003).
37. Belin, P., Fillion-Bilodeau, S. & Gosselin, F. The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* **40,** 531–539 (2008).
38. Kawahara, H. & Morise, M. TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *Sadhana* **36,** 713–722 (2011).
39. Leonard, W. R. *et al.* The Tsimane' Amazonian Panel Study (TAPS): Nine years (2002-2010) of annual data available to the public. *Econ. Hum. Biol.* **19,** 51–61 (2015).
40. Hissink, K. Felsbilder und salz der Chimanen-Indianer. *Paideuma (Wiesb.)* **6,** 60–68 (1955).
41. Huanca, T. *Tsimane' Oral Tradition, Landscape, and Identity in Tropical Forest* (Wa-Gui, 2008).




**Extended Data Figure 1 | Average ratings of individual chords from the synthetic and sung chord experiments from Study 1.** Ten chords were presented: the unison, minor second, major second, major third, perfect fourth, tritone, perfect fifth, major seventh, major triad, and augmented triad. The composite ratings plotted in Fig. 2 were averages of those for the consonant (blue) and dissonant (brown) chords. Ratings are from 23 US musicians (US-M), 25 US non-musicians (US-NM), 24 Bolivian city-dwellers (capital), 26 Bolivian town-dwellers, and 64 Tsimane'. Asterisks denote statistical significance of a repeated-measures ANOVA across all chord ratings. Data are mean and s.e.m.

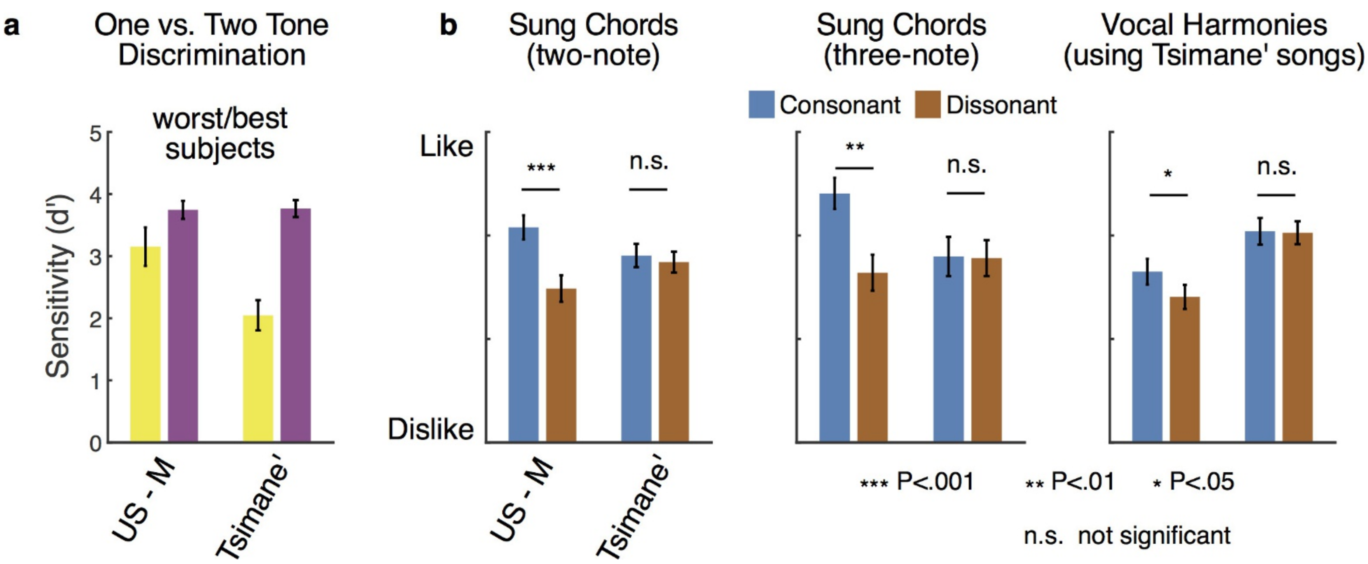



**Extended Data Figure 2 | Average ratings of individual chords and vocal harmonies from Study 2.** Rating variation for US listeners ($n = 47$) was largely determined by consonance and dissonance, whereas for Tsimane' listeners ($n = 50$) it was largely driven by interval size (in the two-note chord experiments in which they exhibited significant variation). 'S' denotes a single vocal phrase in the vocal harmonies experiment, whereas '0' denotes the unison condition (two concurrently presented exemplars of the same phrase at the same pitch). Data are mean and s.e.m.




**Extended Data Figure 3** | See next page for caption.

**Extended Data Figure 3 | Pleasantness ratings of vocal harmonies from Study 2, averaged across pitch intervals or across song excerpts, for US musicians and Tsimane'. a**, Ratings of song excerpts used to create harmonies (averaged across intervals). Each participant heard each excerpt twice, in each of two different randomly assigned interval conditions. The ratings of the two excerpt occurrences were averaged for each participant (47 US musicians and 50 Tsimane'). Top panels plot these mean ratings averaged across all participants. Asterisks denote statistical significance of Friedman's non-parametric test of differences among repeated measures across song excerpts. Ratings of US listeners varied across song excerpts ($\chi^2(25) = 40.92$, $P = 0.02$), as did those of Tsimane' listeners ($\chi^2(25) = 49.01$, $P = 0.003$), but the preferences of the two participant groups were not significantly correlated ($r = -0.19$, $P = 0.35$). Bottom panels plot the mean ratings of each excerpt averaged across the first and last halves of the participants tested. The mean song excerpt ratings of the two half-groups were significantly correlated for Tsimane' but not for US participants, indicating that the preferences in US listeners were not reliable. **b**, Ratings of pitch intervals of the harmonies (averaged across song excerpts). Each participant heard each pitch interval four times, each time generated with a different song excerpt. The ratings of the four interval occurrences were averaged for each participant. Top panels (replicated from Extended Data Fig. 2) plot these mean ratings averaged across all participants. Asterisks denote statistical significance of a repeated-measures ANOVA across all chord ratings. Bottom panels plot the mean rating of each pitch interval averaged across the first and last halves of the participants tested. The mean interval ratings of the two groups were significantly correlated for US but not for Tsimane' participants. Data are mean and s.e.m.




**Extended Data Figure 4 | Discrimination and preference data for worst US and best Tsimane' participant subsets from Study 2.** **a**, Discrimination performance for the worst US-M participants (bottom third, $n = 16$) and best Tsimane' participants (top third, $n = 16$), selected based on performance in the onset asynchrony condition. This selection criterion produced a group of Tsimane' listeners who achieved an average $d'$ of 2.0 on the mistuned condition—good performance in absolute terms, and only slightly worse than the poorly performing Western listeners. **b**, Pleasantness ratings for the subsets of participants from **a** for conventionally consonant and dissonant sung chords and vocal harmonies. This subset of Tsimane' listeners remained indifferent to dissonance, rating consonant and dissonant sung intervals, triads, and harmonies similarly ($t(15) < 1.2$, $P > 0.28$ in all cases). By contrast, the Western subset yielded significant consonant preferences ($t(15) > 2.65$, $P < 0.05$ in all cases). These results suggest that the absence of a consonance preference cannot be explained by a lack of sensitivity to the underlying stimulus distinction—the Tsimane' were able to distinguish harmonic from inharmonic tones despite not having a preference for one over the other. Data are mean and s.e.m.

**Extended Data Table 1 | Summary of participant group demographics for Study 1**

## Participant Summary Table for Study 1

| | US (Musicians) | US (Non-musicians) | La Paz (Capital) | San Borja (Town) | Santa Maria (Tsimane' village) |
|---|---|---|---|---|---|
| **N total** | 23 | 25 | 24 | 26 | 64 |
| **N female** | 19 | 12 | 14 | 15 | 31 |
| **Mean age** | 20.2 | 37.9 | 29.9 | 31.6 | 31.5 |

**Extended Data Table 2 | Summary of participant group demographics for Study 2**

**Participant Summary Table for Study 2**

| | US (Musicians) | Tsimane' |
|---|---|---|
| **N total** | 47 | 50 |
| **N female** | 32 | 27 |
| **Mean age** | 25.9 | 27.4 |