

Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors

Nori Jacoby^{1,2}, Rainer Polak³, Jessica A. Grahn⁴, Daniel J. Cameron⁵, Kyung Myun Lee^{6,7}, Ricardo Godoy⁸, Eduardo A. Undurraga^{9,10}, Tomás Huanca¹¹, Timon Thalwitzer¹², Noumouké Doumbia¹³, Daniel Goldberg¹⁴, Elizabeth Margulis¹⁵, Patrick C. M. Wong¹⁶, Luis Jure¹⁷, Martín Rocamora¹⁸, Shinya Fujii¹⁹, Patrick E. Savage¹⁹, Jun Ajimi²⁰, Rei Konno¹⁹, Sho Oishi¹⁹, Kelly Jakubowski²¹, Andre Holzapfel²², Esra Mungan²³, Ece Kaya^{24,25}, Preeti Rao²⁶, Mattur Ananthanarayana Rohit²⁶, Suvarna Alladi²⁷, Bronwyn Tarr^{28,29}, Manuel Anglada-Tort¹, Peter M. C. Harrison¹, Malinda J. McPherson^{30,31,32}, Sophie Dolan^{30,33}, Alex Durango^{30,32,34} and Josh H. McDermott^{30,31,32,35}

¹Computational Auditory Perception Group, Max Planck Institute for Empirical Aesthetics, Germany, 60322 Frankfurt am Main, Germany.

²Presidential Scholars in Society and Neuroscience, Columbia University, New York, NY 10027, USA.

³Music Department, Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany.

⁴Brain and Mind Institute and Department of Psychology, University of Western Ontario, London, ON, N6A 3K7, Canada.

⁵Department of Psychology, Neuroscience and Behaviour, McMaster University, ON, L8S 4K1, Canada.

⁶School of Humanities and Social Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea.

⁷Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea.

⁸The Heller School for Social Policy and Management, Brandeis University, Waltham, MA 02453, USA.

⁹Escuela de Gobierno, Pontificia Universidad Católica de Chile, Chile, 7820436 Santiago, Región Metropolitana, Chile.

¹⁰CIFAR Azrieli Global Scholars program, CIFAR, Toronto, ON M5G 1M1, Canada.

¹¹Centro Boliviano de Investigación y Desarrollo Socio Integral (CBIDSI), San Borja, 00000, Bolivia.

¹²Department of Musicology, University of Vienna, 1090 Vienna, Austria.

¹³Sciences de l'Education, Université Catholique d'Afrique de l'Ouest UCAO Bamako, Quartier-Mali, Bamako, Mali.

¹⁴Department of Music, University of Connecticut, Storrs, CT 06269, USA.

¹⁵Department of Music, Princeton University, Princeton, NJ 08544, USA.

¹⁶Department of Linguistics & Modern Languages and Brain and Mind Institute, Chinese University of Hong Kong, China, 00000 Shatin, New Territories, Hong Kong SAR, China.

¹⁷School of Music, Universidad de la República, Uruguay, 11200 Montevideo, Uruguay.

¹⁸Signal Processing Department; School of Engineering, Universidad de la República, 11300 Montevideo, Uruguay.

¹⁹Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-0882, Japan.

²⁰Department of Traditional Japanese Music, Tokyo University of the Arts, Japan, Tokyo University of the Arts, Taito City, Tokyo 110-8714, Japan.

²¹Department of Music, Durham University, Durham, DH1 3RL, UK.

²²Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden.

²³Department of Psychology, Bogazici University, 34342 Istanbul, Turkey.

²⁴Max Planck Research Group 'Neural and Environmental Rhythms', Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany.

²⁵Cognitive Science Master Program, Bogazici University, 34342 Istanbul, Turkey.

²⁶Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India.

²⁷Nizam's Institute of Medical Sciences Hyderabad, Punjagutta, Hyderabad, Telangana 500082, India.

²⁸Department of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford OX2 6PN, UK.

²⁹Department of Experimental Psychology, University of Oxford, Oxford, OX2 6GG UK.

³⁰Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

³¹Program in Speech and Hearing Biosciences and Technology, Harvard University, Cambridge, MA 02115, USA.

³²McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

³³Department of Brain and Cognitive Sciences, Wellesley College, Wellesley MA 02481, USA.

³⁴Neurosciences Graduate Program, Stanford University, Stanford, CA 94305, USA.

³⁵Center for Brains, Minds & Machines, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Corresponding authors:

- Nori Jacoby, Computational Auditory Perception Group, Max Planck Institute for Empirical Aesthetics, Germany, 60322 Frankfurt am Main, Germany, nori.jacoby@ae.mpg.de
- Josh McDermott, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA, Cambridge, MA 02139, USA, jhm@mit.edu

Abstract

Music is present in every known society, yet varies from place to place. What is universal to the perception of music? We measured a signature of mental representations of rhythm in 923 participants from 39 participant groups in 15 countries across 5 continents, spanning urban societies, indigenous populations, and online participants. Listeners reproduced random “seed” rhythms; their reproductions were fed back as the stimulus (as in the game of “telephone”), such that their biases (the prior) could be estimated from the distribution of reproductions. Every tested group showed a prior with peaks at integer ratio rhythms, suggesting that discrete rhythm “categories” at small integer ratios are universal. The occurrence and relative importance of different integer ratio categories varied across groups, often reflecting local musical systems. However, university students and online participants in non-Western countries tended to resemble Western participants, underrepresenting the variability otherwise evident across cultures. The results suggest the universality of discrete mental representations of music while showing their interaction with culture-specific traditions.

Introduction

Music, like language, is conceived by Western scholars to consist of combinations of discrete elements (1). Musical notes are grouped into phrases, and described in terms of discrete intervals in pitch and time. In Western music, these intervals are non-arbitrary, often defined by integer ratios between pitches or durations. However, actual musical performances deviate considerably from notated intervals in pitch and time (2-6). Discrete symbolic mental representations of music are thought to be aided by categorical perception (7-10) – the perceptual mapping of continuous spaces of signals onto discrete elements (11).

Are discrete representations fundamental to the experience of music, and to what extent do they vary cross-culturally? Attempts to characterize music from around the world have supported the idea that there are universal properties of music, including composition in terms of discrete elements defined by simple-integer ratios (12-16). However, such analyses of musical corpora have largely relied on Western-trained researchers to annotate what they hear when listening to recordings from other cultures, with the unavoidable possibility that researchers' perceptual biases influence the results. Controlled comparisons of actual perception of music from participants in diverse societies have thus far been limited to comparisons between small numbers of societies (17-29), or participants on the internet who plausibly have extensive exposure to a similar distribution of music as that consumed by typical Western participants. In addition, experiments in Western non-musicians have in some cases failed to find evidence for discrete musical features (30, 31). Thus, the universal existence of psychological mechanisms supporting discrete representations of music has remained unclear, as has the extent to which such representations are biologically constrained to exhibit characteristics found in Western listeners.

Here we present the first large-scale cross-cultural study of global variation in music perception, measuring a signature of discrete mental representations of rhythm in a large number of diverse participant groups around the world. We utilized a paradigm developed in an earlier study (25) to characterize perceptual priors on simple periodic rhythms, which in Westerners exhibit discrete peaks (modes) at rhythms composed of time intervals related by ratios of small integers ("simple" integer ratios such as 1:2 or 2:3 as opposed to 4:7 or 7:12) (32). These modes bias the perception of presented rhythms by shifting them toward the rhythm represented by a mode. Small-integer ratios are believed to be common, if not universal, in musical rhythms around the world (14, 15), but, as with other cross-cultural corpus analyses, have primarily been assessed via the annotations of Western researchers, leaving both their prevalence in non-Western music, as well as their mental representation in non-Western listeners, unclear.

In the experiment, individuals are initially presented with a random "seed" rhythm, which they reproduce by tapping (Figure 1A). Empirically, these reproductions are biased away from the actual stimulus rhythm, as might be expected if the perception of the rhythm were influenced by a prior via Bayesian inference. To reveal the prior, the reproduction is substituted as the stimulus on the following trial, and the process iterated (25, 33, 34). Each time the reproduction is further biased by the prior, such that it is eventually indistinguishable from a sample from the prior. The prior can then be estimated by running the procedure multiple times with different random seed rhythms. Here we use a kernel density estimate from the last iteration of a large set of random seeds to approximate the prior (see Methods).

We ran a version of the experiment in which the rhythms consisted of a repeating cycle of three clicks. When repeated, the clicks were separated by three successive time intervals which we constrained to sum to two seconds. The results of the experiment are plotted in a triangular "rhythm space" of three-interval rhythms (Figure 1B). Each of the three axes of the rhythm space represents one of the three intervals. Because the total duration of the three-element rhythm is constrained to 2000 ms, two of the

intervals are sufficient to uniquely specify the rhythm. The brief and simple rhythms defined by this space are elements that can be composed to form longer and more complex rhythms (8, 26, 35).

Small-integer ratio rhythms are particular points in this space (Figure 1B; red + symbols demarcate small-integer ratio rhythms whose duration ratios are defined by integers less than 3), with isochrony (1:1:1) lying in the middle. We previously found that both musician and non-musician participants in the US exhibited priors with modes at small-integer ratio rhythms (25). We also previously conducted an extensive set of control experiments to substantiate the claim that the measured distribution of responses reflects a perceptual prior rather than production constraints (25). For instance, the distribution measured in US participants is similar irrespective of whether responses are tapped or vocalized. Moreover, the modes of the prior act as rhythm “categories”, conferring the characteristics of categorical perception: discrimination of rhythms is worse for rhythms near a mode than for rhythms distant from modes (25), as would be expected from Bayesian inference (36).

The iterated reproduction paradigm is well suited to cross-cultural experiments. The task is intuitive for participants, with minimal reliance on verbal instructions, and is easy to run in the field, allowing us to run the same experiment across many different groups speaking different languages and with varying levels of education (Figure 1C). The paradigm also has the attractive feature of being independent of any specific hypothesis about the structure of rhythm representation. In particular, the experiment is not limited to testing whether phenomena prominent in Western listeners (discrete modes at small integer ratios) are present in other cultures. A group of people could in principle exhibit a uniform prior (without discrete modes), or one with completely different modal structure from that observed in Westerners.

We previously observed a prior with small-integer-ratio modes in one indigenous Amazonian society (25). However, the integer ratios at which modes occurred were different than in US participants. To test whether discrete modes at small-integer ratios are universal across groups and to assess the nature and extent of cross-cultural variation, we ran the experiment in a large number of participant groups around the world.

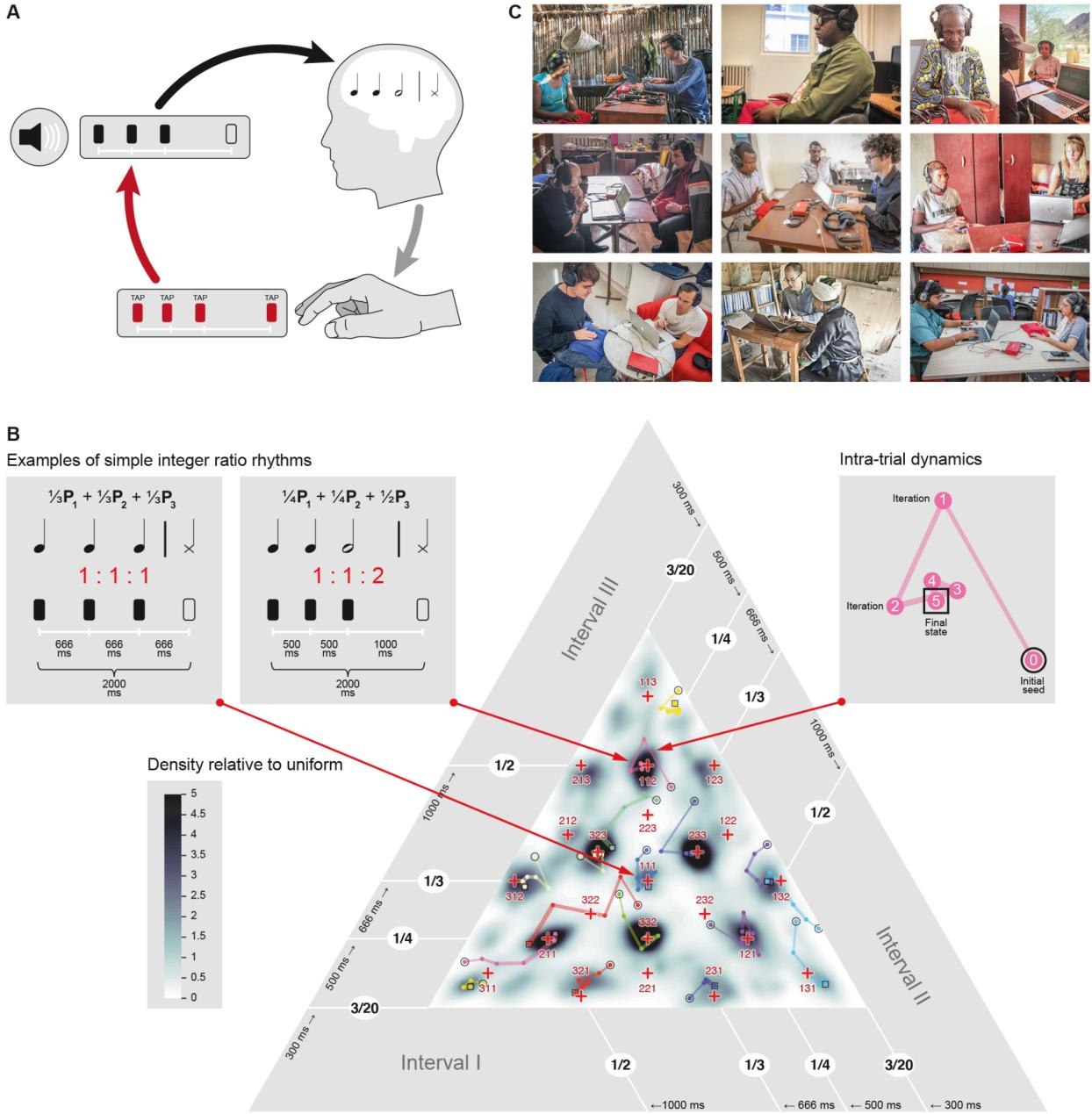


Figure 1. Iterated reproduction experiment paradigm and analysis. A) Schematic of experiment. Participants are presented with initially random seed rhythms (a repeating cycle of three clicks, defined by three inter-beat time intervals) and reproduce them by tapping. The reproduction becomes the stimulus on the following trial. This procedure is iterated 5 times (one “trial”). B) The triangular rhythm space in which results are plotted (shown here for Candombe musicians in Uruguay, for illustration purposes). Each axis of the triangle represents one of the three intervals in a rhythm. We constrained each interval to be at least 300 ms in duration, and the total pattern duration to be 2000 ms, resulting in the space within the inner triangle (see Supplementary Figures S1 and S2 for results of a second experiment conducted at a faster tempo by using a pattern duration of 1000 ms). Small-integer-ratio rhythms, in which the time intervals are related by small-integer ratios (integers less than 3; see Methods), occupy a subset of points in the triangular rhythm space. Two example small-integer ratio rhythms (1:1:2 and 1:1:1) are shown to the left. For these examples, the three intervals are marked with white lines, the solid rectangle marks the click that starts the new repeated cycle. The colored dots connected by lines show trajectories from example experimental trials. Inset shows one example trial in more detail, converging in this case to the 1:1:2 rhythm. C) The paradigm as conducted at several testing sites (First row, from left to right: Yaranda, Bolivia; Montevideo, Uruguay; Spitzkoppe, Namibia; Sagele, Mali; Second row: Pleven, Bulgaria; Bamako, Mali; Etsha, Botswana; Third row: Stockholm, Sweden; Guizhou, China; Mumbai, India).

Results

We tested 39 participant groups, spanning five continents and 15 countries (Figure 2A). The groups were chosen primarily to provide a strong test of 1) potential universality by comparing groups with diverse musical experiences, and 2) the role of musical experience in shaping mental representations. We thus included groups ranging from industrialized to small-scale societies, as well as groups of musicians and dancers from non-Western societies. We also selected groups whose musical traditions were known to have distinct rhythmic characteristics to test whether any cross-cultural variability in mental representations could be explained by exposure to local musical styles. Where possible, we tested multiple groups from the same country that differed in the nature of their presumptive musical exposure. We also tested both university students and online participants in a number of countries. Compared to the other participant groups in the same countries, these latter groups had musical experience that was more like that of typical Western participants, and were intended to assess the potential effects of exposure to Western/globalized music on mental representations of rhythm, and thus the consequences of reliance on student and online participants in cross-cultural research (37, 38). We chose participant groups based on these criteria in conjunction with practical constraints (we tested more groups in some countries than others primarily due to constraints of testing time and access to particular populations).

The key questions we sought to answer were 1) whether all groups would exhibit discrete rhythm categories, 2) whether any discrete “categories” would consistently occur at small-integer ratio rhythms, 3) how rhythm categories would vary across groups, if at all, and 4) whether any cross-cultural variation would be related to musical or other demographic characteristics of the group.

All groups exhibit priors with discrete modes

The measured priors are shown in Figure 2B. Their most obvious feature is that they are non-uniform, being dominated in all cases by a set of relatively discrete modes that form local maxima in two dimensions. This need not have been the case – some or all priors could have been uniform, or could have exhibited one-dimensional ridges rather than discrete two-dimensional modes. We substantiated this observation quantitatively in several ways. First, we found that in all 39 groups, 33% of the triangular rhythm space contained at least 61% of the rhythm reproductions in the fifth iteration (mean 70.1%; range 61.8-81.7%), suggesting that most of the probability mass is concentrated in a small portion of the space. The odds of this happening by chance are very low ($p < .001$ for points randomly positioned in the rhythm triangle). Second, the peak of the distribution in all 39 groups was at least 5 times larger than the uniform density ($p < .001$ for all groups, bootstrapping over participants in each group), suggesting that all distributions were far from uniform and very “peaky”. Finally, a Gaussian mixture model with categories constrained to 22 small-integer-ratio points (see Methods) explained most of the kernel density variability (93.8% on average, ranging from 86.2-97.8% depending on the group).

All groups exhibit modes at small-integer-ratio rhythms

Visual inspection of the modes suggests that they fall on small-integer-ratio rhythms (red crosses superimposed on Figure 2B). To quantify the extent to which this was the case, we computed the average distance between each fifth iteration reproduction by all participants in a given group and the closest small-integer-ratio rhythm. This measure should be small for priors with modes that are perfectly centered on small-integer-ratios. We compared this measure to a null distribution obtained by computing the average minimum distance from a random set of points. All groups produced priors that were closer to small-integer-ratios than would be expected by chance ($p < .001$ in all cases, Bonferroni-corrected; Figure 2C). Given the diversity of the participant groups, this result suggests that small-integer-ratio rhythm categories are a universal feature of human perception. This conclusion is further supported by two other measures of the overlap of the priors with small-integer ratios (see Methods).

We obtained similar results in 13 groups who performed the same experiment at a faster tempo (pattern duration of 1000 ms compared with 2000 ms in the main experiment; Figures S1-S2). In all 13 cases priors were closer to small-integer ratios compared with chance ($p < .001$ in all groups; see Figure S1B).

We also found that some of the categories were systematically biased away from the closest small-integer ratio. Of the 22 analyzed small-integer ratios, 9 had corresponding modes that were slightly but significantly biased away from the integer ratio (Figure S3). These biases were present cross-culturally but were similar to those previously observed in Western participants. Specifically, the ratio between the short and medium intervals in the categories 1:2:3 and 2:1:3 (and their cyclic permutations) was consistently larger than 1:2. This lengthening of the short element is characteristic of rhythm performance studied in European musicians (39). A similar bias was present for 2:2:3 (and its cyclic permutations). This result suggests that some specific deviations from small-integer ratios are also a universal feature of perception.

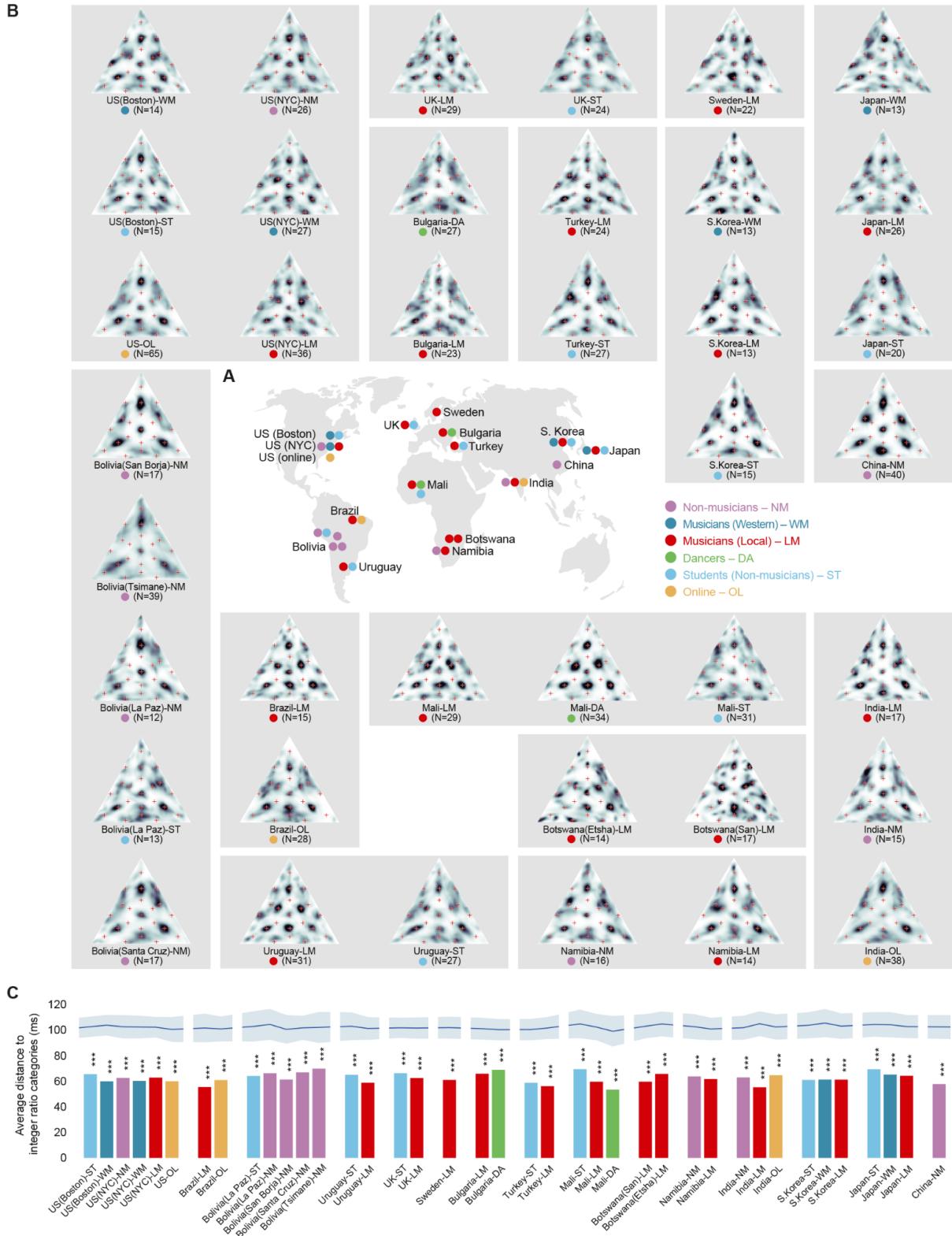


Figure 2. Results of the Global Iterated Rhythm Study. A) Map of test sites. B) Rhythm priors from all 39 groups tested. Musicians are labelled as “Western” if they predominantly play Western music, and “local” if they predominantly play any other style. See Methods for criteria used to classify participants as “musicians” or “non-musicians”. Red + symbols demarcate small-integer ratio rhythms whose duration ratios are defined by integers less than 3. C) Average distance from

nearest small-integer-ratio rhythm, for each participant group. This measure is small for a prior with all its mass at small-integer-ratio rhythms. Shaded region at top plots mean and 95% confidence intervals of a null distribution of the average distance from randomly selected points. Asterisks denote statistical significance relative to this null distribution (**: p<.001; **: p<.01; *: p<.05). Here and elsewhere, p values are Bonferroni corrected for multiple comparisons. All groups have probability mass concentrated closer to integer ratio rhythms than would be expected by chance.

Occurrence and relative importance of specific small-integer ratios varies cross-culturally

Despite the consistent presence of discrete modes that tend to overlap small-integer-ratios, the measured priors varied across groups. To examine the dominant dimensions of variation, we performed multidimensional scaling (40) on the measured priors. Two dimensions of variation captured considerable variance (87.1% of the variance in the inter-group distances measured by Jensen-Shannon divergence).

Figure 3A shows the measured priors for each group arranged according to their position in this space. The four groups lying at the “corners” of the multidimensional scaling space provide a snapshot of the variation across groups, along with features that consistently appear (Figure 3A, insets). All groups show modes at isochrony (1:1:1) and 1:1:2. However, the presence of other small-integer rhythms varies. In particular, the presence or absence of a mode at the 3:3:2 rhythm is a major source of variation, accounting for the horizontal dimension of the multidimensional scaling space (position along this dimension was highly correlated with the weight on the 3:3:2 category; $r=.89$, $p<.001$; Figure 3B&C). The first dimension is also correlated with a reduction in the weight of the “simpler” categories (1:1:1: $r=-.45$ $p=0.03$; 1:1:2 $r=-0.64$ $p<0.001$; $r=-0.61$, $p<0.001$). The second multidimensional scaling dimension is less obviously interpretable, but also correlates significantly with two of the small-integer-ratio rhythms (1:2:3 and 1:3:2; $r=.57$ and $r=.63$, respectively, $p=.001$, $.0001$ after Bonferroni correction; Figure 3C). The groups at the corners also exhibit variation in the presence of 2:2:1, such that all possible combinations of modes at 2:2:1 and 3:3:2 are evident across the Tsimane’, China, Botswana San musician, and Mali dancer groups. We also analyzed the data using principal component analysis (PCA); the first two components captured variation along the same modes highlighted by the multidimensional scaling analysis (see Supplementary Figure S4).

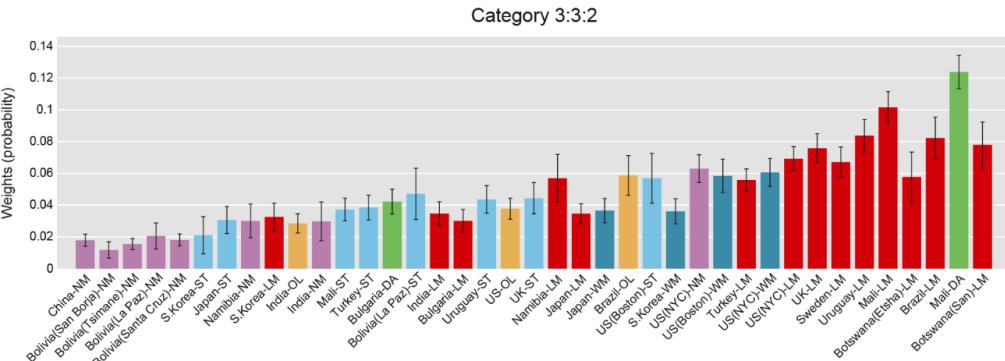
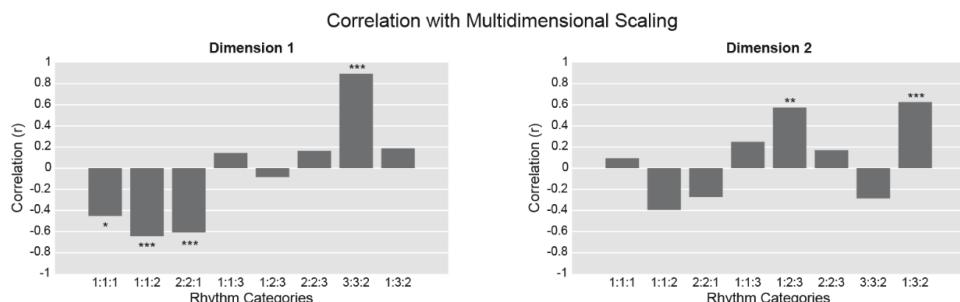
A**B****C**

Figure 3. Cross-cultural variation in mental representations of rhythm. A) Participant groups positioned in two dimensions via multidimensional scaling of their estimated rhythm priors. Group names are colored to indicate musicians, dancers, non-musicians, and university students. Insets show rhythm priors for the four participant groups lying at the “corners” of the multidimensional scaling space. These groups represent the extremes of the variation we observed. Note that for all analyses

of specific rhythms, we averaged across the cyclic permutations of that rhythm (e.g. 1:1:2, 1:2:1, 2:1:1), as the differences between permutations were typically modest. See Supplemental Figure S5 for an analysis of differences between cyclic permutations. B) Strength of the mode at the 3:3:2 rhythm (the average weight of the modes at 3:3:2, 2:3:3, and 3:2:3 in the fitted Gaussian mixture model) for all participant groups, ordered according to their position on the horizontal dimension of the multidimensional scaling space. This rhythm characterizes one dominant dimension of variation across groups. Error bars plot SEM, derived from bootstrap. C) Left: The correlation between the weights given to each of a subset of small-integer-ratio rhythms in a Gaussian mixture model fitted to the prior, and the position along the first (horizontal) multidimensional scaling dimension. Dimension 1 is highly correlated with the strength of category 3:3:2. Asterisks denote statistical significance of the correlations, after Bonferroni correction for multiple comparisons (**: $p < .001$; **: $p < .01$; *: $p < .05$). Right: Same as left panel but for the second (vertical) multidimensional scaling dimension, which is significantly correlated with categories 1:2:3 and 1:3:2.

Cross-cultural variation is much less evident in university students and online participants

The positioning of the groups in the multidimensional scaling space (Figure 3A) also suggests that the priors in university students and online participants in other countries are relatively similar, both to each other and to those of US participants. To quantify this effect in students, we first measured the Jensen-Shannon divergence between the estimated priors for pairs of student groups in different countries and compared it to that for pairs of non-student groups (from the same countries from which the student groups were drawn). This comparison revealed that the distance between student groups was significantly smaller than that of non-student groups ($p = .01$, permutation test; Figure 4B). We conducted an analogous analysis in online participants and obtained a similar result ($p < .001$, permutation test; Figure 4E). This result indicates that both student and online participants underrepresent cross-cultural diversity. This conclusion is supported by the self-reported music listening habits of the participants, which are notably more similar between the student groups, and online groups, compared to the corresponding non-student and non-online groups from the same countries (Figures 4C,D,F&G; Table S1).

To quantify the similarity of these two types of groups to US participants, we measured the average distance between priors of the US student group to those of the other student and online groups by comparing their average distance to that between US students and random sets of groups of the same size (who were not students or online participants). Student and online groups were significantly closer to US participants than were the other groups (students: $p = .0011$, online $p = .015$; via bootstrap).

Although the online participants spanned a wide range of ages, student participants tended to be younger than other groups (Table 1). To control for possible effects of age, we repeated both analyses, restricting participants to the age range of 20-40 (see Methods). The pairwise distance between student and online groups was still significantly smaller than that between non-student and non-online groups (students: $p = .005$, online $p < .001$), as was the distance between US student groups and other student and online groups when compared to that between US student groups and non-student/non-online groups (students: $p < .001$, online: $p < .001$). In addition, we compared priors of younger (under 35) and older (over 35) online participants and found no significant differences in each of the three locations ($p = .09, .09$ and $.39$ for US, India, and Brazilian online groups, respectively).

These findings suggest that certain lifestyle factor(s), such as socioeconomic status or access to global media and internet, involve exposure to globalized culture that is sufficient to produce mental representations of rhythm that are similar to those of Westerners. As a consequence, student populations and online participants (the populations typically studied in psychology, neuroscience, and music cognition (37, 38, 41) underrepresent global variability in music perception (37).

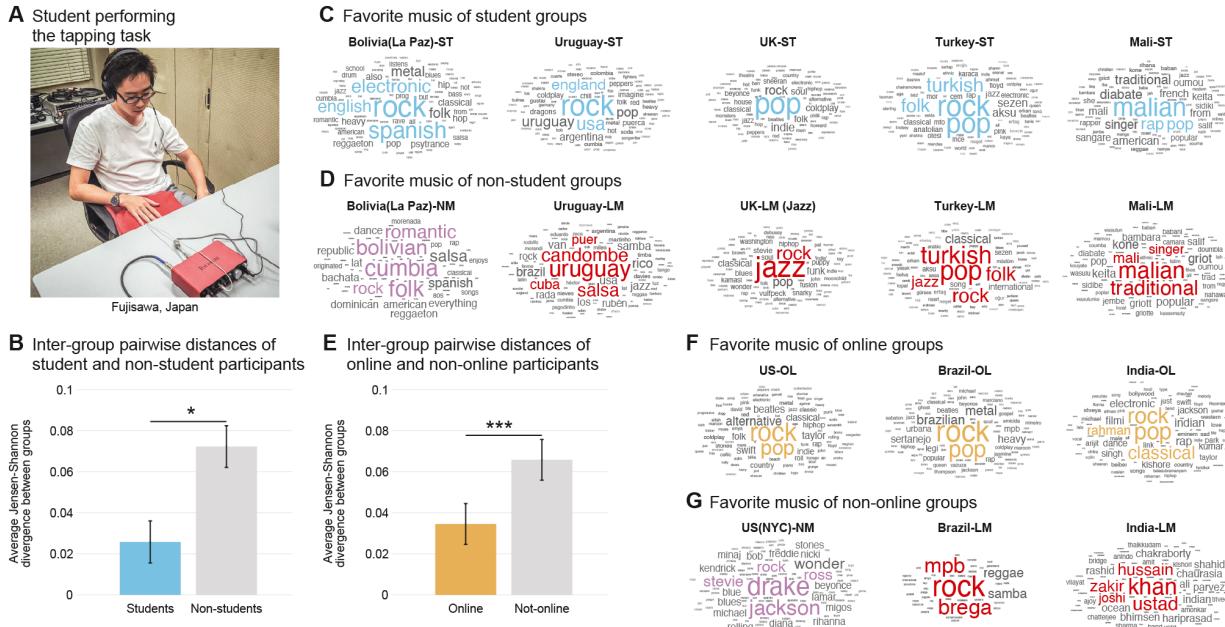


Figure 4 Cross-cultural variation is reduced in university students and online participants **A)** Photograph of a student participant in Japan. **B)** Jensen-Shannon divergence between the estimated priors for pairs of student groups in different countries and pairs of non-student groups in the same set of countries (one for each student group). Error bars plot SEM, computed via bootstrap. Here and in E, asterisks mark statistical significance (***: $p < .001$; **: $p < .01$; *: $p < .05$). **C)** Word clouds of favorite musical genres listed by student groups. **D)** Word clouds of favorite musical genres listed by non-student groups. **E)** Same as B, but for online participants. **F&G)** Same as C&D, but for online and non-online groups.

Culture-specific variation can be linked to the music in a group

What underlies the observed variation across groups? In several cases, variation in rhythm priors had obvious links to rhythms prominent in local musical systems. For instance, the mode at the 2:2:3 rhythm was pronounced in groups that have this rhythm in their local musical tradition (42-44): traditional musicians in Turkey, Botswana, and Bulgaria (Figure 5A-C). As shown in Figure 5A, the weights assigned to the 2:2:3 rhythm in a Gaussian mixture model fit to the prior of each group were substantially higher in these groups than in most other groups ($p = .003$ via a one-sided Wilcoxon rank-sum test). This result is consistent with previous developmental work showing differences in sensitivity to the 2:2:3 rhythm in Western and Turkish adults (45), but not infants (46), suggesting that sensitivity to the rhythm may be lost without exposure.

The 3:3:2 rhythm can also be related to specific musical systems, in this case, those of African and Afro-diasporic music. In particular, it features prominently in many sub-Saharan musical styles (47, 48), is popular under the name of “tresillo” in Afro-Cuban and Latin music, and is characteristic of many Afro-diasporic traditions from Cuba, Brazil, Uruguay, and North America (49-52), among others. In the priors estimated from the experiment, the 3:3:2 mode was strongest in dancers from the Sagele village in Mali (Figure 5D-F). We recorded a representative corpus of their musical repertoire and found that forty-six percent of the excerpts recorded in the corpus featured a prominent 3:3:2 pattern (see Methods for additional details). We similarly found strong 3:3:2 weights in all other groups of musicians and dancers from African and Afro-diaspora traditions (musicians from Botswana, Mali, Uruguay, Brazil, and both US and UK Jazz musicians; Figure 5D; significantly higher weights for this mode when compared to the other participant groups, $p < .001$ via a one-sided Wilcoxon rank-sum test).

Perceptual modes can occur at moderately complex integer ratio rhythms

Another example of an idiosyncratic feature that can be linked to a musical system is the mode at 7:2:3 evident in drummers in Mali (Figure 5G-I). This rhythm is noteworthy because it is defined by a relatively complex integer ratio. We took several steps to establish that this rhythm is prominent in the local musical tradition (see Methods). First, we observed that some participants recognized the rhythm during the experiment session. They named it “Maraka,” which is a popular local piece of dance music that uses a 7:2:3 rhythmic pattern. To validate this recognition of the rhythm in the musician groups we tested, we developed a follow-up experiment that was carried out after the end of the main experimental session with the help of ethnographer team members. We estimated the empirical modes of the distribution, synthesized stimuli corresponding to each of the modes, and played them to participants, asking for their spontaneous response. Participants typically identified the stimulus with a particular musical form/genre, either by naming corresponding pieces, by demonstrating the structure of the rhythm by tapping or singing, or both.

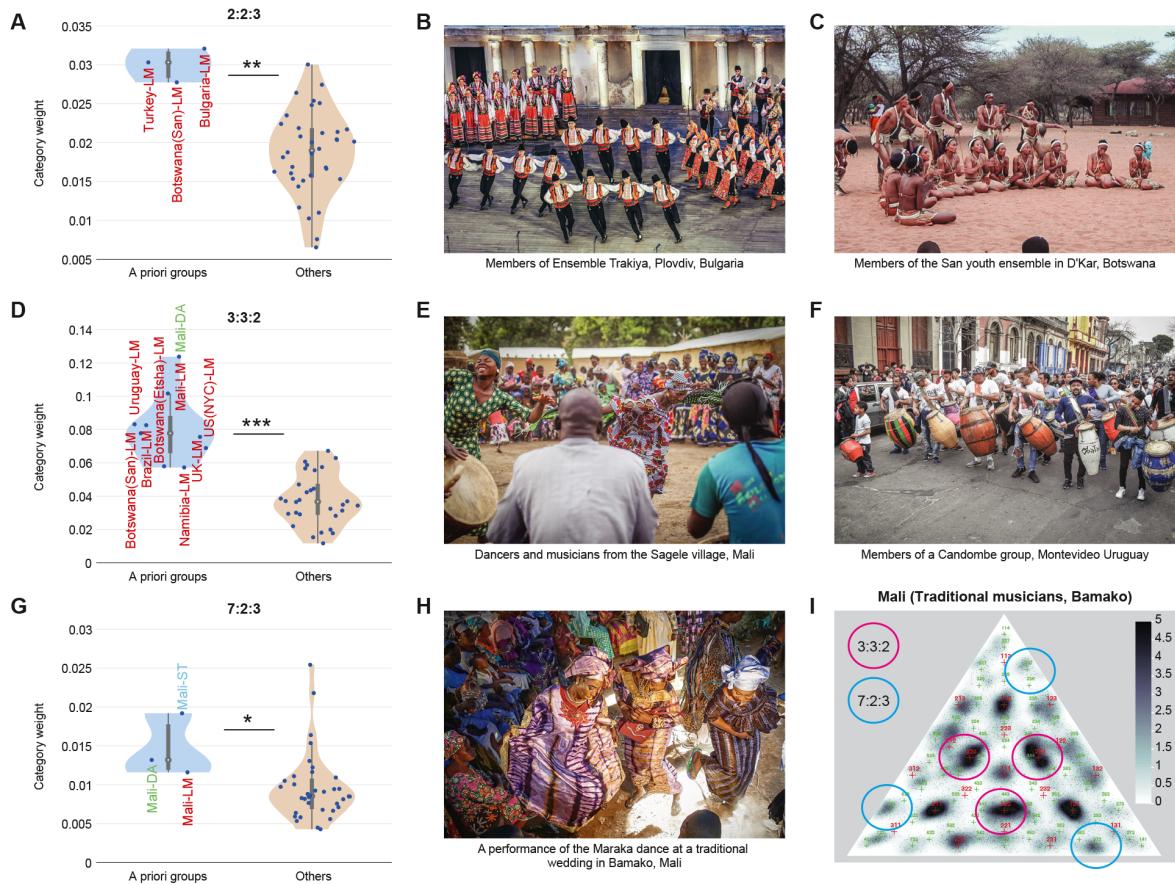


Figure 5. Rhythm priors reflect established culture-specific musical features. A) Violin plots showing the strength of the 2:2:3 rhythm for all tested groups, separated into those in whose music the rhythm is prominent, and all other groups. Here and in D and G, asterisks mark statistical significance (**: p<.001; **: p<.01; *: p<.05). The 2:2:3 rhythm is strongly represented in the priors of traditional musicians in Bulgaria, Turkey, and Botswana, when compared with all other groups. B) Dancers and musicians from Ensemble Trakiya in Plovdiv, Bulgaria. C) Members of the San youth ensemble in D'Kar, Botswana. D) The strength of the 3:3:2 rhythm for all tested groups, separated into those in whose music the rhythm is prominent (i.e. music of African and Afro-diasporic traditions), and all other groups. E) Dancers and musicians from the Sagele village in Mali. F) Members of a Candombe group in Montevideo, Uruguay. G) Strength of 7:2:3 rhythm for all tested groups. H) A performance of the Maraka dance at a traditional wedding in Bamako, Mali. I) Rhythm prior for drummers from Bamako showing modes at 3:3:2 and 7:2:3.

As shown in Figure 5G, the 7:2:3 mode was stronger in the three groups from Mali than in all other groups ($p=.016$ via a one-sided Wilcoxon rank-sum test). This result suggests that even relatively complex rhythms can form perceptual categories, but that they are strongly dependent on musical experience within cultural environments where the corresponding rhythm prevails.

Musicianship improves tapping accuracy but does not obviously alter rhythm priors

We previously found that rhythm priors were similar for musicians and non-musicians in Boston, USA (25), suggesting that priors are driven by “passive” exposure to music rather than explicit training. We replicated this finding here in multiple other groups of musicians who play Western music and corresponding groups of non-musician students with exposure to Western music (in South Korea, Japan, and the USA; no significant difference in each case: $p=.062$, .1, .21; Figure S6). For the sake of brevity, here and elsewhere we classify and refer to participants as “musicians” or “non-musicians” (see Methods), cognizant that the Western concept of musicianship may not fully apply to other societies, and that there can be a continuum of levels of participation in musical activities. We nonetheless found that years of self-reported musical experience differed between groups classified as “musicians” or “non-musicians” (Figure 6A); pairwise comparisons between musician and non-musician groups in the same country were statistically significant in all cases (Figure 6B).

We did find multiple cases where musicians and non-musicians in the same country exhibited distinct priors, but these all seem explainable by differences in musical “diet”. Specifically, local musician groups in Mali, Uruguay, and Turkey differed from other groups in those countries (MA-LM vs. MA-DA $p=.002$, MA-LM vs. MA-ST $p=.004$, UY-LM vs. UY-ST $p=.013$ TK-LM vs. TK-ST $p=.023$; Figure S6), presumably because they have internalized different types of music in the cultural milieu of the tradition they specialize in (see Figure 4C,D,F&G). We thus have no reason to suppose that musicianship per se qualitatively alters mental representations of rhythm. The signature properties of discrete modes at small-integer ratios are clearly evident in both musicians and non-musicians.

Strong effects of musicianship were, by contrast, evident in the fidelity of tapping (Figure 6C). We compared tapping error (quantified as the standard deviation of the asynchrony between stimulus and response) between 16 pairs of musician and non-musician groups in the same country. Tapping error was lower in musicians than non-musicians (Figure 6D) in 13 of the 16 pairwise comparisons ($p<.001$ via Wilcoxon test). This result is consistent with a large literature in Western musicians showing perceptual and production advantages in musical tasks (53-56), including tapping to a beat (3), but provides evidence that these musicianship advantages are present cross-culturally. The results support a distinction between musicians and non-musicians in at least some non-Western societies (26, 57).

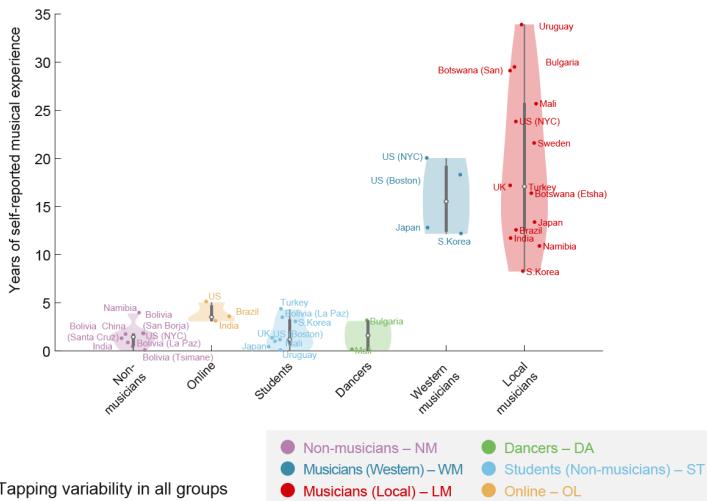
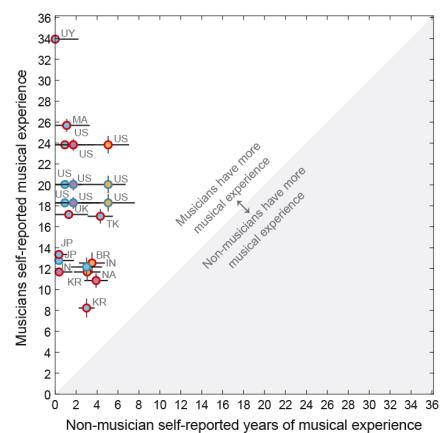
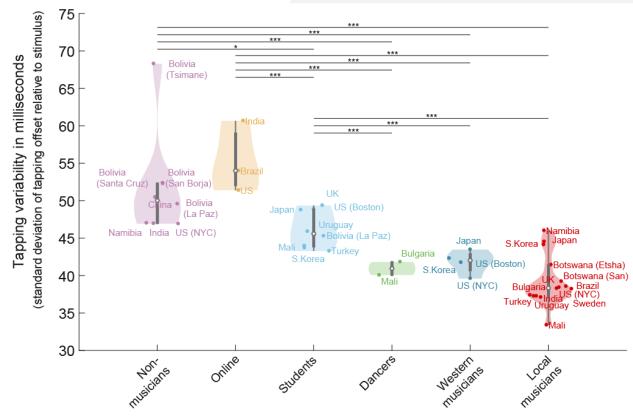
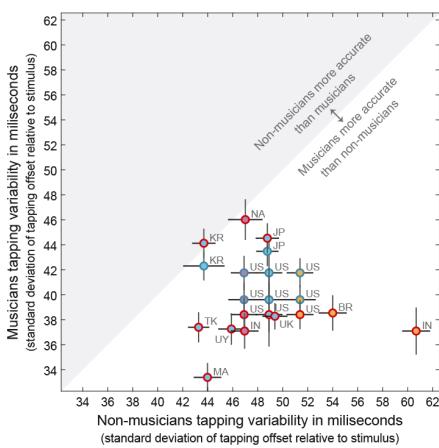
A Musical experience of all groups**B** Self-reported experience in musicians and non-musicians**C** Tapping variability in all groups**D** Effect of musicianship on tapping variability

Figure 6. Tapping variability and musical experience. A) Average self-reported years of musical experience for each participant group. Both traditional and Western musicians self-reported significantly more years of experience than other participant groups. B) Scatter plot of self-reported years of musical experience for musician and non-musician groups from the same country. Each point represents a pair of non-musician (x-axis) and musician (y-axis) groups. C) Tapping error (standard deviation of the asynchrony) as a function of musical experience. Asterisks denote statistical significance of comparisons between musician and non-musician groups (** = <0.001, ** = <0.01, * = <0.05). D) Scatter plot of tapping variability for musician and non-musician groups from the same country. In 13 of 16 comparisons (pairs of musicians and non-musician from the same country) the musicians were significantly more accurate (smaller tapping variability) than the non-musician comparison group ($p<0.001$ in all cases after correcting for multiple comparisons).

No evidence that language influences rhythm perception

It is natural to wonder whether the prior on rhythm might be influenced by experiential factors outside of music, most obviously spoken language. Speech has previously been argued to influence rhythm perception (58), but most recent evidence indicates segregation of speech and music analysis in the brain (59). Several findings from our study suggest that language does not strongly influence rhythm representations. First, we observed clear examples of groups who speak the same language but whose rhythm priors were quite different. The clearest example is the three groups in Mali (local musicians, dancers, and students). All three groups speak a nearly identical set of languages but have obviously distinct rhythm priors (Figure S6A; $p<.015$ for all pairwise comparisons, comparing priors derived from splits of the data within and across groups, via bootstrap). Specifically, both the local musicians and dancers in Mali had significantly higher weights on the 3:3:2 modes than did Malian students. We found a similar result for comparisons of students and local musicians in Uruguay (Figure S6B; $p=.011$, again differing in the 3:3:2 mode). An analogous result was evident in Turkey, where local musicians had

significantly higher weights on the 2:2:3 rhythm that is associated with Balkan and Turkish music, compared to Turkish students (who did not have significant weights for these modes).

Second, we observed several examples of similar rhythm priors despite distinct spoken language experience. Both university students and Western-trained musicians in Japan and South Korea exhibited priors that were not obviously different from those of Westerners (Figure S6D-F; no significant difference in either case: Japanese students and US students $p=.75$; Japanese Western musicians and US Western musicians $p=.12$; South Korean students and US students $p=.26$; South Korean Western musicians and US musicians $p=.42$, comparing priors derived from splits of the data within and across groups, via bootstrap). Although the student groups in these countries undoubtedly had some exposure to English, it was not their native language. We cannot exclude the possibility that language could in some cases have some influence on rhythm representations, but our study provides no evidence for such influences.

Discussion

We conducted the first large-scale cross-cultural study of music perception. Our aim was to assess whether discrete representations of musical rhythm are a universal feature of perception and whether the associated perceptual categories are fixed or dependent on some aspect of life experience. We ran the same iterated reproduction experiment on 39 different groups around the world. We found that all groups tested imposed discrete structure on the continuous space of simple rhythms. Mental “categories” (modes) tended to be small-integer-ratio rhythms, but the categories varied across groups and could often be linked to specific rhythms prominent in the local musical culture. Our study provides unambiguous evidence for universal properties of music perception, but also demonstrates substantial cultural variation linked to culture-specific musical experience.

Universality and cultural specificity

Given that music is present in all known societies but also varies considerably across cultures (13–15, 60), it must result from interactions between biological constraints and culture-specific experience. What do our results reveal about these constraints and experiential influences? The most salient feature of our results is the universal presence of relatively discrete modes within the mental prior over rhythms. We refer to these modes as “categories” because they yield characteristics of categorical perception, biasing the perception of nearby rhythms toward the mode center (25, 36). These discrete categories were present in every group we tested. Discrete mental representations likely help to stabilize musical systems, in that a reproduction of a heard piece of music will be attracted to the discrete categories in the listener’s mind (61, 62). Inaccuracies in musical production are less likely to be heard by the listener, and thus less likely to be transmitted as a song is passed along between individuals (6, 63).

The prevalence of discrete categories across such diverse groups suggests that the conception of music as combinations of discrete elements is not merely an invention of the Western academic tradition, but is instead fundamental to the human experience of music around the world. The quantization of a continuous space into discrete categories allows signals to be represented and stored more efficiently (64, 65). This efficiency gain likely makes patterns composed of the rhythm categories easier to learn and share, which may further aid cultural transmission.

The particular categories we observed were also non-arbitrary. Categories were present at small-integer-ratio rhythms in every group tested, and thus appear to be universal features of human perception (14). This universality could in principle reflect innate perceptual biases favoring integer ratio rhythms. However, it is also clear that any such biases at best only partially constrain adult perceptual systems, because the specific integer ratios that were present as categories varied substantially across

groups. We also cannot fully exclude the possibility that perception simply mirrors the musical systems to which listeners are exposed, which might feature integer ratios for other reasons (e.g., production constraints (66) from the periodic nature of motor behaviors).

We found clear evidence that much of the variation in categories is linked to the local musical systems the participants had grown up with (67) (Figure 5). Moreover, the relevant experience appears to primarily be consumption rather than production of music – musicianship appears to influence priors primarily insofar as it alters musical experience. For instance, musicians who play classical Western music and Western non-musicians exhibited similar priors, presumably because both groups are exposed to similar distributions of music, whereas traditional musicians in some other countries were substantially different from students in cities (e.g., Mali, Uruguay, Turkey). Rhythm categories thus appear to be somewhat flexible and culturally dependent (similar to other domains of perception in which categorical perception has been documented cross-culturally, such as color (68), speech (69) and smell (70)). One difference relative to other examples of cross-cultural categorical perception is that the rhythm categories are typically non-verbal, consistent with the idea that discrete representations do not necessarily depend on linguistic labels.

How might other aspects of culture influence rhythm priors? We found evidence that rhythm priors could be dissociated from language, in that we observed several sets of groups that spoke the same language but exhibited distinct priors. The question of whether language influences rhythm priors at all remains open; our groups were not selected to definitively address this issue, in that language covaried along with music across most groups. It remains possible that groups specifically selected to dissociate language from music might show such an influence. We also note that some of the most complex priors that we measured (those in Malian and Uruguayan musicians, with modes at relatively large integer ratio) were in groups without much formal education (see Table S1).

Small-integer ratios

Small-integer ratios are often proposed to be perceptually favored in various ways as a consequence of their mathematical simplicity (51, 71). We found that the perceptual categories that were common across all groups were indeed the simplest possible in mathematical terms (1:1:1 and 1:1:2). However, we also found instances of categories at relatively complex ratios (e.g., 7:2:3). We note that although 7:2:3 is a complex rhythm from the perspective of mathematical relations, it is relatively simple to generate from a fixed train of 12 pulses grouped hierarchically into groups of 2 and 3 (7:2:3=[2+2+3]+2+3). Consistent with this idea, the five-interval pattern of 2:2:3:2:3 is common in West African music (47, 48). This observation raises the possibility that small-integer ratios tend to arise due to biases for groupings of isochrony rather than simplicity per se. The simple rhythms that are present everywhere (1:1:1 and 1:1:2) could also result from isochrony combined with perceptual grouping, as they are consistent with multiple hierarchical groupings (for example, meters of 4 (1:1:2), 8 (2:2:4), 12 (3:3:6), 16 (4:4:8) etc.).

Lack of diversity evident in university students and online participants

Our results illustrate that perceptual phenomena seen in Westerners may not generalize to other populations, consistent with the growing awareness of the underrepresentation of human cultural diversity in psychology and cognitive neuroscience (37, 38). Our results also suggest that global studies involving convenient participant samples from university communities or online cohorts are similarly likely to underrepresent cultural diversity. We found university students and online participants in non-Western countries to be substantially more similar to US participants than corresponding non-student groups recruited in person, underscoring the problematic reliance on student and online participants in cognitive science. Our results highlight the importance of testing participants from a variety of backgrounds. A wide range of participant groups extending from small-scale societies to villagers to small-town dwellers to urbanites, as well as multiple groups per society with different backgrounds, was

critical to providing a strong test of universality and for revealing the extent of diversity in human perception (38, 72, 73).

Limitations

Due to the practical constraints of our experimental paradigm, our experiments were limited to periodic three-interval rhythms, two seconds in duration (though see Supplementary Figure S1-S2 for results at a faster tempo in a subset of groups). We think it plausible that these simple rhythms are the building blocks for more complicated rhythms (74), but they are nonetheless small in scale relative to many natural musical rhythmic motifs, and to the complex phenomena that are evident in such longer patterns (e.g., meter). It is likely that additional principles govern more extended musical pieces, such as memory constraints, that could be revealed with experiments using longer patterns (63).

Our paradigm is also limited by the absence of melody. The method could be naturally extended to incorporate melody, for instance, by having participants sing back a note sequence defined by both time and pitch intervals (27). Such experiments could address whether priors for pitch and time are independent, and whether rhythm and melody processing are interdependent (75).

Although we interpret the results in terms of perceptual priors, our method relies on participants' reproductions, and thus in principle could reflect production constraints. We previously conducted a series of experiments in Western participants (25) that provided evidence that the results of the paradigm primarily reflect perception rather than production. For instance, priors obtained using vocal responses were very similar to those obtained with tapping, suggesting that the result is not specific to a particular production modality, and in particular to a production domain dominated by periodic motor rhythms (tapping). We also validated the measured priors using a discrimination experiment that did not involve production of any sort. Our results here provide additional evidence that our method reveals a perceptual representation, in that priors in musicians were often similar to those in non-musician groups from the same country, despite the greater precision of musician tapping (Figure 6).

One could also take issue with our selection of participant groups, which were chosen using prior knowledge of local musical cultures to provide a strong test of universality subject to practical constraints. With unlimited resources, one could envision sampling groups uniformly (e.g., across geography or linguistic clades) to have the sampling be independent of any particular hypothesis. However, because of the process of cultural globalization, selecting groups at random would likely produce a prevalence of relatively Western-like participants, and would require a much larger number of participant groups to reveal the sort of diversity evident in our results. And given the recent drastic rise of globalization (72), it is not obvious that uniform sampling would reveal anything fundamental about the diversity of music. Nonetheless, our sampling of groups is clearly not exhaustive and additional groups might reveal additional constraints on rhythm priors.

Finally, previous literature suggests that tempo is an important factor in rhythmic representation (76). To explore the role of tempo we ran the experiment with a faster tempo in 13 of the 39 groups (in 6 of the 15 countries). We found that the results for the fast tempo were qualitatively similar to those in the slower tempo used in the main experiment, but that there were some instances where category weights changed with tempo (Figures S1-S2). This finding is consistent with previous work on tempo dependence in rhythm perception but suggests that our main conclusions are not specific to the particular tempo we studied (26, 77, 78).

Relation to prior work

Experimental efforts to compare aspects of music perception across cultures have revealed striking differences along with commonalities (17-29), but have in all previous cases been limited to comparisons of small numbers of societies. As a result, the existence of universal perceptual

phenomena that might constrain music, as well as the extent of cross-cultural variation in perception, has remained unclear. For instance, across a series of studies, one indigenous society (the Tsimane') was shown to exhibit notable differences in music perception compared to typical Westerners, along with some similarities (24, 25, 27, 28). These studies have left open the extent to which the particular group studied was unusual relative to other world cultures. Our present results show that at least for the rhythm priors studied here, the Tsimane' are not unusually different from Western listeners – other groups in the study were equally distinct. This result suggests that diversity evident with the Tsimane' is likely to be the norm rather than an exception for non-Western participants that are relatively isolated from Western and globalized culture.

Our study provides an example of how it can be possible to conduct cross-cultural experiments at a relatively large scale. Our method is ideal for cross-cultural experiments in that it is largely non-verbal, and simple for participants to understand. It also has the advantage relative to some other methods of making no prior assumption about the structures that might be perceptually important. For instance, small-integer ratios in music have been widely discussed by Western scholars for centuries (79-82), and it is natural to wonder whether this emphasis reflects intellectual biases rather than the nature of perception (83). However, we found small-integer ratios to emerge from the data in each participant group. Small-integer ratios thus indeed appear to be a prominent feature of the way most humans hear music.

Previous attempts to characterize universality in music have relied on corpus studies that analyze collections of recorded music from different societies (13-15). Our approach is complementary to such efforts, in that we directly assess perception rather than cultural artifacts (music performances). Music as a phenomenon involves both perception and production; characterizing both is essential to understanding music. Our method also avoids a weakness of current corpus studies, in that we do not require a Western-trained musical expert to score the results, which likely introduces biases that are difficult to quantify or control for. Advances in machine perception are likely to enable corpus studies that automate the transcription process (84), but biases from the training of such algorithms will likely remain a challenge (85).

Future directions

The cross-cultural differences we observed in rhythm priors raise the question of their developmental trajectory. The simplicity of our method should enable cross-cultural experiments in children that could address this. Previous results have suggested that infants initially possess undifferentiated sensitivity to rhythms that then narrow over development dependent on the musical system one is exposed to (46). Accordingly, priors measured in young children of different societies might be more similar than those in adults. One key question is whether priors are initially uniform, or whether small-integer-ratio modes are evident early on. Such experiments could help reveal the origins of the small-integer ratio sensitivity that we found in all groups we tested.

The large-scale, cross-cultural nature of our study illustrates the value of collaboration between science and the humanities, as well as of global cooperation between research groups (41). To interpret our results it was essential to consult the ethnomusicological literature and to work closely with ethnomusicologists with a deep understanding of specific musical cultures. And the clarity provided by a relatively large number of participant groups around the world is most readily attained by large international teams. The results here suggest that large-scale multi-culture studies with other tasks could clarify similar issues of universality and diversity in other aspects of perception.

Methods

Overview of Procedure

Participants were instructed to synchronize their finger tapping to an auditory stimulus presented over headphones. They first completed a short training session (about 10 minutes in duration) familiarizing them with the apparatus and task (described below). The main experiment consisted of a series of trials, each of which contained 5 iterations.

On each trial we sampled a random seed uniformly from the triangular rhythm space, corresponding to a three-interval rhythm (s_1, s_2, s_3). We then generated a sequence of clicks from the seed by repeating the three-interval seed pattern 10 times. After a few clicks (typically a bit more than one cycle), participants began to synchronize to the click sequence. A Matlab script extracted response onsets from an audio recording of the participant's taps (see "Onset extraction" below). We averaged the inter-response intervals across the 10 repetitions, obtaining an average three-interval response (r_1, r_2, r_3). We defined a valid iteration as one which had a three-tap response for 7 or more of the 10 repetitions and whose average response (r_1, r_2, r_3) was not situated far beyond the region we defined for human-producible rhythms (i.e., did not contain an interval shorter than 285 ms of the overall duration). If the iteration satisfied these two criteria, we set the seed pattern for the subsequent iteration to the response pattern: $(s_1, s_2, s_3) \leftarrow (r_1, r_2, r_3)$. If the iteration was invalid, the data from that iteration was omitted from analysis and the seed remained unchanged. We repeated this process 5 times. Trials with 3 or more invalid iterations were excluded from the analysis. For trials with 2 or fewer invalid iterations, the n^{th} iteration was analyzed as the n^{th} iteration even if the iteration that preceded it was invalid. There was a fixed interval of approximately 4 s between iterations within a trial, and a fixed interval of approximately 9 s between trials (both varied by up to 200 ms in either direction due to slight variation in computer systems across sites).

The number of trials that could be run in an experimental session varied depending on the location and is reported in Table S1. In 13 locations we also performed an additional experiment with a faster tempo (pattern duration of 1000 ms). This additional experiment was always performed after the main experiment (with pattern duration of 2000 ms). The demographic information for this experiment is provided in Table 2 and Table S2.

Apparatus

Data were collected with between 1 and 4 computerized stations for each testing location. Each station included a computer, a Focusrite Scarlett 2i2 USB sound card, two sets of Sennheiser HD 280 Pro headphones, and a tapping sensor (Figure 1). This design is identical to that reported in Jacoby and McDermott (2017) (25). The tapping sensors were covered in a soft cloth that minimized the sound of the participant's tapping. Each sensor contained a microphone hidden within sound isolation materials and was further covered with a fluffy blanket for additional isolation. The microphone in each sensor was highly sensitive, and light touches generated bursts of noise that were recorded by the microphone. The sound card simultaneously recorded the microphone output and the audio stimulus played out by the participant's headphones, so that the latency of the response recording relative to the stimulus was nearly eliminated (less than 1 ms). The specification of the hardware and instructions for building the sensor are provided in the OSF repository associated with this paper (see Data Availability).

Stimuli

The stimulus on each trial was a rhythmic pattern composed of short percussive sounds (bursts of white noise) 65 ms in duration with an attack time of 5 ms (linear ramp) and that decayed gradually over the remaining 60 ms, with the envelope hand-designed to mimic common percussion instruments. These patterns contain 10 repetitions of a particular sequence of three-interval rhythms. The stimuli

were identical to those used in Jacoby and McDermott (2017). Software to replicate the experiment and sound material can be found in the OSF repository associated with this paper (see Data Availability).

Three-interval rhythms

Each stimulus was defined by a pattern of three intervals (s_1, s_2, s_3) constrained such that the overall pattern duration was 2000 ms: $s = s_1 + s_2 + s_3 = 2000$. In addition, to avoid rhythms that were too fast for humans to reproduce, we constrained the initial seeds so that the smallest interval was larger than 300 ms ($(s_1, s_2, s_3) > 300$). We then repeated the interval pattern 10 times, thereby forming a sequence of 30 intervals $\{S\}_{1 \leq i \leq 30} = (s_1, s_2, s_3, s_1, s_2, s_3, s_1, \dots, s_3)$. From this sequence, we created a sequence of 31 onsets $\{O\}_{0 \leq t \leq 30}$ with intervals corresponding to S . The 31 onsets (“clicks”) were defined with respect to the initial onset O_0 , $O_t = O_0 + \sum_{1 \leq i \leq t} S_i$ for $1 \leq t \leq 30$.

Projection to the rhythm triangle

We mapped a three-interval rhythm with inter-onset intervals (s_1, s_2, s_3) to a point in a triangular rhythm space spanning all linear combinations of three extremal rhythms : $\frac{s_1}{s} \vec{P}_1 + \frac{s_2}{s} \vec{P}_2 + \frac{s_3}{s} \vec{P}_3$ where $s = s_1 + s_2 + s_3$, and where \vec{P}_i are the vertices of the triangle (a simplex) (Figure 1B; see (9, 25)). For visualization we used an equilateral triangle, with $\vec{P}_1 = (0,0)$, $\vec{P}_2 = (1,0)$, $\vec{P}_3 = (1/2, \sqrt{3}/2)$. Note that since the initial seed intervals satisfied the constraint that $(s_1, s_2, s_3) > 300$ □□, initial seeds were located within an inner triangular region with vertices $(f,f), (1-2f,f), (\frac{1}{2}, \frac{\sqrt{3}}{2}-f)$, where $f=300/2000$ (see Figure 1B, inner region).

Onset extraction

We processed the microphone recording from each trial in non-overlapping windows of 15 sec, detecting all samples exceeding a relative threshold of 1.45% of the maximal power of the recorded waveform in the window. This threshold was slightly more sensitive than the 2.25% threshold used in Jacoby and McDermott (2017); this change was made to accommodate a small number of participants (less than 3%) who tended to produce very light taps. In most cases, onsets were detected with minimal errors (as evaluated by comparing the detected onsets to what was audible from listening to example trials). We nevertheless took several steps to ensure that the detected onsets corresponded to actual taps. First, we discarded onsets that were too close to one another (less than 80 ms apart) as humans generally cannot produce two taps in such close proximity (see Repp (2005) (3) for a justification of this threshold). Second, we discarded responses that were too far from any stimulus click, regarding them as errors. Here we took into account an important characteristic of human tapping known as “negative mean asynchrony”, namely that tapping in time to a beat tends to be biased compared to the beat onset (typically occurring before the beat) (3). We first matched each onset to its closest stimulus click and computed the mean asynchrony m as the average difference between a response and its corresponding stimulus click. We then excluded all events such that $|R - m| > 300$ ms (namely, a window of 300 ms centered around the perceptual center defined by the mean asynchrony) from further analysis.

When participants tap, they sometimes omit one of the taps within one or more of the 10 repetitions of the pattern. In such cases, we simply averaged the taps that were successfully produced and detected (because each tap was associated with a stimulus click, it was straightforward to average the tap onsets corresponding to each of the three clicks). However, we only did so if there were at least 3 successful taps per stimulus position (i.e. at each of the first, second and third click of the stimulus pattern); if not the iteration was counted as invalid and discarded (and the seed from the previous iteration was kept unchanged).

Apart from the change in the detection threshold mentioned above, this procedure was identical to the one reported by Jacoby and McDermott (25). Code for the procedure is provided in the OSF repository associated with this project.

Training of participants

Participants were asked to position the sensor in a comfortable way (see OSF repository for the full instructions). In most cases participants preferred to sit on a chair, positioning the sensor on their lap and using their entire hand or one finger for tapping. However, we allowed for different postures in different groups. For example, some participants in Mali preferred to sit on the floor with their legs stretched out in front of them (Figure 1A). In all cases, participants were encouraged to change posture or switch hands if they were fatigued, however they were not permitted to tap with two hands simultaneously or to tap with two alternating fingers. In cases where participants nevertheless tried to do so, the experimenter would stop the experiment and repeat the instructions. The experiment was conducted using a fixed set of steps:

1. The first step of the experiment was to ask the participant to tap “a steady beat” at any tempo that they like. The idea of this task was to familiarize the participant with the sensor, as well as to characterize the spontaneous motor tempo of the participants (86). This task was generally easy for participants, however, the concept of “steady beat” varied slightly across participant groups. For some participants in Mali and Uruguay, a “steady beat” was a non-isochronous rhythmic pattern rather than an isochronous beat. In cases where participants continued to tap a rhythmic pattern rather than an isochronous beat, the experimenter repeated the instruction. In some cases, participants did not change their behavior according to the repeated instructions and continued to tap a non-isochronous pattern; in these cases, we did not stop the experiment again and proceeded to the next step and recorded their performance as-is.
2. Participants were then asked to “tap faster”.
3. Next, participants were instructed to tap “as fast as they can” for a couple of seconds. This step aimed to test that the participant did not have any severe motor constraint that would limit them in performing the task. After the participant tapped for about 3-6 seconds, the experimenter thanked the participant and stopped tapping to avoid fatigue.
4. The next step was to familiarize the participant with the stimulus. The experimenter played a few isochronous clicks (inter-stimulus interval of 800 ms) over the headphones and asked the participant to report if the sound was too loud or too soft. The level of the sound was then adjusted until the participants felt that the sound was at a comfortable level.
5. Participants were then instructed to tap along to an isochronous beat (with inter-stimulus-interval of 800 ms). This step could be repeated multiple times until the experimenter felt that the participant understood the task and was able to provide a synchronized response. Note that this was not easy for all participants, for example, some participants naturally tapped in antiphase (“offbeat”; halfway in between the beat). In case of difficulties, the experimenter would first check that this was not caused by the participant’s posture, in which case the experimenter would suggest the participant change their position to enable easier tapping. In other cases, the experimenter would demonstrate synchronous tapping to the participant. These additional steps allowed nearly all participants to successfully perform isochronous synchronous tapping. The rare participant who could not successfully perform isochronous tapping did not continue to the main experiment.
6. We then performed an isochronous tapping task in which participants tapped to a sequence of 56 clicks lasting 44 seconds with ISI = 800 ms.
7. We next performed an additional isochronous tapping task with 75 clicks lasting 45 seconds with ISI = 600 ms.
8. Finally, we performed a tempo-changing tapping task in which the inter-stimulus interval alternated between 546 and 654 ms every 8-13 clicks (chosen pseudorandomly), lasting 74

- clicks / 45 seconds (the exact tapping sequence, identical for all participants, can be found in the OSF repository, see below).
9. In the next step of the experiment, participants were told that they would now tap a rhythm. The experimenter emphasized that, as before, the participant needed to “tap once for every click that they hear”. Participants were given 5 trials of random three-interval rhythms (each with 5 iterations) to get them familiarized with the task. The experimenter provided feedback to the participant only if they were performing the task in a qualitatively incorrect way, such as tapping on the off-beats, or omitting a beat.
 10. Participants performed 10-30 trials of the tapping experiment (each containing 5 iterations, where each iteration contained 10 repetitions of the three-interval pattern, as described above). Due to the long duration, the participants were informed that they could ask for a short break at any time, and the experimenter included additional breaks at various times during the experiment.
 11. Participants answered a set of demographic questions (see OSF repository for full list) during one of the breaks and/or at the beginning/end of the experiment.
 12. In some locations we performed an additional experiment after the completion of the main experiment, in which we repeated the main experiment with an overall pattern duration of 1000 ms.

Procedures for replicability across sites

Testing stations were created by local research team members according to a set of specifications (see OSF repository), or created according to the same specifications by the first author and sent to teams in different locations. A written procedure describing the process of hardware preparation, software installation, task instruction and participant training was delivered to each group (see OSF repository for the details). Task instructions were translated into local languages. Experimenters were either highly fluent in the local language or accompanied by a translator that was a native speaker. In most locations, the data collection team included an anthropologist or an ethnomusicologist with an expert understanding of the local culture, social groups, and music. To ensure that the same procedures were used across sites, all teams were walked through the procedures by the first author, either directly or via video conferencing. Pilot data were collected and analyzed at each site, and the first author inspected the quality of the data before the collection of additional data. To assist this process, the Matlab script generates images with small file size (about 150 KB) that summarize the main statistics of data collection (the validity of the trial, mean and standard deviation of tapping asynchrony, indicative of tapping accuracy, and plots showing the microphone recording levels). The same script also generates small binary files (about 4 KB) with summaries of the data (onset times for stimuli and responses). These files were sent to the first author via low-bandwidth internet from remote data collection sites, which assisted in troubleshooting technical and data collection errors.

Testing conditions

Experiments in the US (Boston and NYC) were run in sound-attenuating booths. Elsewhere, where possible, experiments were run indoors in rooms without other activities (Brazil, Uruguay, UK, Sweden, Bulgaria, Turkey, Mali-Bamako, India, South Korea, Japan). When run outdoors, the experimenter chose a location away from community activities that was relatively free of distractions and noise (Bolivia, Mali-Sagele, Botswana, Namibia and China).

Online measurement of tapped responses

To run the experiment online we used REPP (87), a software package for measuring sensorimotor synchronization in online experiments that works efficiently using hardware and software available to most online participants. To achieve temporal accuracy superior to that obtained by a web-browser

(which would have been inadequate for our experiment (88, 89)), the software plays the audio stimulus through the participant's laptop speakers and records the original signal simultaneously with participants' responses (which they supply by tapping on the laptop case) using the built-in laptop microphone. The resulting recording is then analyzed to extract the participant's taps. The method has been validated in a series of calibration and behavioral experiments (87), and achieves high temporal accuracy (latency and jitter within 2 ms on average). In addition, the method has been shown to provide results that are consistent with those obtained in a laboratory setup (e.g., for isochronous tapping the lab-online correlation for tapping accuracy of individual participants is $r = .94$; See Experiment 2 in (87)).

To measure stimulus-coordinated tapping by online participants, we made some modifications to the original paradigm. One set of differences involved the stimulus. First, because the recorded audio contains both the stimulus and response, we filtered the stimulus with a high-pass filter (cut-off frequency of 1000 Hz) in order to avoid overlap with the frequency range typically occupied by participants' tapping responses (80-500 Hz). Next, we added three custom audio "markers" with known temporal locations at the beginning and end of each stimulus (six in total). These markers enabled us to unambiguously identify the position of the stimulus onsets in the recorded audio, and facilitate precise measurement of participants' asynchronies. The markers were designed to be robustly detected across a variety of hardware and software setups, including cases of noise-cancellation technologies and ambient room noise. The marker sounds were generated from 15 ms bursts of bandpass filtered white noise in the range of 200-340 Hz, to which we applied linear ramps at the onset and offset (2 ms in length). We choose a very short ISI between the markers—280 ms and 230 ms for the first and second interval, respectively—in order to avoid participants confusing the markers with the repeated rhythm (the rhythm pulses and the markers also differed substantially in timbre due to the different frequency ranges).

A second set of differences involved the response recording. The online experiments used free-field recording whereby the audio stimulus is played through the laptop speakers and simultaneously recorded along with the participant's tapping response using the built-in laptop microphone. This returns an audio file where both the audio stimulus and participant tapping are superimposed. To separate this recording into the different components of the stimulus and response, we used bandpass filters. Since most of the energy in the tapping signal occurs at low frequencies, filtering the recording around the tapping range (80-500Hz) isolated the tapping response with a high degree of efficacy. We were also able to isolate the markers by filtering in their frequency range (200-340 Hz). In addition, we applied a filter in the 100-170 Hz range, the output of which was used for calibration (because the markers had no energy in this range, this helped to determine the noise level in the recording, which we used to adaptively set the marker detection thresholds. This allowed us to reliably estimate the marker locations even with very noisy or low-quality recordings, as characterize some laptop models and brands.

The detected stimulus markers were used to estimate and compensate for the latency of the recording, and to estimate the jitter in the recording. This enabled us to monitor the timing accuracy of each individual trial, crucial to ensure that timing accuracy remained high in all trials. See Experiment 1 in Anglada-Tort et al. (87) for full details of the calibration experiments and their validation.

After the online-specific pre-processing to isolate the tapped response and compensate for the recording latency, we used the same pipeline as in the main experiment in order to align the tap onsets and perform the tapping analysis.

Online iterated reproduction procedure

To meet the challenges of online data collection, such as poor control over participants' hardware and software and a higher risk of fraudulent responders, we made two minor changes to the iterated reproduction procedure.

1. In the online version of the experiment, the analysis of the recording may take a few seconds to complete (e.g., from uploading the audio recording, performing the signal processing, synthesizing the new stimulus etc.). To avoid unnecessary wait times, we did not run one iterated reproduction chain at a time, as in the laboratory. Instead, we ran 6 chains in parallel. On each trial, participants were randomly assigned to one of the chains that was not used for the previous trial. Each participant completed all six chains during an experimental block. In Jacoby and McDermott (2017; Experiment 4) (25), we showed that the results of this parallel chain procedure do not differ substantially from those of the original paradigm, where the iterations from different chains are not intermixed.
2. The change in trial order also necessitated a change in the failing criteria: if participants failed a trial, they repeated the trial with a maximum of 10 possible additional attempts, but these repeated trials were randomly drawn from the 6 chains (the 10 allowed failed trials were tallied globally within the block). In those cases where this limit was exceeded, the experiment was terminated.

In addition, the following seven modifications were made to the overall procedure to ensure that online participants met the technical requirements for the online experiment, and were able to provide good tapping data consistently throughout the experiment. The different steps in this procedure have been extensively piloted and optimized to ensure high data quality when collecting tapping data in online settings (87):

1. First, participants were instructed that the experiment could only be performed using the laptop speakers and that they should unplug any headphones/earphones or disconnect any wireless devices. They were also instructed to remain in a quiet environment.
2. Participants were then asked to set the volume of their speakers to a level that was sufficiently high to be detected by the microphone. A sound meter was used to visually indicate when the level was appropriate.
3. After the volume test, participants completed a short recording test to detect hardware and software that did not meet the technical requirements of the experiment, such as malfunctioning speakers or microphones. The recording test played a test stimulus with six marker sounds. The markers were recorded with the laptop's microphone and analyzed using our signal processing pipeline. During the playback, participants were supposed to remain silent. There were a total of three such test recording trials, and we provided feedback after the first trial based on the recording quality: if the markers were not recorded (for example, this could occur if the participant forgot to unplug their headphones) we reminded participants that they need to unplug any headphones. If, despite these reminders, marker sounds could not be detected in two of the three trials the participant was excluded from the experiment. Note that this process also serves as a basic test of task compliance, as participants must follow the instructions (e.g., accept the enabling of the microphone in the browser, unplug any headphones or wireless devices, adjust the volume of the computer) in order to pass the test.
4. Participants who passed the recording test were then directed to a tapping calibration test. Here, participants were asked to tap on the surface of their laptops with their index finger to test whether the microphone could detect their taps, using a sound meter to visually provide feedback. In cases where the signal was too low, participants were asked to tap on different locations of the laptop or to try to tap harder.

5. Next, participants performed a practice phase to acquaint themselves with the main tapping task. The practice phase consisted of four trials using the stimulus sampling procedure of the main experiment (e.g., three-interval rhythms randomly sampled from the triangular simplex with a fixed duration of 2,000 ms and repeated 10 times). In the first two trials, we provided feedback to participants based on their recording quality and tapping performance. In the remaining two trials, we excluded participants who were still unable to provide good tapping data. In both cases we used two exclusion criteria. First, we excluded all trials in which we could not detect all marker sounds, or where the markers were displaced relative to each other for more than 15 ms. Second, we excluded all trials where the percentage of detected taps (i.e., the number of detected tapping onsets out of the total number of stimulus onsets) was less than 50% or more than 200%. Note that none of these criteria excluded trials based on actual accuracy in replicating the target rhythm, but only based on whether the signal could be correctly recorded and processed, and whether participants produced a minimally/maximally acceptable number of tapping responses.

6. Participants who passed the practice phase were then able to start the main tapping task, which used the same procedure described for the in-person experiments with the exception of the modifications mentioned above.

7. As mentioned above, a main difference between the in-person and online experiments was that the latter consisted of shorter and more flexible experimental sessions. Namely, the experiment was divided in different blocks of 6 chains per block. After completing one block, participants could decide whether to continue with the next block or to instead end the experiment. There was a maximum of 3 blocks per session (18 chains). The motivation behind this design choice was to keep online experimental sessions engaging and short, always allowing participants to decide whether to take more trials or not.

After completing the first block, participants answered the same set of demographic questions used in the in-person experiments.

Criteria for group selection

We tested 39 participant groups spanning five continents and 15 countries (Table 1). Overall, we recruited 923 participants (792 were run face to face and 131 online) who completed a total of 20,287 trials (seeds), and 2,319,095 individual taps. Participant groups were chosen to provide a strong test of any potentially universal features of the results. We included groups from both industrialized and non-industrialized societies, as well as groups of local musicians from non-Western societies. We also tested groups of musicians and dancers where possible, as these populations would be expected to have substantial exposure to particular musical styles. In addition, we tested university students in a number of countries in order to assess potential effects of Westernization. The groups tested were also determined in part by practical constraints (testing time and access to particular populations). Age and gender could not be fully equalized across groups. For example, Malian professional jembe drummers and Uruguayan candombe drummers (the populations recruited for MA-LM and UY-LM) are both relatively small groups—less than 50 individuals—composed of highly skilled professionals, and were predominantly male. In both cases, only one participant in each group was female. The significant experience required for membership in these groups also resulted in these participants being relatively older (Mali: mean age = 40.5 years, std= 11.9; Uruguay: mean age = 45.5 years, std= 12.8). At the other extreme, dancers in the Sagele village in Mali (MA-DA) were predominantly female.

Sample sizes

We conducted a power analysis using data from US participants collected for a previous publication (25). The approach was to try to collect enough data that the test-retest reliability of the estimated prior

for a group was likely to be relatively high (with the goal of having enough data that the results of the experiment would be similar in a hypothetical future replication). The test-retest reliability was estimated using the split-half reliability of our previously collected data following Spearman-Brown correction. We simulated different amounts of data by subsampling the number of trials used to estimate the prior (resampling without replacement). We found that 250 trials produced a test-retest reliability greater than 0.8, and thus targeted this number for the sample size of each group. In practice, we often ran more participants if circumstances permitted (between 261-948 and therefore up to 3.8x the target number of trials; see Table S1). The only exceptions were the two groups in Botswana, for whom we did not reach this recruitment target because of practical constraints on testing time (BO.SA=170 and BO.EA=127 chains). However, the post-hoc reliability of the data collected for these groups was not that far below our target value (0.75 and 0.67 for BO.SA and BO.EA respectively). The post hoc reliabilities of all other groups were high, meeting or exceeding the predicted value of 0.8 (ranging from 0.8-0.96; mean 0.9, std=0.03).

Definition of group types

Students: We defined students as members of local universities in either undergraduate or graduate programs.

Musicians. For brevity, we used the term “musicians” to describe participants with relatively extensive musical experience, acknowledging that musicianship is a concept that changes from place to place (41). Based on previous work, we defined three recruiting criteria that generalize more broadly for different cultural contexts (26): 1) professional: do you make most or part of your living from music? 2) training: did you undergo music training (such as an apprenticeship or formal study)? 3) identity: do you see yourself as a musician? In some locations (in Uruguay, Mali, US, UK, Bulgaria, Turkey, S. Korea, Japan) all musicians satisfied all criteria, while in other locations we also included people who satisfied the first two criteria but not the last one. For all participants, we recorded self-reported years of regularly playing an instrument or singing, as is common in music cognition literature (Table S1).

Dancers. In Bulgaria, we recruited dancers who were members of the same professional ensembles from which we recruited the musician groups. In Mali, we recruited group members of a local recreational dance association that promotes events featuring traditional dance and music.

Non-musicians. Non-musicians were people who did not satisfy any of the inclusion criteria for the other groups. Their self-reported years of musical experience was substantially less than the musician groups in all cases (Table S1).

Online. Online participants were recruited from Amazon Mechanical Turk. Their geographical location was determined by the Amazon qualification system and verified with IP geo-location. We tested participants from the three countries (US, Brazil, and India) with substantial worker pools that overlapped with the 15 countries from which the other participant groups were drawn (90).

Participants

All participants provided informed consent in accordance with the Ethics Council of the Max Planck Society, Columbia University Institutional Review Board, KAIST Institutional Review Board, and the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects.

Table 1. Demographic information. Types: ST = student; NM = non-musician; WM = Musicians who play Western classical music; LM = musicians who play a local musical style that is not Western classical music ; DA = dancer; OL = Online participants.

#	Group Name	Group abbreviation	Type	Continent	Country	City/Cities	N	Females	Ages
1	US (BOSTON)-ST	US.BO	ST	North America	USA	Boston	15	8	18-37 (mean = 25.93 std= 6.55)
2	US (BOSTON)-WM	US.BO	WM	North America	USA	Boston	14	8	19-61 (mean = 29.36 std= 11.28)
3	US (NYC)-NM	US.NY	NM	North America	USA	NYC	26	15	18-61 (mean = 38.77 std= 13.00)
4	US (NYC)-WM	US.NY	WM	North America	USA	NYC	27	15	18-64 (mean = 30.70 std= 11.11)
5	US (NYC)-LM	US.NY	LM	North America	USA	NYC	36	4	18-64 (mean = 36.56 std= 13.40)
6	Brazil	BR	LM	South America	Brazil	Recife	15	7	16-40 (mean = 28.13 std= 7.85)
7	Bolivia (La Paz)-ST	BO.LP	ST	South America	Bolivia	La Paz	13	7	21-31 (mean = 25.31 std= 2.78)
8	Bolivia (La Paz)-NM	BO.LP	NM	South America	Bolivia	La Paz	12	10	18-58 (mean = 34.83 std= 11.50)
9	Bolivia (San Borja)-NM	BO.SB	NM	South America	Bolivia	San Borja	17	13	16-51 (mean = 29.53 std= 10.12)
10	Bolivia (Santa Cruz)-NM	BO.SC	NM	South America	Bolivia	Santa Cruz	17	7	17-45 (mean = 27.29 std= 7.24)
11	Bolivia (Tsimane')-NM	BO.TS	NM	South America	Bolivia	Tsimane' villages	39	8	16-49 (mean = 28.33 std= 9.83)
12	Uruguay-ST	UY	ST	South America	Uruguay	Montevideo	27	14	21-32 (mean = 24.56 std= 2.64)
13	Uruguay-LM	UY	LM	South America	Uruguay	Montevideo	31	1	18-74 (mean = 45.48 std= 12.83)
14	UK-ST	UK	ST	Europe	UK	Durham	24	19	20-36 (mean = 24.00 std= 3.87)
			LM	Europe		Durham/ Newcastle/ Edinburgh	29	2	19-59 (mean = 29.90 std= 13.05)
15	UK-LM	UK		Europe	UK		22	11	19-64 (mean = 34.45 std= 14.44)
16	Sweden-LM	SE	LM	Europe	Sweden	Stockholm	23	3	20-59 (mean = 41.48 std= 10.51)
17	Bulgaria-LM	BG	LM	Europe	Bulgaria	Pleven/Plovdiv/Sofia	27	12	19-53 (mean = 33.63 std= 9.08)
18	Bulgaria-DA	BG	DA	Europe	Bulgaria	Pleven/Plovdiv/Sofia	27	18	18-28 (mean = 20.67 std= 2.09)
19	Turkey-ST	TR	ST	Europe	Turkey	Istanbul	24	7	22-44 (mean = 31.75 std= 6.19)
20	Turkey-LM	TR	LM	Europe	Turkey	Istanbul/ Izmir	31	15	19-33 (mean = 25.74 std= 3.02)
21	Mali-ST	ML	ST	Africa	Mali	Bamako	29	1	24-69 (mean = 40.52 std= 11.92)
22	Mali-LM	ML	LM	Africa	Mali	Bamako	34	31	18-62 (mean = 33.24 std= 11.94)
23	Mali-DA	ML	DA	Africa	Mali	Sagele	17	13	22-69 (mean = 44.50 std= 13.32)
24	Botswana (San)-LM	BW.SA	LM	Africa	Botswana	San	14	7	22-61 (mean = 28.93 std= 10.20)
25	Botswana (Etsha)-LM	BW.ET	LM	Africa	Botswana	Etsha	16	9	18-46 (mean = 30.88 std= 8.45)
26	Namibia-NM	NA	NM	Africa	Namibia	Spitzkoppe	15	10	18-48 (mean = 27.79 std= 8.68)
27	Namibia-LM	NA	LM	Africa	Namibia	Spitzkoppe	15	3	20-44 (mean = 27.13 std= 7.30)
28	India-NM	IN	NM	Asia	India	Mumbai	17	6	17-49 (mean = 28.75 std= 9.10)
29	India-LM	IN	LM	Asia	India	Mumbai	15	6	20-24 (mean = 22.33 std= 1.45)
30	S. Korea-ST	KR	ST	Asia	S. Korea	Daejeon	13	11	18-25 (mean = 21.54 std= 1.81)
31	S. Korea-WM	KR	WM	Asia	S. Korea	Jeonju	13	4	20-25 (mean = 22.69 std= 1.80)
32	S. Korea-LM	KR	LM	Asia	S. Korea	Tokyo / Fujisawa	20	5	18-23 (mean = 20.80 std= 1.47)
33	Japan-ST	JP	ST	Asia	Japan	Tokyo / Fujisawa	13	9	18-28 (mean = 21.15 std= 2.51)
34	Japan-WM	JP	WM	Asia	Japan	Tokyo/ Fujisawa	26	13	18-25 (mean = 20.69 std= 1.59)
35	Japan-LM	JP	LM	Asia	Japan	Guizhou Province	40	36	20-51 (mean = 34.70 std= 9.76)
36	China-NM	CN	NM	Asia	China	online	65	27	18-67 (mean = 35.48 std= 10.77)
37	US-OL	US-OL	OL	North America	US	online	38	7	19-51 (mean = 31.25 std= 8.50)
38	Brazil-OL	BR-OL	OL	South America	Brazil	online	28	11	21-58 (mean = 33.45 std= 8.09)
39	India-OL	IN-OL	OL	Asia	India	online			

Recruiting locations

Basic demographic information for each participant group is provided in Table 1. Table S1 provides additional information about each group including the number of languages spoken, the languages spoken, years of self-reported musical experience, instruments played, and favorite artists or musical genres. Here we provide additional information about each group, ordered according to their geographical location.

USA

Boston – Students and Western Classical Musicians (US.BO-ST, US.BO-WM)

The student participants were students from local universities, recruited using the MIT Brain and Cognitive Sciences Department participant mailing list, and through additional online advertisement. The musician group was recruited from the same departmental mailing list as well as via a social media ad targeting conservatory students from the Boston area. Some of the participants were professional musicians. Most participants in the musician group played Western classical music, though some also played other styles such as pop and jazz.

NYC – Non-musicians, Western Classical Musicians, and Jazz Musicians (US.NY-NM, US.NY-WM, US.NY-LM)

NYC participants were recruited by word of mouth, campus advertisements at Columbia University, and online advertisements. We recruited three groups: non-musicians, musicians specializing in Western classical music (the WM group), and jazz musicians (the LM group). Both musician groups were a mix of music students and professional musicians.

Bolivia (BO.TS-NM)

Tsimane’ – Non-musicians: The Tsimane’ are an indigenous people of lowland Bolivia, comprising about 17,000 individuals who live in about 120 small villages located around the Maniqui river in the Beni department. They subsist mostly on hunting, fishing, and farming. Tsimane’ have traditional indigenous music, familiarity with which varies across individuals. As reported by Riester (91) their traditional songs have characteristic rhythmic patterns. The most common such pattern reported by Riester can be written in ratios as 1:1:2. Traditional Tsimane’ musical culture also once included Shamanic practices with drum playing, but these practices are no longer in use (92). The region where the Tsimane’ live is undergoing rapid modernization due to a push by the Bolivian government to provide services to the indigenous peoples. Radio usage is now fairly common, and villages near the local town of San Borja tend to have electricity. In addition, evangelism has spread within many villages through Christian missionaries. Thus, in addition to their knowledge of traditional music, nowadays most Tsimane’ villagers are somewhat familiar with religious Christian hymns. These hymns are monophonic and sung in Tsimane’. They are similar to traditional Tsimane’ music in relying on small intervals and a narrow vocal range and are sometimes accompanied by other instruments played by community members. Group singing appears to be rare, irrespective of whether the material is traditional songs or hymns. We recruited participants in three Tsimane’ villages. Two of the villages (Mara and Moseruna) were a 2-day walk or a 3-hour car ride from San Borja, along a road that was only accessible to high-clearance vehicles and motorcycles if recent weather had been dry. The other village (Yaranda) was located along the Maniqui river and only accessible by a one-day trip on a motorized canoe. All three villages have relatively little communal church singing. Participants had varied musical experience, but none regarded music as a profession. Two participants reported regularly playing an instrument. 34 participants reported playing an instrument at least once.

San Borja – Non-musicians (BO.SB-NM): San Borja is a small town in the Bolivian department of Beni in the Amazon basin. San Borja can be reached by car during dry months of the year but is accessible only by small planes during much of the rainy season. Participants were recruited by word of mouth and had resided for most of their life in San Borja.

Santa Cruz de la Sierra – Non-musicians (BO.SC-NM): Santa Cruz de la Sierra is the largest city in Bolivia. Participants were recruited using an online advertisement and word of mouth. All participants had been born and raised in the Bolivian department of Santa Cruz, and resided in the city at the time of the experiment.

La Paz – Students and Non-musicians (BO.LP-ST, BO.TS-NM): La Paz is the third-largest city in Bolivia, located in the Andes far above sea level. We recruited the student group from local universities. Most student participants were recruited by a student that was a research assistant. For the non-musician group, we recruited blue-collar workers employed by a hotel and their relatives. These participants were mostly from El Alto, a city adjacent to La Paz, and many of them were indigenous (Aymara or Quechua). Many participants reported experience with traditional music and dance in childhood or as adults.

Brazil – Local Musicians (BR-LM)

We recruited percussionists in the Recife metropolitan area, a city in Pernambuco, part of Northeast Brazil. The percussionists practice a local style of music that is part of the *Maracatu-nação* (translation from Portuguese: “nation *maracatu*”) cultural tradition. This regionally-specific tradition, usually perceived as Afro-Brazilian, involves religion, music, song, dance, and elaborate costumes. Its history is disputed but is most often linked to the colonial coronations of queens and kings among enslaved Africans in Brazil (in the 16th-19th centuries). However, the earliest descriptions date only from the beginning of the 20th century (93, 94). The music is performed by community-based groups in the slums of the metropolitan area of Recife termed *maracatus-nação*, approximately 30 of which are currently active. These groups consist of a costumed dance group, a large percussion ensemble, and people working in the production. The groups perform in parades that take place primarily during carnival. The percussionists we recruited either grew up within one of these community groups or had participated in one for at least several years.

Uruguay – Students and Local Musicians (UY-ST, UY-LM)

Participants in Uruguay were all recruited in Montevideo, the capital of Uruguay. The musicians were performers of Uruguayan candombe drumming (52, 95). All participants were born and raised in neighborhoods with a strong tradition of candombe drumming, and were “native players”, having acquired their competence in the style by direct transmission. The majority of the participants were outstanding players, regarded as master drummers by the community. The students were members of the local university with no formal musical training, but with passive exposure to music, including Uruguayan candombe drumming.

United Kingdom – Students and Jazz Musicians (UK-ST, UK-LM)

Two groups of participants were recruited in the Northeast of the UK. The first group consisted of students from Durham University. The second group consisted of instrumental jazz musicians, comprising a mix of professional musicians and students currently studying music at university (students were recruited from Durham University and the University of Edinburgh). All participants in this group reported that they perform jazz in public and earn money from performing. These participants played a range of instruments (including piano, saxophone, guitar, trumpet, drums, and double bass), and most reported performing in a range of different sub-genres and groups, most commonly big bands and smaller ensembles (e.g., trios) playing jazz standards. Some participants reported liking/performing musical traditions from other cultures, including Latin and Balkan-influenced music.

Sweden – Local Musicians (SE-LM)

The recruitment of musicians in Stockholm, Sweden, focused on students and teachers of folk music performance at the Royal College of Music, and on dance students at the School of Dance and Circus at the Stockholm University of the Arts, as the latter also had extensive experience playing music. Among the latter group, we required the participants' focus to be Swedish folk dance. Out of the recruited 22 participants from the Stockholm area, nine considered themselves mainly as dancers and thirteen mainly as musicians, and one participant identified as a dancer and musician to the same degree. Independent of this self-categorization, all participants asserted that they dance and have experience either with instrumental or vocal music making.

Bulgaria – Local Musicians and Dancers (BG-LM, BG-DA)

The recruitment of musicians and dancers in Bulgaria focused on members of a type of professional folk ensemble that developed once the country adopted a communist system of government after the Second World War (96, 97). These ensembles typically consist of an orchestra of folk instruments, a women's choir, and a dance troupe, and give stage performances of arranged and newly composed Bulgarian folk music and elaborately choreographed folk dances. Most performers in these ensembles have studied folk music performance or choreography in the Bulgarian conservatory system. Some of

the dancers who participated in the study belonged to professional or semi-professional dance troupes that perform Bulgarian folk dances with recorded rather than live music. We recruited participants in three Bulgarian cities: Pleven, Plovdiv, and Sofia.

Turkey – Students and Local Musicians (TR-ST, TR-LM)

The group of student participants was recruited from Istanbul Technical University Turkish Music State Conservatory and Bogazici University. The musician participant group was recruited from the cities of Izmir and Istanbul and included professional musicians who studied at institutes or conservatories for traditional Turkish music, such as Istanbul Technical University Turkish Music State Conservatory. These musicians were experienced in Turkish folk music or dance through formal training or extensive practice in groups or ensembles. They had experience in various sub-genres of Turkish music, ranging from Aegean to Black Sea regions, also including Sufi music. Musicians were involved in ensembles performing traditional, religious, classical or modern Turkish music, as well as Western music with eastern influences.

Mali – Students, Local Musicians, and Dancers (MA-ST, MA-LM, MA-DA)

The group of university students was recruited in the capital city of Bamako, and comprised both BA and MA level students as well as recent graduates of the University of Bamako. The musician group was also recruited in Bamako. Their main performance work occurred in local music and dance events, primarily wedding celebrations, but they also worked in the national and international scenes of staged folk dance and percussion music (98). By contrast, Malian dancers were not active as specialized musicians, but regularly danced at wedding celebrations in which professional musicians performed, and thus were highly familiar with local styles of music. The dancer participants were recruited based on their membership in a local dance organization in the peasant village of Sagele, approximately 75 km to the southwest of Bamako.

Botswana

San – Local Musicians (BW.EA-LM): This group was recruited in D'Kar, a village ~640 km NW of Gabarone (the capital of Botswana). Participants were either members of a local organization that performed traditional songs and dances of the San culture, primarily for nearby events or for tourists, or were local residents with significant musical experience (at least 10 years of self-reported musical experience).

Etsha – Local Musicians (BW.EA-LM): This group was recruited in Etsha, a village ~320 km north of D'kar. Participants were members of two groups who performed traditional songs, or local residents with significant musical experience (more than 7 years of self-reported years of musical experience). Participants came from two closely related Okavango Delta region subcultures: Hambukushu and Bayei. Although the Hambukushu and Bayei are culturally distinct, their geographically proximity means they are each exposed to each other's music, with far less exposure to the music of the San.

Namibia – Non-musicians and Local Musicians (NM-NM, NM-LM)

Participants were recruited from a small local Damara community (population approx. 500) in the Spitzkoppe region via word of mouth during the week prior to the start of the study. Recruitment and data collection occurred with help from two local research assistants, who also acted as translators. The musician group was drawn from two active musical groups, both of whom had extensive experience with traditional music of the region: a 'cultural singing group' (who frequently practice and perform 'lais' (old traditional folk songs and dances), in addition to other styles of performance), and a 'youth choir' (who typically rehearse and perform 'elob mis', a form of gospel music, sung in Khoekhoegowab (the local language), English and Afrikaans). Both groups perform at local and regional community events (such as weddings, funerals, and at an annual Damara traditional festival). The singing group intermittently performs for visiting tourists, and the youth choir also performs weekly

in church. The first group typically earned money in exchange for these performances, though most members would not be construed as professional performers. The non-musician participants were members of the local community who were not part of either group. Although almost all participants reported engaging in some form of singing (e.g. joining in at weddings or other community events, or simply while listening to the radio), these non-musician participants did not regularly practice or perform with any group, and many had limited or no knowledge of old traditional song lyrics and dances.

India – Non-musicians and Local Musicians (IN-NM, IN-LM)

The experiment was conducted at I.I.T. Bombay in the city of Mumbai. The non-musicians were graduate students in Engineering at I.I.T. Bombay who had not had any musical training or significant exposure to Indian classical music. Most of the musician participants were professional musicians living in the city, but a few were students at I.I.T. Bombay. Most of them were trained in the North Indian (Hindustani) form of art music. Formal training typically involves taking one-on-one lessons from a teacher (the appointed guru) and performing in public concerts. Some of the musician participants were primarily vocalists but also played an instrument, and about half of them were tabla (percussion) players who accompanied vocalists in concerts.

S. Korea – Students, Western Classical Musicians, and Local Musicians (KR-ST, KR-WM, KR-LM)

The student participant group was recruited from Chungnam National University. The Western musician group consisted of students from the Department of Music at Chungnam National University. These participants had been exposed to both Western music and K-pop music via mass media. The local musician group comprised students studying traditional Korean instruments in the Department of Traditional Korean Music at Jeonbuk National University. These participants had trained on Korean traditional instruments for many years, but also had extensive exposure to both Western and K-pop music.

Japan – Students, Western Classical Musicians, and Local Musicians (JP-ST, JP-WM, JP-LM)

The student and Western musician groups were recruited from Keio University Shonan Fujisawa Campus in Fujisawa, Kanagawa near Tokyo. Participants in the student group had no formal musical training. The Western musician group had formal music training in Western instruments. The local musician group was recruited from the Tokyo University of the Arts (Japan's leading music conservatory). We recruited students studying traditional Japanese instruments (*shamisen*, *koto*, *shakuhachi*, or *hayashi* instruments from the *noh* ensemble) in the Department of Traditional Japanese Music. These students had to train on these instruments for many years to qualify for acceptance into the department, and can be considered "bi-musical" (99) in that they had extensive exposure to both popular Western and traditional Japanese musical systems. All Japanese participants had extensive passive exposure to both Western music and Western-influenced Japanese popular music via radio, TV, and other media.

China – Non-musicians (CN-NM)

Participants in China were recruited from a cluster of Dong minority villages in the Guizhou Province in southwestern China. Singing features prominently in many village activities, but the most famous and distinctive tradition of Dong song is *Dage*, or Big Song, recognized by UNESCO in 2009 on their list of humanity's Intangible Cultural Heritage. There is wide participation throughout the villages in learning and performing *Dage*, a genre of two-part group polyphonic singing, occasionally metered, where words are central and pitch height and contour carefully follow the lyrics (100). *Dage* is performed informally within people's homes and more formally within the drum-tower.

Online participants from US, Brazil and India (US-OL, BR-OL, IN-OL)

All participants were recruited online using Amazon Mechanical Turk. In the online advertisement, we required participants meet the following five criteria: Participants had to (i) be at least 18 years old, (ii) speak English, (iii) use a laptop to complete the experiment (no desktop computers allowed), (iv) use an up-to-date Google Chrome browser (due to compatibility with the technology), and (iv) sit in a quiet environment, to ensure that their tapping could be recorded precisely. To test participants from the US, India, and Brazil, we used Amazon Mechanical Turk's qualification system, which allows researchers to recruit participants registered in each location. To ensure participants were undertaking the experiment from the registered location, we only included participants whose registered location matched their IP-based geo-location, which we verified using the service "ipinfo.io."

Analysis

Small-integer ratios

Following Jacoby and McDermott (2017), we considered "small-integer-ratio rhythms" to be only those with ratios composed of the integers 1, 2, and 3. This results in the following 22 unique ratios: $\Omega_{22} = \{1:1:1, 1:1:2, 1:2:1, 2:1:1, 1:2:2, 2:1:2, 2:2:1, 1:1:3, 1:3:1, 3:1:1, 1:2:3, 2:3:1, 3:1:2, 1:3:2, 2:1:3, 3:2:1, 2:2:3, 2:3:2, 3:2:2, 2:3:3, 3:2:3 and 3:3:2\}$. We also grouped together categories that are equivalent under cyclic permutation, resulting in eight categories: $\underline{\Omega}_8 = \{111, 112, 122, 113, 123, 223, 233, 132\}$.

Cyclic permutations

One feature of the experimental results is that the response distributions are typically symmetric across cyclic permutations (Figure 2). For example, the modes at 1:1:2, 1:2:1, and 2:1:1 have about the same weight for a given participant group. To quantify this symmetry, we counted the percentage of seeds in the final iterations that are in each of the three possibly cyclic permutations, which can be identified by whether the longest interval is either the first, second, or third position (Figure S5A).

As is evident in Figure S5B, the deviations from perfect symmetry were relatively modest (perfect symmetry would yield 33.3% of tapped responses in each third of the space; actual proportions ranged from 24.3% to 43.6%; the standard deviation of the difference from 33.3% was 3.2%). However, these deviations appear to be non-random, with a tendency for more weight on permutations in which the third interval is the longest (red region). Previous literature (101-103) in fact predicts that the most frequently occurring permutations would be those where the long interval occurs at the end, because if this configuration is played cyclically, the long interval provides a gap that helps the pattern to group according to Gestalt principles (101). This was the case in 29 of 39 groups (those for which the yellow area in Figure S5B extends beyond the horizontal line; $p < 0.001$ via a one-sided sign test).

What can explain the overall tendency towards symmetry? In principle the symmetry could reflect the fact that the beginning of a repeating cycle is ambiguous if participants ignore or forget the initial interval. In an earlier paper (25), we tested whether this ambiguity underlies the symmetry evident in the response distributions. Specifically, we performed an experiment in which the first click of the repeating stimulus was given a 7dB level increment to render the permutations distinct. The results (Figure S5 in that paper) indicated that the symmetry of the response distribution was largely maintained, suggesting that perceptual ambiguity is not the only reason for symmetry in the prior.

Another possibility is that participants do not always hear the first element of the repeating three-element pattern as the "beginning" of the cycle (i.e., a metrically stable "downbeat"), instead hearing the first element as an "offbeat" (a syncopated element that precedes or follows the metric downbeat). This interpretation is supported by the cross-cultural presence of "offbeats" in music (104-107). For instance,

offbeats enable a singer to fit a varied number of syllables within a given (strophic) repeated rhythmic pattern; also, offbeats allow musicians to create specific rhythmic interest or tension with otherwise simple rhythmic patterns (104, 106, 108). Under this interpretation, cyclic permutations of an interval pattern would have different beginnings but would otherwise be perceptually similar, such that the distinct peaks in the prior for different permutations might reflect the same underlying cyclic rhythm representation but with distinct offbeat patterns (78, 109, 110).

Kernel density estimates of the prior

The experiment consisted of a number of trials, each of which consisted of iterated reproduction of a random seed rhythm. We estimated a participant's prior using the data from the fifth and final iteration from each trial, having demonstrated in Jacoby and McDermott (2017) that five iterations are sufficient for the iterative procedure to converge to the prior (Figures S1-S2 and S7 in Jacoby and McDermott (2017); see Figure S7 of this paper for analyses of convergence in each group) (25). The prior was estimated by adding together Gaussian kernels, with mean μ_i and covariance Σ_i empirically computed from the repetitions of the rhythm (there were up to 10 repetitions depending on the number that the participant correctly produced during the fifth iteration; for repetitions that had one missing tap, the missing tap was replaced by the mean onset of the successfully produced taps at that stimulus position). Since this covariance matrix is estimated based on small numbers of samples, we added a regularization term: $\Sigma'_i = \Sigma_i + \gamma I$, where I is the identity matrix, and $\gamma = 15$ ms (we increased slightly this value compared with the $\gamma = 10$ ms of Jacoby and McDermott (2017) since some participant groups had lower numbers of correctly reproduced repetitions). We averaged these kernels ($G_i(x) \sim N(\mu_i, \Sigma'_i)$, one per trial) across all trials completed by a participant, obtaining a distribution $P(x) = \frac{1}{N} \sum_{i=1}^N G_i(x)$ over the triangle. For statistical analyses, we represented these distributions in bins spanning 0.006 in each dimension of the triangle (i.e., 12ms with given the 2000ms pattern duration). To generate high-resolution images for the paper figures, we used bins of size 0.003 in each dimension.

Normalizing density compared with uniform distribution

As described above, the random seeds were constrained to have the smallest interval exceed 15% of the pattern duration (300 ms), corresponding to a smaller triangular region within the full rhythm triangle. We defined the uniform distribution over this smaller region as U . To avoid working with small numbers, we pointwise-normalized the kernel density estimate P with respect to U , namely $P'(x) = P(x)/U(x)$.

Jensen-Shannon divergence

To compare distances between distributions, we used the Jensen-Shannon divergence. The Jensen-Shannon divergence of two distributions P and Q is defined as

$$JSD(P, Q) = \frac{1}{2} D_{KL}(P, M) + \frac{1}{2} D_{KL}(Q, M) \text{ where } M = \frac{1}{2}(P + Q) \text{ and } D_{KL}(P, Q) = \int_x P(x) \frac{P(x)}{Q(x)} dx .$$

Gaussian mixture model fits

To measure the relative weight of each category in a group's prior we used a Gaussian mixture model in which the mean of each mixture component was constrained to be close to an small-integer-ratio rhythm. We imposed additional constraints on the standard Gaussian mixture model fitting procedure of the model to ensure that the mapping of mixture components to integer ratios was fixed across groups, to avoid artifacts associated with small sample size, and to avoid uninterpretable overlap between the modes.

We define a Gaussian mixture model with category centers $\{\mu_i\}_{i=1\dots K}$ covariance matrices $\{C_i\}_{i=1\dots K}$ and weights $\{w_i\}_{i=1\dots K}$ as follows:

$$Q(x) = \sum_i \frac{w_i}{2\pi\sqrt{|C_k|}} \exp \left(-\frac{1}{2}(x - \mu_i)^T C_k^{-1} (x - \mu_i) \right)$$

To fit the model, we used a modified expectation-maximization algorithm (111). We initialized the algorithm by assigning the mixture components to the small-integer-ratio rhythms within Ω_{22} . We then proceeded by alternating between the expectation and maximization steps. After each maximization step, we applied additional constraints, namely:

- (1) Mode identity: to guarantee that each mode was associated with the same category across groups we required that $\|\mu_i - \mu_i^0\|_2 < \frac{1}{2}d_{min}$ where μ_i^0 is the i th category in Ω_{22} , and where d_{min} is the minimal distance between categories in Ω_{22} .
- (2) Overlap: to avoid overlap between the modes we required that $|\lambda_i| < \frac{1}{2}d_{min}$.
- (3) Additional constraints on the overlap between the modes: we also required that the two eigenvalues of the covariance matrix λ_1 and λ_2 be limited by: $1/A < \frac{\lambda_i}{\lambda_1 + \lambda_2} < A$, where A is a constant. We fixed $A = 5$, which intuitively corresponds to a constraint on the aspect ratio of the ellipsoid defined by the covariance matrix.

We applied these constraints after the maximization step. We applied constraint 1 by projecting the μ_i resulting from maximization step to the closest point in Euclidean distance that satisfied constraint 1. Similarly, we applied constraints (2)-(3) on the eigenvalues of the covariance matrix by truncating them so they satisfied both constraints:

$$\begin{aligned}\lambda_i &= \left(\lambda_i, \frac{1}{2}d_{min} \right) \\ \lambda_i &= (\lambda_i, A(\lambda_1 + \lambda_2)), \frac{1}{A}(\lambda_1 + \lambda_2)\end{aligned}$$

We then iterated these steps until convergence (using a convergence threshold of $1e-6$). The result was an estimate of $\{\mu_i\}_{i=1\dots K}$; $\{C_i\}_{i=1\dots K}$ and $\{w_i\}_{i=1\dots K}$.

This procedure is different from the one reported in Jacoby and McDermott (2017), where we performed a numeric constraint optimization with the Matlab *fmincon* function on the KL divergence of the kernel density estimate defined by the model and the kernel density estimate of the data. Other than this difference, we had similar constraints on the optimization. The 2017 procedure provided comparable results but was significantly slower than the method used here. Considering the large amount of data in this project, we considered the efficient expectation-maximization method to be preferable to direct optimization.

Gaussian mixtures fitting with additional category at 7:2:3 (Figure 5G,H&I)

In the case of Malian drummers and dancers we added the additional ratios: 2:3:7, 7:2:3 and 3:7:2 (Figure 5). We denote by $\Omega_{25} = \Omega_{22} \cup \{2:3:7, 7:2:3, 3:7:2\}$ and used it instead of Ω_{22} for the Gaussian mixture model estimate. In all other respects the analysis was identical to the previous section (Gaussian mixture model fit with Ω_{22}).

Average category weights

In all cases, we fitted the Gaussian mixture to the categories (Ω_{22} ; except for Figure 5G,H&I where we used Ω_{25}). In some cases, we wanted to display or analyze the results ignoring cyclic permutations of the same category (for example, 2:2:3, 2:3:2 and 3:2:2 would be mapped to the same category 223). We then computed the Gaussian mixture model fit to Ω_{22} and averaged the weight across the three

permutations. The one exception was the isochronous rhythm 1:1:1 which has no variants; in this case we used the original fit of 1:1:1 in Ω_{22} . This resulted in eight weights per group corresponding to the eight categories in: $\underline{\Omega}_8$ (defined above in the “Small-integer ratio” section). These category weights are reported in Figure S8 and are provided as part of the OSF repository associated with this publication (see below).

Significant distance between two groups

In this analysis, we evaluated whether two distributions P_1 and P_2 associated with groups i_1 and i_2 had significantly different kernel density estimates. Since the Jensen-Shannon divergence is always positive, it deviates from zero when the kernel density estimates being compared are computed from a finite sample. We used bootstrap to estimate whether the distance between P_1 and P_2 was greater than what is expected from this finite-sampling effect. We created 1000 simulated split halves of the trials from each participant group. From these bootstrap samples we estimated the 1000 kernel density estimates associated with group i_1 (we denote them by $P_1^{j,1}$ and $P_1^{j,2}$ where j indexes the 1000 split halves) and group i_2 (denoted by $P_2^{j,1}$ and $P_2^{j,2}$). We then computed $JSD(P_1^{j,1}, P_2^{j,1})$ (the distance between split halves across groups) and compared it with $JSD(P_1^{j,1}, P_1^{j,2})$ and $JSD(P_2^{j,1}, P_2^{j,2})$. We obtained a p-value from the minimum of the rank order of ($JSD(P_1^{j,1}, P_2^{j,1})$) within $JSD(P_1^{j,1}, P_1^{j,1})$ and $JSD(P_2^{j,1}, P_2^{j,1})$. Namely, to declare that two groups are significantly different, their mean JSD distance had to be significant with respect to both within-group distances.

Discrete mode (“peakiness”) analysis

We performed three analyses to substantiate the presence of discrete modes in the measured priors.

Analysis 1: To show that the mass of the estimated density is centered in a small part of the space, we computed for each group the 33% of the bins with the highest kernel density, and then computed the sum of the density in these bins relative to the sum over all bins. This resulted in numbers ranging between 60.5-82.3% (mean 70.1%) over the 36 groups. To obtain a null distribution, for each group we sampled points (equal in number to the total number of trials for that participant group) randomly on the triangle and estimated the empirical kernel density estimate for this random distribution. We then repeated the selection process described above, picking the 33% of the bins with the highest kernel density and computing the proportion of the summed density in these bins. This analysis showed that the percentage obtained in this way from the empirical data was significantly larger than would be expected from the null distribution computed from uniform sampling.

Analysis 2: We also estimated the peak density with respect to a uniform distribution. We identified the bin in the kernel density estimate with the highest density and found that in all 36 groups this bin had a density that was over five times larger than the density of the same bin under a uniform distribution (range 5.1-13.1; mean 8.8). We estimated the statistical significance of this ratio using a null distribution obtained by sampling from a uniform distribution (using the same number of points per group) and measuring the peak density from the resulting kernel density estimate. We found that the empirical peak ratios were significantly larger than the null distributions of these peak ratios for all 36 groups ($p<0.001$).

Analysis 3: We fit a Gaussian mixture model with mixture components constrained to be near small-integer ratios (see section “Gaussian mixture model fits” for details). This model explained most of the variance in the kernel density (92.2% on average, ranging from 83.0-97.7% depending on the group). Explained variance was measured here by treating the kernel density estimates of both the empirical data and the models as vectors, and squaring the correlation between the vectors.

Overlap with small-integer ratios (Figure 2C)

We computed the average minimal distance between each fifth iteration reproduction (for all participants in a given group) and the closest small-integer-ratio rhythm:

$L_2(\Omega_{22}, p_i)$ where p_i are the fifth iteration reproductions represented on the rhythm triangle and Ω_{22} is the set of all small-integer ratios involving the numbers 1-3 (defined above), and L_2 is the minimum of the Euclidean distances between the point p_i and each of the 22 points in Ω_{22} . To show that these distances are significantly smaller than would be obtained by chance, we generated a null distribution by randomly sampling sets of 22 points uniformly from the triangle and computing the same mean minimal distance between the points p_i and these randomized sets. We then compared the empirical distance D_{22} to the null distribution from randomized sets. These results are reported in Figure 2C.

We performed an additional control analysis where instead of sampling the sets of 22 points uniformly we constrained them so that each point fell within a circle of radius d around the integer points, where d is $\frac{1}{2}$ of the minimal distance between two points in Ω_{22} . This guaranteed that the null sets were spaced similarly to Ω_{22} . The results of this alternative analysis were similar to the simpler analysis described above, and in all 36 cases the empirical distance was significantly smaller ($p<0.001$) than would be expected from the null distribution, even when Bonferroni correction was applied.

Finally, we also applied the integerness score reported in Jacoby and McDermott (2017). In this analysis we compared the Jensen-Shannon divergence distance between the empirical KDE of the fifth iteration (P) and the normalized indicating function $I_{\Omega_{22}}(x) = \frac{1}{22} \sum_{\omega \in \Omega_{22}} \delta(x - \omega)$, where δ is the Dirac delta function on the triangle. This Jensen-Shannon divergence is maximal if all probability mass is located at the small-integer ratios. We initially fitted an unconstrained Gaussian mixture model with 22 components. We then randomized the means of the components of this mixture. This simulates a response distribution that is similar in statistical characteristics to the data, but that is not centered around integer ratios. We obtained a null distribution by generating 1000 such randomized distributions, each time computing the Jensen-Shannon divergence between $I_{\Omega_{22}}(x)$. We then compared the distance of $I_{\Omega_{22}}(x)$ and P to this null distribution. We found that in all cases the distance between $I_{\Omega_{22}}(x)$ and P was significantly smaller ($p<.01$) than what would be expected by chance. In Figure 2C, we used the first analysis described in this section, as it is simpler to describe.

Multidimensional Scaling (MDS) analysis (Figure 3A)

To visualize the similarity relations between the rhythm priors for each participant group we first estimated the priors as the kernel density estimate P_i from the fifth iteration of the experiment (aggregated for all participants in each group; see above). We then computed the Jensen-Shannon divergence between all pairs of groups $M_{ij} = (P_i, P_j)$. We used Matlab's `mdscale` function with default parameters to obtain a two-dimensional space in which the rhythm prior for each group was positioned so as to best match the measured distances. Note that we used the distances between the full distributions (i.e., the kernel density estimates) for the multidimensional scaling analysis (as opposed to the Gaussian mixture models used in other analyses, or any other parametric model).

Category weight for the 3:3:2 rhythm (Figure 3B)

To compute the weight of the 3:3:2 rhythm for each group, we computed the Gaussian mixture model weights as explained above for the 22 rhythm categories, and then averaged the weights over the three cyclic permutations of 3:3:2 (3:3:2, 3:2:3 and 2:3:3). We obtained error bars via bootstrap, sampling 1000 datasets with replacement for each group and computing the weights of the Gaussian mixture model for each of these datasets. Error bars plot one standard deviation of the resulting distribution above and below the mean (i.e., the standard error of the mean). The order of the groups in the bar graph of Figure 3B is drawn from the first dimension of the multidimensional scaling analysis, and it is

obvious that this dimension is correlated with the 3:3:2 category weight (which increases nearly monotonically across the first multidimensional scaling dimension).

Multidimensional scaling analysis and category weights (Figure 3C)

We computed the correlation between the multidimensional scaling dimensions and each of the category weights for each group, averaged across cyclic permutations. The multidimensional scaling projections and Gaussian mixture model weights were computed as described above.

Analysis of student groups (Figure 4)

In the first analysis, we computed the average distance (Jensen-Shannon divergence) between the estimated priors of all pairs of student or online groups in different countries and compared it to that of pairs of non-musician and local musician groups (from the same countries as the student/online groups). The pairs we considered were:

Bolivia(La Paz)-ST & Bolivia(La Paz)-NM, Bolivia(La Paz)-ST & Bolivia(San Borja)-NM, Bolivia(La Paz)-ST & Bolivia(Santa Cruz)-NM, Bolivia(La Paz)-ST & Bolivia(Tsimane)-NM, Brazil-OL & Brazil-LM, India-OL & India-NM, India-OL & India-LM, Japan-ST & Japan-LM, S.Korea-ST & S.Korea-LM, Mali-ST & Mali-LM, Turkey-ST & Turkey-LM, UK-ST & UK-LM, US(Boston)-ST & US(NYC)-NM, US(Boston)-ST & US(NYC)-LM, US-OL & US(NYC)-NM, US-OL & US(NYC)-LM, Uruguay-ST & Uruguay-LM.

To evaluate the statistical significance of the difference in distances we created shuffled datasets where two sets of groups (one the same size as the student/online set and one the same size as the non-student/non-online set) were sampled without replacement from the union of the student/online and non-student/non-online sets. We then computed the difference between the average Jensen-Shannon divergences of these shuffled groups for each resampling, and evaluated the probability of the actual difference under this null distribution.

In the second analysis, we computed the average distance (Jensen-Shannon divergence) between the US student group (US(Boston)-ST) and the priors of all other student/online groups (Bolivia(La Paz)-ST, Uruguay-ST, UK-ST, Turkey-ST, Mali-ST, S.Korea-ST, Japan-ST, US-OL, Brazil-OL, India-OL). We compared this average distance to a null distribution obtained by sampling sets of non-student/non-online groups of the same size (10 groups) and measuring the average distance of each set of groups and the US student group.

To control for the fact that student groups tended to be younger than other groups, we repeated the above two analyses restricted to participants younger than 40. The group differences in mean age were not eliminated by this restriction, but they were significantly reduced (with this restriction, all groups had mean ages between 21-33.7 years, and could all be considered “young”). The pairwise distance between student groups was still significantly smaller than that between non-student groups (students: $p=.005$; the analogous control analysis for online groups also remained significant: $p<.001$), as was the distance between US student groups and other student groups when compared to that between US student groups and non-student groups (students: $p<.001$; the analogous control analysis for online groups again remained significant: $p<.001$).

We also performed a control analysis testing for effects of age by comparing the differences in the kernel densities between young and old subsets of the online groups. Given that over a third—specifically, 46%, 34%, and 39% in the US, Indian and Brazilian groups, respectively—of participants in the online groups were older than 35, we divided all groups into young and older subsets using the age of 35 as a threshold. We found that kernel density estimates for the US, Indian and Brazilian groups were not significantly different ($p=.09$, $.09$ and $.39$, respectively). These results suggest that age does not explain the increased similarity of student populations.

Word clouds of favorite music (Figure 4C,D,F&G)

In the demography questionnaire, we asked participants to list their three favorite bands or musical artists, and to indicate the genre for each. The level of detail varied somewhat between individuals and groups (e.g., some individuals specified sub-genres such as “indie-rock” whereas others indicated “rock”). Due to site-specific limitations on the experiment session duration, this question was only asked of 31 of the 39 groups. We analyzed the results using single word histograms. Some of these histograms are presented as word clouds in Figure 4, with the font size proportional to the frequency of occurrence. In addition, the four most common words for each group are presented in Table S2.

Analysis of specific modes (Figure 5)

Figure 5 analyzes the weight of particular rhythm modes in different participant groups. To compute these weights for each group, we computed the Gaussian mixture model weights as explained above for the 22 rhythm categories, and then averaged the weights over the three cyclic permutations of the rhythm in question. We then examined these weights for participant groups whose local musical tradition was known to feature the rhythm. We asked whether the weights were higher than in the remaining participant groups using a Wilcoxon rank-sum test.

The 2:2:3 and 3:3:2 rhythms had been previously associated with specific musical traditions. 2:2:3 has been documented in Balkan, Turkish, and Botswanan music (42-44), which often employ meters with a signature of 7/8. Balkan and Turkish listeners have also been shown to better discriminate this pattern than US and Canadian participants without such familiarity (46, 112). The 3:3:2 rhythm is similarly ubiquitous across sub-Saharan Africa (47, 48) and the African diasporas (49-52) in the Americas. We confirmed its presence in the musical culture of our Malian dancer participants (Mali-DA; recruited among farmers from Sagele village in Southern Mali), by recording and analyzing a representative corpus of their musical repertoire. The pieces chosen were ones to which they frequently danced in the context of wedding celebrations and other local events. We found that forty-six percent of the recorded excerpts prominently featured a 3:3:2 pattern, making it one of the most characteristic rhythmic patterns in this repertoire.

The 7:2:3 rhythm evident in the priors measured from drummers in Mali (Figure 5G,H&I) is popular in West Africa; a slightly denser, five-interval variant (2:2:3:2:3) constitutes a signature rhythm that is emblematic of the musical culture area(47, 48). Additionally, we substantiated its presence in Malian music as described in the Results section. We used the same procedure to validate participant responses in Bulgaria (2:2:3 rhythm).

Validation of musicianship (Figure 6)

We observed that self-reported years of musical experience consistently distinguished the musician and non-musician groups as intended (Figure 6 A-B). We also observed that objective accuracy in our task (tapping standard deviation) was higher for musicians (both those playing Western music and those playing local musical styles) compared with students and non-musicians ($p < 0.001$ see Figure 6C). In countries where we ran both musicians and non-musicians, musicians were generally more accurate than non-musicians (only one exception; Figure 6D). These results suggest that the concept of musicianship translates cross-culturally (32, 38).

Fast tempo experiment (Figures S1 and S2)

Procedure

When experimental conditions allowed for longer sessions, we ran an additional experiment to explore whether the results would be similar at other tempos. The experiment was always run last, and was identical to the main experiment except that the pattern duration was 1000 ms. The other experimental constants (for example, the fastest allowed interval) were scaled accordingly (the experiment was

identical to the fast tempo experiment in Jacoby and McDermott (2017), experiment S2, shown in Figure S3 of that paper).

Participants

293 participants from 13 groups (6 countries) participated in the fast tempo experiment. These participants performed 7587 trials (seeds) with 911,564 taps. The demographic information of these participants is summarized in Table 2.

Table 2: Demographic information for the fast tempo experiment.

#	Group Name	Group abbreviation	Type	Continent	Country	City/Cities	N	Females	Ages
1	US(NYC)-NM	US.NY	NM	North America	USA	NYC	26	15	18-61 (mean = 38.77 std= 13.00)
2	US(NYC)-WM	US.NY	WM	North America	USA	NYC	26	14	18-64 (mean = 30.54 std= 11.29)
3	US(NYC)-LM	US.NY	LM	North America	USA	NYC	34	4	18-64 (mean = 36.94 std= 13.62)
4	Bulgaria-LM	BG	LM	Europe	Bulgaria	Pleven/Plovdiv/Sofia	23	3	20-59 (mean = 41.48 std= 10.51)
5	Bulgaria-DA	BG	DA	Europe	Bulgaria	Pleven/Plovdiv/Sofia	32	15	19-53 (mean = 33.31 std= 8.76)
6	Turkey-ST	TK	ST	Europe	Turkey	Istanbul	27	18	18-28 (mean = 20.67 std= 2.09)
7	Turkey-LM	TK	LM	Europe	Turkey	Istanbul/ Izmir	24	7	22-44 (mean = 31.75 std= 6.19)
8	Mali-LM	MA	NM	Africa	Mali	Bamako	20	1	25-58 (mean = 38.40 std= 9.95)
9	Mali-DA	MA	NM	Africa	Mali	Sagele	24	21	20-62 (mean = 37.67 std= 9.91)
10	Namibia-NM	NA	NM	Africa	Namibia	Spitzkoppe	16	9	18-46 (mean = 30.88 std= 8.45)
11	Namibia-LM	NA	NM	Africa	Namibia	Spitzkoppe	14	10	18-48 (mean = 27.79 std= 8.68)
12	India-NM	IN	ST	Asia	India	Mumbai	15	14	20-44 (mean = 27.13 std= 7.30)
13	India-LM	IN	LM	Asia	India	Mumbai	12	1	17-43 (mean = 25.67 std= 7.09)

Analysis of results

The kernel density estimates of the 13 groups are provided in Figure S1. We provide the raw data of the experiment in the OSF repository associated with this publication (see below).

Overall, results at the faster tempo were similar to those at the slower tempo. All 13 groups who performed the fast tempo experiment produced priors that were closer to integer ratios than could be obtained by chance ($p < .001$ in all cases), even with Bonferroni correction.

Figure S2 shows an analysis of the weights of the modes in the 13 groups. The weights of the 22 categories were correlated across the two tempi ($r=0.35-0.72$; mean=0.57; std= 0.11 across the 13 groups). As expected from previous literature, there were also some subtle differences between the category weights for the two tempos (Figure S2). Three of the four largest effect sizes were found in dancers (Bulgarian dancers: effect size of 5.5 – more weight on category 2:2:3 in the fast tempo; Malian dancers: effect size of 5.8 – more weight on 1:2:3 in the fast tempo; Malian dancers: effect size of 4.6 – decreased weight on 3:3:2 in the fast tempo). These tempo-dependent effects in dancers are consistent with the idea that dancers have an increased sensitivity to tempo and to embodied aspects of music (113, 114). For instance, Bulgarian dancers showed much more weight on 2:2:3 at the faster tempo. Bulgarian folklorists have long recognized tempo as an important factor in metrical patterns that feature a 2:2:3 ratio, such that the metric durations are only considered fundamentally unequal when performed at fast tempos (115, 116). This idea was also reflected in one of the interviews we conducted with the musicians after the experiment. When we asked one participant whether she recognized the 2:2:3 pattern with a period of 2000 ms, she identified it as the rhythm of a Bulgarian dance type called *ruchenitsa*, but slower than usual. The effect of tempo in Malian dancers' 1:2:3 category is similarly consistent with Polak et al. (2018) (26) who showed that reproductions of short-long patterns in two interval rhythms (the first part of the 1:2:3 pattern) strongly vary with tempo. This pattern is characteristic of the three most common Malian jembe musical pieces: Maraka, Suku and Manjanin, which are typically performed at a very fast tempo (100-200 beats per minute) (117).

Bias analysis (Figure S3).

Figure S3 displays an analysis testing whether perceptual category centers are systematically biased away from the corresponding small-integer ratio. The small black dots plot the modes of the fitted

Gaussian mixture models for each category and participant group. We then calculated the empirical means (indicated in larger circles) of each category across all groups. We applied a non-parametric test analogous to an ANOVA to test whether each category was biased. The test statistic was the ratio of 1) the average squared Mahalanobis distance between all points and the empirical mean and 2) the average squared Mahalanobis distance between all points and the corresponding small-integer ratio. We compared this test statistic to its null distribution computed from 10,000 bootstrapped samples where the data were randomly sampled from a Gaussian distribution with the same empirical covariance matrix as the experimental data but with the mean set to the integer ratio category (i.e., with zero bias).

9 categories showed small but significant deviation from unbiased integer ratio categories after Bonferroni correction ($p < 0.002$ for all cases; see red circles in Figure S3). The biased categories consisted of the three cyclic permutations of each of 1:2:3, 2:1:3, and 2:2:3 (123/231/312, 213/321/132, and 223/232/322). The bias of the “6/8” categories 1:2:3 and 2:1:3 is consistent with lengthening of short elements in rhythm performance studied in European musicians (39). It may also be related to the phenomenon of “swing” found in jazz music, in which the short interval in ratios of 1:2 often is slightly elongated (118).

Principal component analysis (Figure S4).

In Figure 3, we used multidimensional scaling to perform dimension reduction, computing the Jensen-Shannon divergences between the kernel density estimates of all pairs of groups. As an alternative, we performed an analogous analysis using principal components analysis (Figure S4). We treated the triangle images generated from the kernel density estimates as large feature vectors, where each pixel is a feature (using kernels with a resolution of 12 ms; i.e. $0.006 \times$ the pattern duration of 2000ms). We computed the principal components of these vectors across 39 groups. The projections of the 39 groups' priors onto the first two principal components show a structure very similar to what we obtained with multidimensional scaling (Figure S4A). For example, it is apparent that student groups are again centered in the middle. As with multidimensional scaling, the projection of each group's estimated prior onto the first principal component is positively correlated with 2:3:3 category ($r = .94$ $p < .0001$), and negatively correlated with the simpler categories (1:1:1, 1:1:2, 1:2:2; $r = -.44, -.59, -.7$, respectively; $p = .043, .0007, <0.0001$; Figure S4B). The projection onto the second principal component is correlated with 6/8 rhythms (1:3:2: $r = -0.75$, 1:2:3: $r = -0.67$; $p < 0.001$ in both cases) as well as the 1:1:2 rhythm ($r = 0.63$, $p < 0.001$). The components can also be visualized (Figure S4C), revealing that their minima and maxima overlap with small-integer-ratio categories. The consistency between the different dimensionality reduction methods indicates the robustness of the results that we report in this paper. The raw data for the category fitting and category weight for each group is also provided in the OSF repository associated with this publication (see below).

Symmetry analysis (Figure S5)

To quantify the symmetry of the response distributions, we counted the percentage of tapped rhythms from the final iteration that occurred in each third of the triangular rhythm space, corresponding to whether the longest interval is in the first, second, or third position (Figure S5A).

Influence of language (Figure S6)

In this analysis, we evaluated if two groups that spoke the same or different languages had significantly different kernel density estimates. We used the procedure described in “Significant distance between two groups” (bootstrapped JSD distance).

Transmission error (Figure S7)

Transmission error is the magnitude of the difference between the stimulus and response seeds in each iteration. It is used in the serial reproduction literature to monitor convergence dynamics (119,

120). As an error measure we computed the average across trials of $e = \sqrt{(s_1 - r_1)^2 + (s_2 - r_2)^2 + (s_3 - r_3)^2}$, where (s_1, s_2, s_3) and (r_1, r_2, r_3) are the stimulus and average response of each iteration, respectively (i.e., the response is averaged across the 10 repetitions within each iteration). In our previous work we showed that convergence occurred after about five iterations for both Tsimane' and Westerners. Here we show similar dynamics for all groups.

Category weights (Figure S8).

Categorization weights were computed for each group for the seven category types. As part of the OSF repository, we have provided raw data for these category estimates.

Data availability

An OSF repository with all raw data for all groups, the experimental paradigm code, instruction materials to run the paradigm, and analysis scripts will be made available with the final published version of this manuscript.

Demography questionnaires

We employed a demographic questionnaire to characterize musical experience, experience with dance, and basic demographic information (age, gender, education, spoken languages). We used a baseline demographic questionnaire (see OSF repository) that was translated and adapted to different languages and recruiting populations by the researchers. There was some customization of the questions based on their relevance to the local culture. In each location we consulted with ethnographers and translators regarding the relevance and translations of each demographic item.

Image credits

Figure 1: First row: Eduardo A. Undurraga, Martín Rocamora, Bronwyn Tarr, Rainer Polak, Second row: Nori Jacoby, Rainer Polak, Daniel Cameron. Third row: Andre Holzapfel, Elizabeth Margulis, Preeti Rao.

Figure 4: Shinya Fujii.

Figure 5: Ivan Banchev, Jessica Grahn, Rainer Polak, and Luis Jure.

Ethics

Informed consent was obtained using the following approved protocols: Columbia University IRB-AAAR3726, Ethics Council of the Max Planck Society (2017_12, 2020_05, 2020_11), , University of Western Ontario Health Science Research Ethics Board 108477, Korea Advanced Institute of Science and Technology (KAIST IRB -KH2017-15), Chinese University of Hong Kong SBRE-19-695, BCPHS-12301 (Brandeis), Durham University Music Department Ethics Committee (February 2018), Committee on the Use of Humans as Experimental Subjects at MIT (COUHES) Protocol Number 1209005242R004. Boğaziçi University Ethical Board for Human Research SBB-EAK 2017/1.

Funding

This work was supported by the following funding sources: grants from the James S. McDonnell Foundation and National Science and Engineering Research Council to J.Grahn; grant number NRF-2017R1C1B2010004 to K. Lee; ANID, Millennium Science Initiative Program, grant NCS17_06 awarded to E. Undurraga; a DOC Fellowship of the Austrian Academy of Sciences awarded to T. Thalwitzer; NSF BCS award number 1734025 to EH Margulis and P. Wong; Japan Society for the Promotion of Science Grant-in-aid for Young Scientists (Grant No. 16K16483 and 20H04092) to S. Fujii; Japan Society for the Promotion of Science Grant-in-aid (Grant No. 19KK0064) and Keio Global Research Institute Startup Grant to P. Savage; Arts & Humanities Research Council (grant no. AH/N00308X/1) to Martin Clayton; Vetenskapsrådet (grant no. 2019-03694) to A. Holzapfel; John Fell Fund (grant no. 0005546) to B. Tarr; and a McDonnell Scholar award to J. McDermott.

Author contributions

N. Jacoby and J. McDermott designed the paradigm, analyzed the data, created the figures and drafted the manuscript. N. Jacoby, J. McDermott and R. Polak created the training and recruiting materials. N. Jacoby and R. Polak interpreted the data with respect to music-cultural correlations. N. Jacoby trained the group leaders in the different locations. J. Grahn and D. Cameron collected data in Botswana (9/2016-10/2016). K. Lee collected data in South Korea (1/2017-9/2017). R. Godoy, E. Undurraga, T. Huanca, M. McPherson, S. Dolan, A. Durango, N. Jacoby and J. McDermott collected data in Bolivia (6/2015-8/2019). T. Thalwitzer collected data in Brazil (9/2017-10/2017), R. Polak, N. Jacoby and N. Doumbia collected data in Mali (12/2017-1/2018). D. Goldberg and N. Jacoby collected data in Bulgaria (3/2018-7/2018). E. Margulis and P. Wong collected data in China (5/2018). L. Jure, M. Rocamora and N. Jacoby collected data in Uruguay (4/2018-5/2018). S. Fujii, P. Savage, J. Ajimi, R. Konno, and S. Oishi collected data in Japan (6/2018). K. Jakubowski collected data in the UK (5/2018-8/2018). A. Holzapfel collected data in Sweden (10/2018), E. Mungan and E. Kaya collected data in Turkey (4/2018-1/2019). P. Rao, S. Alladi, and M. Rohit collected data in India (9/2018-5/2019). B. Tarr collected data in Namibia (4/2019). N. Jacoby collected data in Boston (5/2015-12/2015) and in NYC (6/2017-9/2019). M. Anglada-Tort, P. Harrison and N. Jacoby collected online data (10/2020-3/2021). All authors reviewed the manuscript before submitting it for publication.

Acknowledgments

We thank the following individuals and organizations for their help: Carmel Raz, Pamela Smith, Sarah Nihad Nedjar, Alexandra Shana Collender Kapilian, Sarah C. Woolley, Ana María Ochoa Gautier, Felix Bernoulli, Madou Diakite, Van Yang, Flow Creative Studios, Jae Min Yu, Yeeun Lee, Esther Conde, Tsimane' council, Ted E. Gibson, John-Elmer Canfield, Pastor Roca, Salomon Hiza, Dino Nate, Ivan Domouztschiev, Ivan Ivanov, Senyao Shen, Yuchen He, Xuepei Tang, Jieqiong Che, Shuya Ma, Mingwei Zhao, Joe Lau, Wai Kit Lee, Claudia Deluca, Franco Toscano, Momoka Yamauchi, Arisa Ideguchi, Yurie Matsugi, Matthias Lichtenfeld, Martin Clayton, Tuomas Eerola, Lisa Andersson Lopéz, Adrian Berger, Emil James Robert Dickson, Karl Johannes Jondell, Edward Lindén, Jonatan Lindström, Patrik Odqvist, Thelma Svenss, Tron Gawub, Chiron Byl, Fundación Natura Bolivia. We gratefully acknowledge the memory of Prof. emeritus (University of Bamako) Dr. Salabary Doumbia (1952-2020), who contributed much to this and other related research in southern Mali, but who passed away from COVID-19 before the publication was completed.

Supplemental Information for Jacoby et al. (2021)

Figure S1: Results of the fast tempo experiment

Figure S2: Category weights from the fast tempo experiment

Figure S3: Category bias analysis

Figure S4: Principal component analysis

Figure S5: Analysis of symmetry

Figure S6: Comparison of rhythm priors in groups speaking different languages or with differing degrees of musical expertise.

Figure S7: Transmission error in all groups

Figure S8: Category weights for all groups

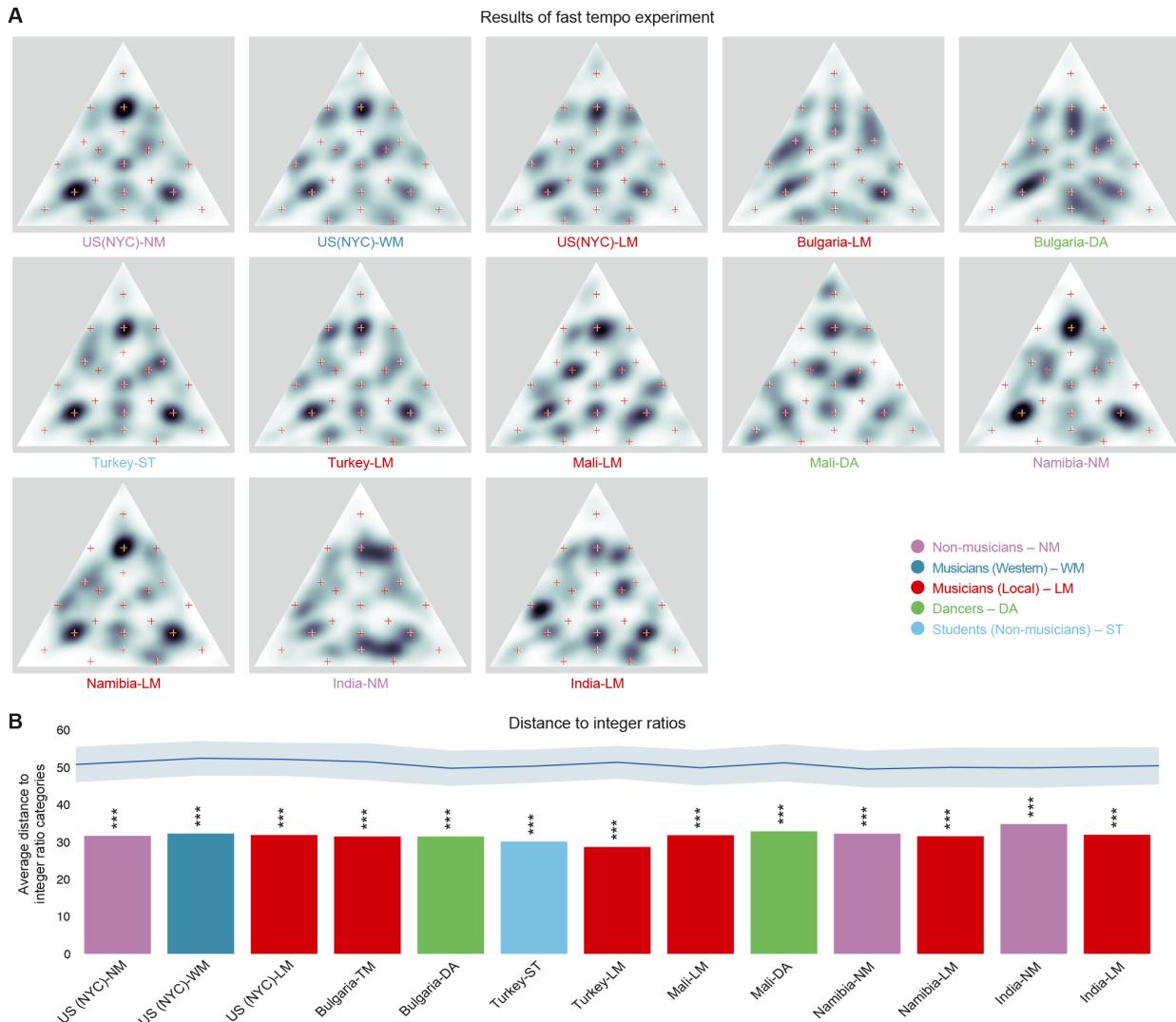


Figure S1: Results of the fast tempo experiment. **A)** Estimated priors for the 13 groups that performed an additional version of the experiment with a faster stimulus tempo (overall pattern duration of 1000 ms compared with 2000 ms in the main experiment). **B)** Average distance from nearest integer ratio rhythm, for each participant group. This measure is small for a prior with all its mass at integer ratio rhythms. Shaded region at top plots mean and 95% confidence intervals of a null distribution of the average distance from randomly selected points. Asterisks mark statistical significance after Bonferroni correction for multiple comparisons (***: $p < .001$; **: $p < .01$; *: $p < .05$). All groups have more probability mass concentrated closer to integer ratio rhythms than would be expected by chance.

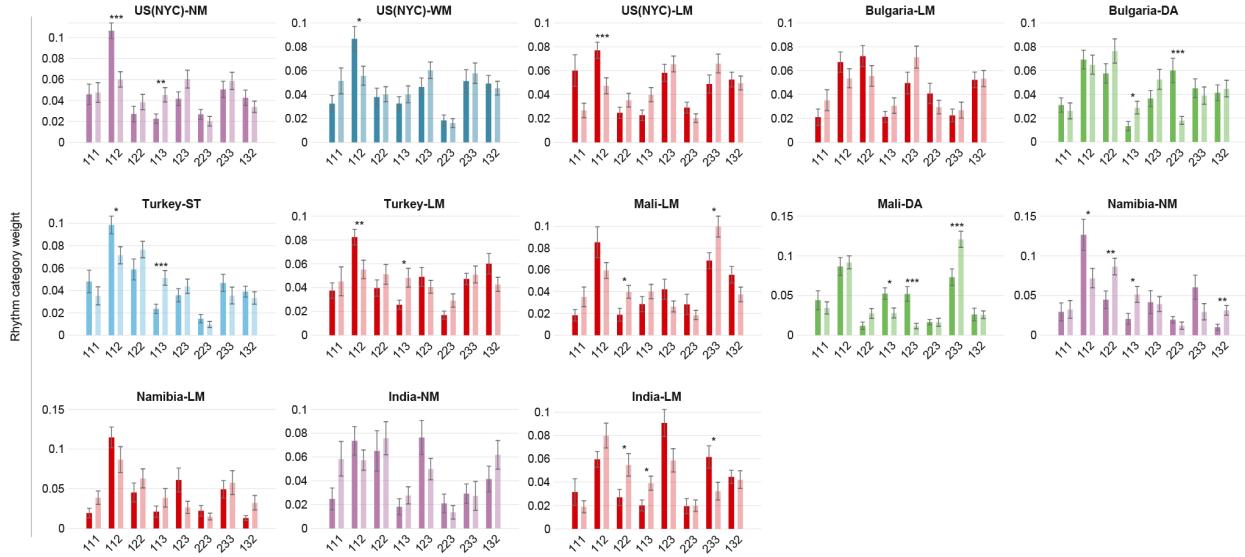


Figure S2: Category weights of the fast tempo experiment. Category weights from the fast and slow tempo experiments in 13 groups that performed both experiments. Darker colored bars plot category weights for the fast tempo; lighter colored bars plot category weights for the slow tempo. Asterisks mark statistical significance after Bonferroni correction for multiple comparisons (***: p<.001; **: p<.01; *: p<.05).

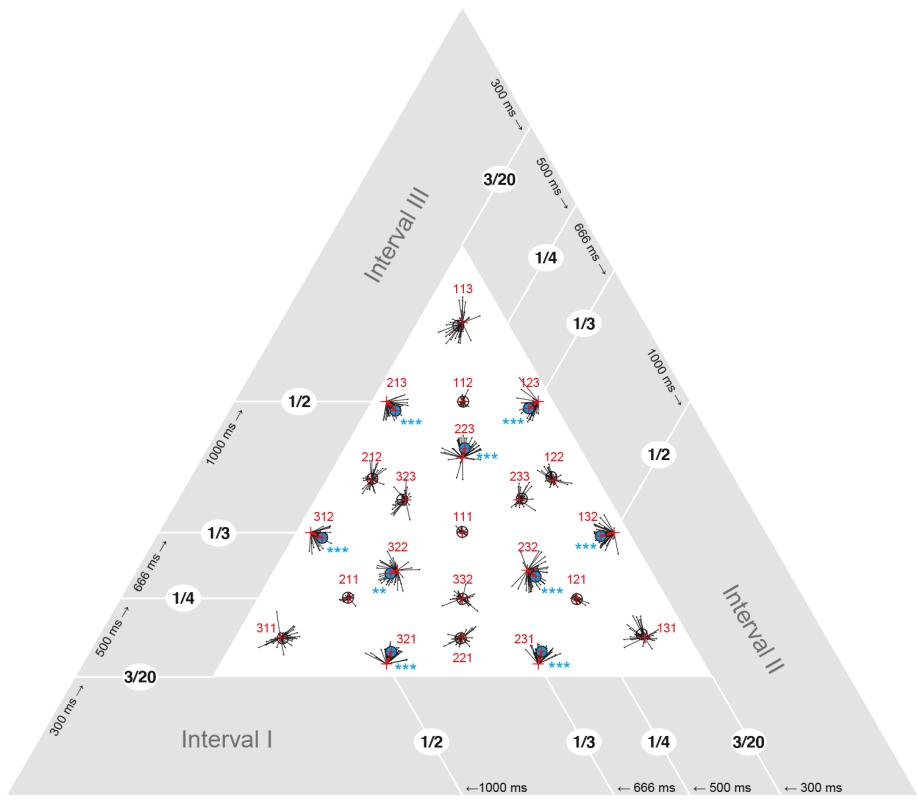


Figure S3: Category Bias analysis. Black dots plot the modes of the GMM computed for each of the 39 groups. Large circles plot the average location over the 39 groups. Asterisks denote biases that are statistically significant across groups (***: $p < .001$; **: $p < .01$; *: $p < .05$). The results show that 9 of the 22 categories exhibit consistent biases. However, the biases are relatively small.

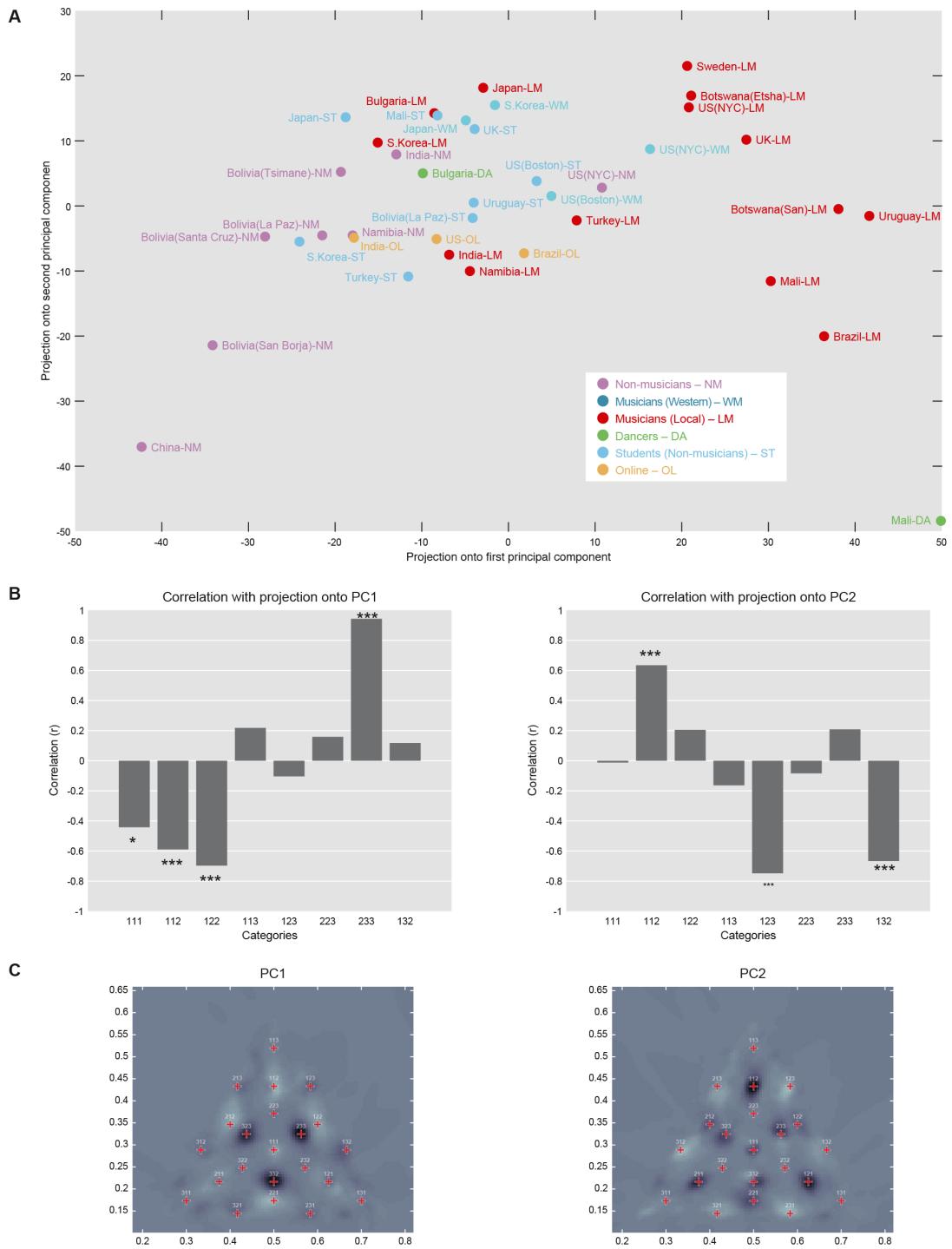


Figure S4: Principal component analysis alternative to the Multidimensional Scaling analysis of Figure 3. A) Projection of each group's prior onto the first two principal components of the measured priors. B) Correlation of the PCA projections and GMM category weights. Asterisks mark statistical significance after Bonferroni correction for multiple comparisons (***: p<.001; **: p<.01; *: p<.05). C) The first two principal components capture variation in the strength of particular integer ratio categories.

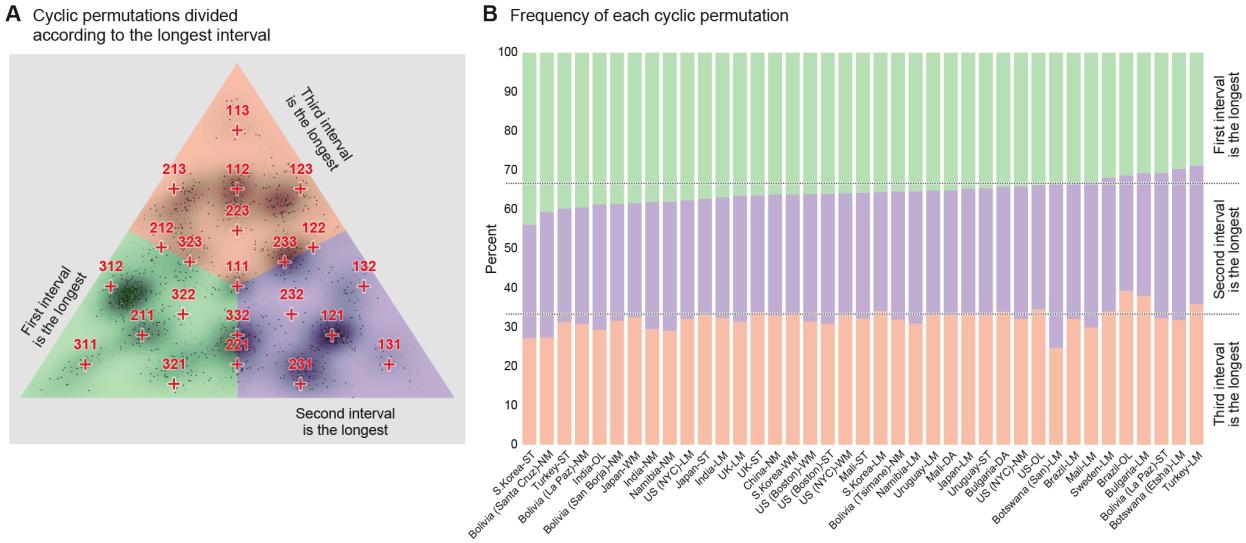


Figure S5: Analysis of symmetry. A) A partition of the triangular rhythm space into rhythms where the first, second, or third interval is the longest (green, pink, and red, respectively). The prior shown here for illustration purposes is that from Candombe musicians in Uruguay. B) Relative frequency of tapped responses in the final iteration that fell into each of the sections.

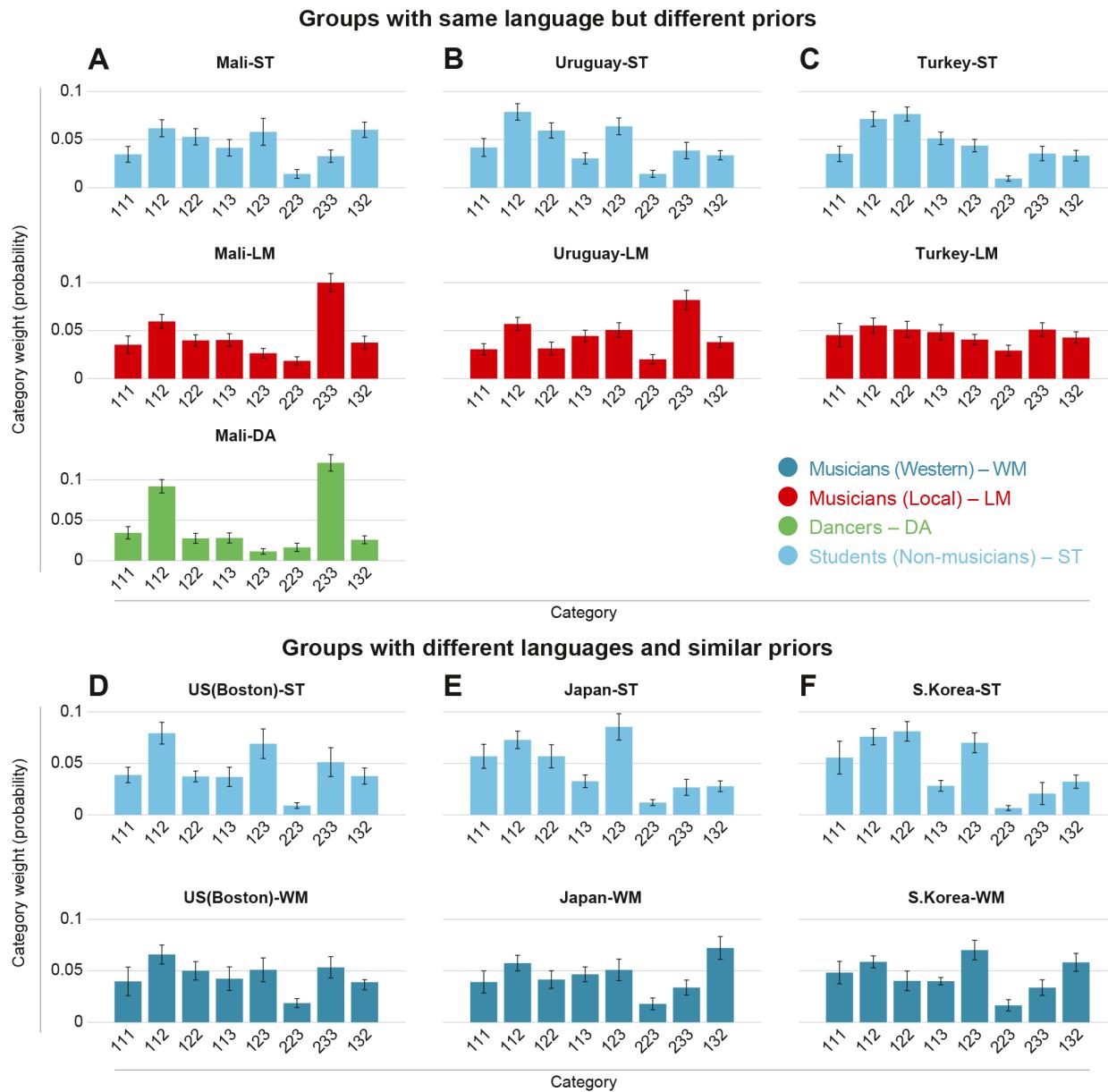


Figure S6: Comparison of rhythm priors in groups speaking different languages or with differing degrees of musical expertise. A-C) Groups speaking the same language that exhibit different priors. Graphs plot weights of different rhythm categories in a Gaussian mixture model fit to the data of each group (see section on Gaussian mixture model fits in the Methods). Error bars plot SEM, computed via bootstrap. D-F) Groups speaking different languages that exhibit similar priors. The comparison of these groups also provides an example where priors can be similar between musicians and non-musicians if the music they consume is plausibly similar (compare the student non-musicians to the Western musicians). Same plotting conventions as in A-C.

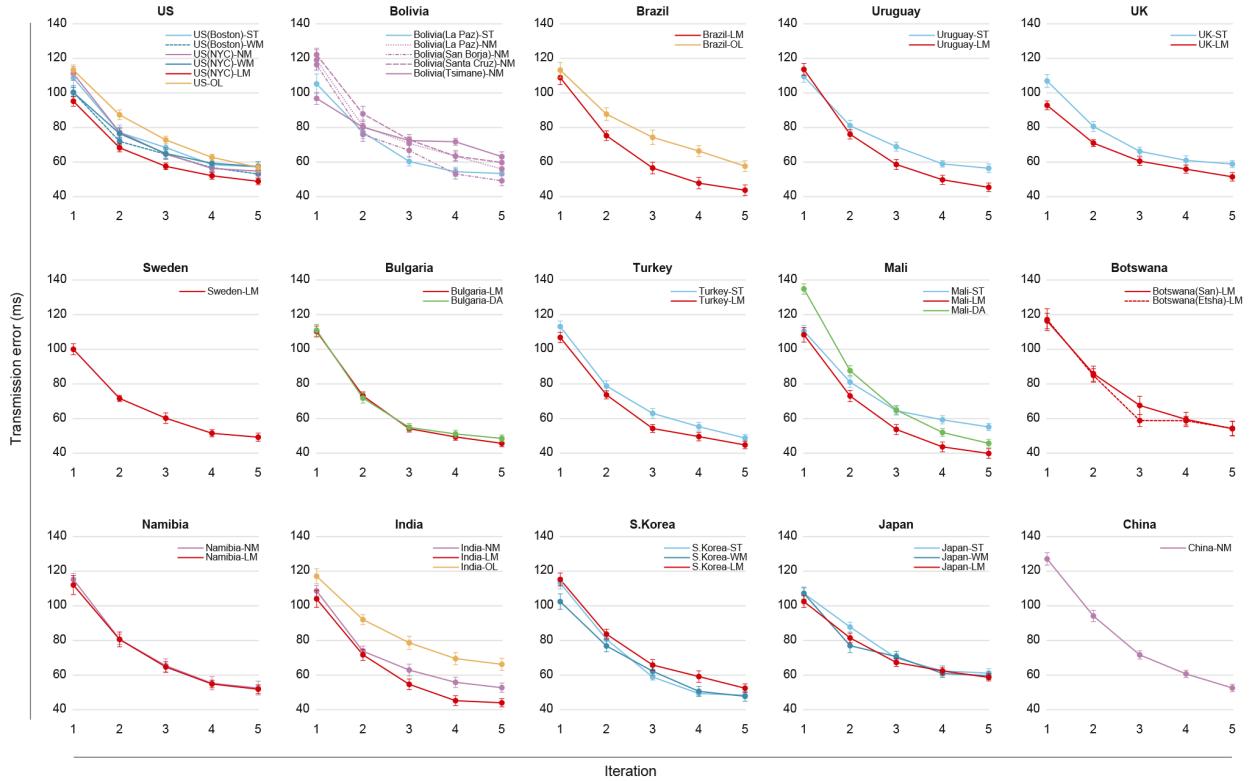


Figure S7: Transmission error in 36 groups. Transmission error is a standard way to monitor the dynamics of serial reproduction experiments. As an error measure, we use the average distance in milliseconds between the seed and response in each iteration. Namely, the average across trials of $e = \sqrt{(s_1 - r_1)^2 + (s_2 - r_2)^2 + (s_3 - r_3)^2}$, where (s_1, s_2, s_3) and (r_1, r_2, r_3) are the stimulus and average response of each iteration, respectively (i.e., the response is averaged across the 10 repetitions within each iteration). Transmission error is expected to remain constant after convergence. The figure shows that the dynamics converged or nearly converged in all 36 groups by the fifth iteration.

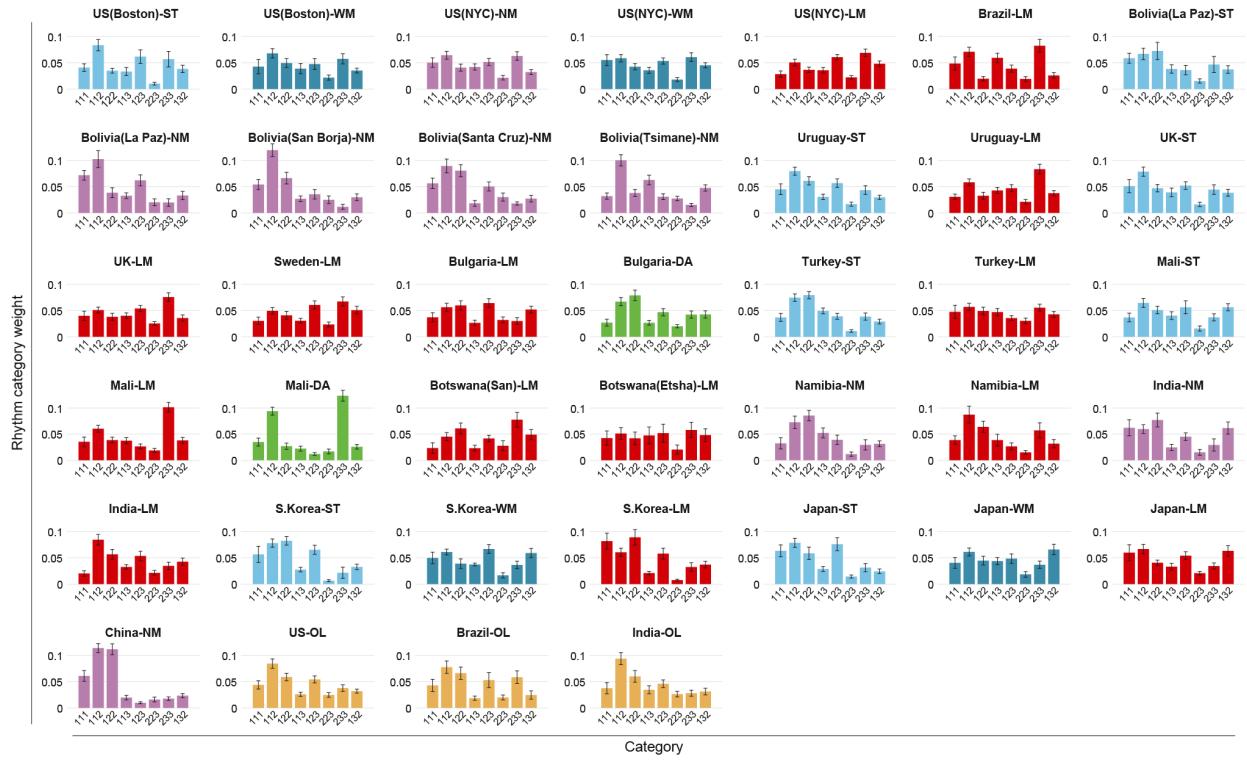


Figure S8: Category weights for all 36 groups. Category weights were obtained from fitting a constrained GMM model to the data; error bars reflect standard deviation of the weight estimate via bootstrapping (see methods).

Supplemental Tables

Table S1: Supplemental demographic information (main experiment)

#	Group	Type	Country	City	Number of Languages Spoken	3 Main Languages	Years Playing Instrument/Singing	3 Main Instruments Played	word cloud of favorite music	Years of formal education	Total Trials	Std Asynchrony
1	US.BO	ST	USA	Boston	1- 4 (mean = 1.87 std= 0.92)	English (100%), Spanish (47%), French (20%)	0- 4 (mean = 0.93 std= 1.17)	None (47%), Piano (27%), Guitar (20%)	Not Reported	12-17 (mean = 15.13 std= 1.55)	431	40-72 (mean = 50.06 std= 10.41)
2	US.BO	WM	USA	Boston	1- 3 (mean = 1.57 std= 0.76)	English (100%), German (14%), Mandarin (14%)	9-47 (mean = 18.29 std= 9.54)	Piano (43%), Guitar (29%), Clarinet (14%)	Not Reported	12-23 (mean = 16.36 std= 3.37)	529	37-51 (mean = 42.54 std= 4.23)
3	US.NY	NM	USA	NYC	1- 3 (mean = 1.23 std= 0.59)	English (100%), Spanish (8%), French (4%)	0- 9 (mean = 1.73 std= 2.86)	None (58%), Guitar (12%), Piano (12%)	drake, jackson, rock, ross	12-21 (mean = 15.62 std= 2.39)	633	29-63 (mean = 47.74 std= 7.10)
4	US.NY	WM	USA	NYC	1- 5 (mean = 2.07 std= 1.17)	English (100%), French (30%), Spanish (26%)	10-42 (mean = 20.04 std= 8.66)	Piano (70%), Violin (30%), Guitar (15%)	classical, pop, rock, beethoven	12-24 (mean = 17.63 std= 2.96)	672	27-56 (mean = 40.13 std= 6.63)
5	US.NY	LM	USA	NYC	1- 3 (mean = 1.61 std= 0.77)	English (100%), Spanish (31%), French (6%)	8-59 (mean = 24.08 std= 12.62)	Piano (33%), Guitar (19%), Trombone (14%)	jazz, american, classical, davis	12-25 (mean = 17.81 std= 2.86)	811	28-58 (mean = 38.93 std= 6.82)
6	BR	LM	Brazil	Recife	1- 2 (mean = 1.07 std= 0.26)	Portuguese (100%), English (7%)	5-20 (mean = 12.53 std= 4.78)	Brazilian percussion (100%),Guitar (13%), Cavaquinho (7%),	tudo, rock, brega, menos	9-20 (mean = 12.85 std= 3.39)	433	32-46 (mean = 38.96 std= 4.49)
7	BO.LP	ST	Bolivia	La Paz	2- 4 (mean = 2.38 std= 0.65)	English (100%), Spanish (100%), Portuguese (23%)	0-10 (mean = 3.42 std= 3.48)	Piano (38%), Guitar (31%), None (31%)	rock, spanish, electronic, english	13-18 (mean = 15.62 std= 1.80)	299	37-53 (mean = 46.15 std= 5.13)
8	BO.LP	NM	Bolivia	La Paz	1- 3 (mean = 2.17 std= 0.72)	Spanish (100%), Aymara (50%), English (42%), Quechua (17%)	0- 3 (mean = 0.79 std= 1.12)	None (58%), Drums (17%), Charango (8%)	cumbia, folk, bolivian, romantic	4-20 (mean = 12.96 std= 5.06)	304	41-60 (mean = 51.42 std= 6.65)
9	BO.SB	NM	Bolivia	San Borja	1- 1 (mean = 1.00 std= 0.00)	Spanish (100%)	0-12 (mean = 1.76 std= 3.67)	Sing (65%), Non (24%), Drum (12%)	romantic, cumbia, ranchera, pop	5-19 (mean = 11.35 std= 4.03)	386	40-69 (mean = 53.07 std= 8.12)
10	BO.SC	NM	Bolivia	Santa Cruz	1- 1 (mean = 1.00 std= 0.00)	Spanish (100%)	0- 9 (mean = 1.24 std= 2.41)	Sing (65%), None (24%), Wind (24%)	rock, pop, romantic, reggeton	12-22 (mean = 15.41 std= 2.62)	376	46-62 (mean = 53.06 std= 4.79)
11	BO.TS	NM	Bolivia	Tsimane' Villages	1- 2 (mean = 1.23 std= 0.43)	Tsimane' (100%), Spanish (23%)	2 participants reported regularly playing an instrument. 34 participants reported playing an instrument at least once.	Drum (64%), Ricarica (64%), None (31%)	hymns, traditional	0-12 (mean = 3.42 std= 2.87)	792	54-83 (mean = 70.32 std= 8.11)
12	UY	ST	Uruguay	Montevideo	1- 4 (mean = 2.44 std= 0.64)	Spanish (100%), English (8%), Portuguese (26%)	0- 0 (mean = 0.00 std= 0.00)	None (100%)	rock, usa, england, uruguay	15-22 (mean = 17.41 std= 1.80)	675	33-62 (mean = 46.66 std= 6.81)
13	UY	LM	Uruguay	Montevideo	1- 3 (mean = 1.32 std= 0.60)	Spanish (100%), Portuguese (19%), English (10%), Yoruba (3%)	7-60 (mean = 34.65 std= 13.51)	Candombe-Drums (77%), Percussion (61%), Piano (10%)	uruguay, candombe, salsa, cuba	6-18 (mean = 9.45 std= 2.63)	776	28-51 (mean = 37.55 std= 5.97)
14	UK	ST	UK	Durham	1- 4 (mean = 2.12 std= 0.95)	English (100%), French (21%), Mandarin (17%)	0- 8 (mean = 1.29 std= 1.88)	None (75%), Guitar (17%), Piano (17%)	pop, rock, indie, coldplay	12-21 (mean = 16.25 std= 2.17)	480	39-61 (mean = 50.55 std= 5.37)
15	UK	LM	UK	Durham	1- 4 (mean = 1.48 std= 0.83)	English (100%), French (17%), German (7%)	6-45 (mean = 17.17 std= 9.93)	Piano (45%), Guitar (28%), Drums (21%)	jazz, rock, pop, funk	12-20 (mean = 16.41 std= 2.26)	580	27-51 (mean = 38.83 std= 5.70)
16	UK	LM	Sweden	Stockholm	2- 6 (mean = 2.95 std= 1.09)	English (100%), Swedish (100%), German (27%)	10-40 (mean = 21.59 std= 10.28)	Violin (59%), Voice (32%), Piano (23%)	folk, all, swedish, bach	12-2.2e+02 (mean = 27.40 std= 43.15)	455	30-47 (mean = 38.88 std= 4.33)
17	BG	LM	Bulgaria	Pleven/Pl ovdv/Sofia	1- 4 (mean = 2.30 std= 0.88)	Bulgarian (100%), English (61%), Russian (48%)	10-50 (mean = 29.52 std= 11.41)	Gadulka (35%), Piano (26%), Kaval (22%)	bulgarian, rock, jazz, player	12-18 (mean = 15.13 std= 2.40)	687	31-44 (mean = 37.78 std= 3.76)
18	BG	DA	Bulgaria	Pleven/Pl ovdv/Sofia	1- 5 (mean = 1.93 std= 0.87)	Bulgarian (100%), English (37%), Russian (26%)	0-34 (mean = 3.11 std= 7.29)	None (56%), Tupan (15%), Kaval (11%)	bulgarian, rock, folk, folklore	12-32 (mean = 17.04 std= 5.61)	780	33-56 (mean = 42.29 std= 6.12)
19	TK	ST	Turkey	Istanbul	2- 3 (mean = 2.19 std= 0.40)	English (100%), Turkish	0-11 (mean = 4.31 std= 3.59)	None (15%), Baglama (11%), Guitar (11%)	rock, pop, turkish, folk	13-21 (mean = 14.89 std= 1.89)	746	31-59 (mean = 43.84 std= 6.58)

					(100%), French (11%),							
20	TK	LM	Turkey	Istanbul / Ismir	1- 4 (mean = 1.96 std= 0.86)	Turkish (100%), English (67%), German (8%)	9-30 (mean = 17.00 std= 5.85)	Baglama (25%), Oud (21%), Bendir (13%),	pop, turkish, folk, rock	14-25 (mean = 18.71 std= 3.38)	617	31-45 (mean = 37.87 std= 4.34)
21	MA	ST	Mali	Bamako	2- 7 (mean = 3.45 std= 1.09)	French (100%), Bambara (97%), English (61%), Sounrai (16%)	0-15 (mean = 1.10 std= 3.45)	None (84%), Voice (6%),	malian, rap, pop, diabate	13-20 (mean = 15.68 std= 1.62)	636	33-60 (mean = 45.17 std= 6.43)
22	MA	LM	Mali	Bamako	1- 4 (mean = 2.03 std= 0.78)	Bambara (100%), French (69%), Kasonka (14%)	7-57 (mean = 25.93 std= 12.45)	Dundun (97%), Jembe (90%), Ngoni (14%),	malian, traditional, mali, singer	0-17 (mean = 6.10 std= 4.20)	589	22-51 (mean = 33.83 std= 6.16)
23	MA	DA	Mali	Sagele	1- 2 (mean = 1.09 std= 0.29)	Maninka (100%), French (9%)	0- 3 (mean = 0.09 std= 0.51)	None (47%), Ji-Dunun (29%), Voice (18%),	malian, singer, diabate, keita	0-12 (mean = 3.85 std= 3.89)	834	32-57 (mean = 40.47 std= 5.32)
24	BW.SA	LM	Botswana	San	1- 2 (mean = 1.59 std= 0.51)	Other (100%), English (59%)	10-59 (mean = 29.73 std= 14.73)	Voice (59%), Tagadire (35%), Thumbpiano (18%)	Not Reported	Not reported	170	26-49 (mean = 39.72 std= 5.75)
25	BW.EA	LM	Botswana	Etsha	1- 2 (mean = 1.14 std= 0.36)	Other (100%), English (14%)	7-46 (mean = 16.36 std= 9.79)	Drums (71%), Voice (50%), Fingerguitar (7%)	Not Reported	Not reported	127	34-52 (mean = 42.53 std= 5.02)
26	NA	NM	Namibia	Spitzkopp e	1- 3 (mean = 1.75 std= 0.58)	Afrikaans (94%), English (75%), Oshivambo (6%)	0- 8 (mean = 3.91 std= 2.76)	Singing (88%), Drums (13%), None (13%)	malgaisa, damara, gospel, staika	Not reported	370	41-67 (mean = 47.82 std= 6.51)
27	NA	LM	Namibia	Spitzkopp e	1- 3 (mean = 2.00 std= 0.55)	English (93%), Afrikaans (79%), German (14%)	4-20 (mean = 10.86 std= 4.19)	Singing (100%), Drums (21%), Trumpet (14%),	damara, staika, rnb, choir	Not reported	344	39-59 (mean = 46.45 std= 5.21)
28	IN	NM	India	Mumbai	2- 5 (mean = 2.73 std= 0.96)	English (100%), Hindi (80%), Marathi (33%),	0- 1 (mean = 0.38 std= 0.48)	None (87%), Guitar (7%), Keyboard (7%)	arijit, rahman, singh, ghoshal	14-23 (mean = 17.60 std= 2.97)	378	38-56 (mean = 47.71 std= 5.80)
29	IN	LM	India	Mumbai	1- 4 (mean = 2.88 std= 0.86)	English (94%), Hindi (94%), Marathi (35%)	6-25 (mean = 11.68 std= 5.24)	Tabla (47%), Voice (18%), Flute (12%),	khan, ustad, hussain, zakir	4-26 (mean = 16.76 std= 5.47)	426	32-47 (mean = 37.64 std= 4.77)
30	KR	ST	S.Korea	Daejeon	1- 1 (mean = 1.00 std= 0.00)	Not-reported (100%)	0- 9 (mean = 3.00 std= 3.48)	None (40%), Piano (40%), Guitar (7%)	pop, classical, hip, hop	Not reported	450	35-55 (mean = 44.33 std= 4.83)
31	KR	WM	S.Korea	Daejeon	1- 1 (mean = 1.00 std= 0.00)	Not-reported (100%)	5-22 (mean = 12.15 std= 5.34)	Piano (31%), String (31%), Vocal (31%)	classical, pop, jazz, hip	Not reported	390	33-59 (mean = 42.95 std= 6.25)
32	KR	LM	S.Korea	Jeonju	1- 1 (mean = 1.00 std= 0.00)	Not-reported (100%)	5-15 (mean = 8.23 std= 2.74)	Wind (54%), Percussion (15%), String (15%)	pop, hip, hop, jazz	Not reported	388	38-50 (mean = 44.96 std= 3.55)
33	JP	ST	Japan	Tokyo / Fujisawa	1- 4 (mean = 1.65 std= 0.99)	Japanese (100%), English (30%), Chinese (5%)	0- 3 (mean = 0.35 std= 0.88)	None (75%), Acoustic Guitar (5%), Piano (5%)	Not Reported	Not reported	428	40-62 (mean = 50.13 std= 5.41)
34	JP	WM	Japan	Tokyo / Fujisawa	1- 4 (mean = 2.23 std= 1.09)	Japanese (100%), English (69%), Chinese (8%)	5-20 (mean = 12.77 std= 5.18)	Piano (77%), Vocal (23%), Wadaiko (23%)	Not Reported	Not reported	261	39-52 (mean = 44.51 std= 3.73)
35	JP	LM	Japan	Tokyo / Fujisawa	1- 4 (mean = 1.58 std= 0.99)	Japanese (100%), English (27%), French (4%)	7-21 (mean = 13.35 std= 3.97)	Shamisen (73%), Koto (46%), Piano (38%),	Not Reported	Not reported	529	37-56 (mean = 45.69 std= 4.94)
36	CN	NM	China	Guizhou Province	1- 2 (mean = 1.77 std= 0.42)	Dong (100%), Mandarin (78%)	0-26 (mean = 1.68 std= 5.23)	None (80%), Voice (20%)	Not Reported	Not reported	575	40-64 (mean = 50.25 std= 5.93)
37	US	OL	USA	Online	1- 4 (mean = 1.46 std= 0.79)	english (100%), spanish (11%), german (6%),	0- 30 (mean = 5.06 std= 6.75)	none (46%), piano (26%), guitar (17%),	rock, pop, alternative, taylor	8 - 22 (mean = 16.25 std= 2.78)	948	30.6 - 71.5 (mean = 52.41 std= 9.48)
38	BR	OL	Brazil	Online	1- 4 (mean = 2.39 std= 0.69)	portuguese (100%), english (93%), spanish (36%),	0 - 20 (mean = 3.52 std= 5.27)	none (43%), guitar (32%), piano (14%),	rock, pop, brazilian, metal	11 - 25 (mean = 16.19 std= 3.57)	413	40 - 75.2 (mean = 55.79 std= 8.28)
39	IN	OL	India	Online	2- 5 (mean = 3.11 std= 0.92)	english (100%), hindi (76%), tamil (47%), malayalam (24%)	0 - 20 (mean = 3.06 std= 5.02)	none (53%), guitar (21%), harmonium (8%)	pop, rock, classical, rahman	8 - 22 (mean = 17.48 std= 2.84)	569	31.9 - 87.3 (mean = 62.63 std= 11.84)

Table S3: Supplemental demographic information (fast tempo experiment)

#	Group	Type	Country	City	Number Of Languages Spoken	3 Main Languages	Years Playing Instrument/Singing	Main Instruments Played	Most Mentioned Favorite Artists or Musical Genres	years of formal education	Total Trials	Std Asynchrony
1	US.NY	NM	USA	NYC	1- 3 (mean = 1.23 std= 0.59)	English (100%), Spanish (8%), French (4%), Hebrew (4%), Italian (4%), Portuguese (4%)	0- 9 (mean = 1.73 std= 2.86)	None (58%), guitar (12%), piano (12%)	Drake (19%), Stevie Wonder (12%), Beyoncé (8%), Diana Ross (8%), Kendrick Lamar (8%),	12-21 (mean = 15.62 std= 2.39)	631	18-33 (mean = 25.74 std= 3.55)
2	US.NY	WM	USA	NYC	1- 4 (mean = 1.96 std= 1.04)	English (100%), French (27%), Spanish (23%), German (12%), Italian (12%), Mandarin (12%), Arabic (4%), Cantonese (4%), Chinese (4%)	10-42 (mean = 19.69 std= 8.64)	Piano (73%), violin (27%), guitar (15%)	Classical (27%), Bruno Mars (8%), jazz (8%), Lady Gaga (8%), Madonna (8%)	12-24 (mean = 17.54 std= 2.98)	636	17-30 (mean = 22.24 std= 3.17)
3	US.NY	LM	USA	NYC	1- 3 (mean = 1.59 std= 0.74)	English (100%), Spanish (29%), French (6%), Hebrew (6%), Russian (6%), Chinese (3%), Dutch (3%), German (3%), Italian (3%)	8-59 (mean = 24.65 std= 12.76)	Piano (32%), guitar (21%), bass (12%)	Bach (9%), Miles Davis (9%), Charles Mingus (6%), Frank Ocean (6%), Jazz (6%), Prince (6%)	12-25 (mean = 17.83 std= 2.95)	773	14-31 (mean = 21.94 std= 3.57)
4	BG	LM	Bulgaria	Pleven/Pl ovdv/Sof ia	1- 4 (mean = 2.30 std= 0.88)	Bulgarian (100%), English (61%), Russian (48%), German (13%), Hungarian (4%), Turkish (4%)	10-50 (mean = 29.52 std= 11.41)	Gadulka (35%), piano (26%), kaval (22%), voice (17%)	Bulgarian Folklore (87%), Jazz (39%), Rock (39%), Pop (26%), Bulgarian Prog Rock (4%), Bulgarian Rock	12-18 (mean = 15.13 std= 2.40)	641	17-25 (mean = 20.27 std= 2.01)
5	BG	DA	Bulgaria	Pleven/Pl ovdv/Sof ia	1- 5 (mean = 2.03 std= 1.00)	Bulgarian (100%), English (44%), Russian (31%), Dutch (6%), German (6%), French (3%), Greek (3%), Italian (3%), Turkish (3%), Ukrainian (3%)	0-34 (mean = 3.03 std= 6.76)	None (53%), tupan (19%), kaval (9%), voice (6%)	Bulgarian Folklore (44%), Pop (25%), Rock (16%), Greek Folklore (9%), Metal (9%)	12-32 (mean = 17.25 std= 5.25)	961	15-28 (mean = 22.43 std= 2.98)
6	TK	ST	Turkey	Istanbul	2- 3 (mean = 2.19 std= 0.40)	Turkish (100%), English (100%), French (11%), German (7%)	0-11 (mean = 4.31 std= 3.59)	None (15%), baglama (11%), guitar (11%), ney (11%)	Turkish Folk (30%), Pop/Rock (26%), Folk (7%), Pop (7%), Turkish Pop (7%), Turkish-Rock (7%)	13-21 (mean = 14.89 std= 1.89)	741	17-29 (mean = 23.14 std= 3.14)
7	TK	LM	Turkey	Istanbul / Izmir	1- 4 (mean = 1.96 std= 0.86)	Turkish (100%), English (67%), German (8%), (4%), Arabic (4%), Bosniak (4%), Greek (4%), Japanese (4%)	9-30 (mean = 17.00 std= 5.85)	Baglama (25%), oud (21%), bendir (13%), clarinet (13%)	Turkish Folk (21%), Turkish Pop (17%), Pop/Rock (13%), Turkish-Classical (13%), Turkish-Rock/Folk (13%)	14-25 (mean = 18.71 std= 3.38)	616	16-24 (mean = 20.49 std= 2.11)
8	MA	LM	Mali	Bamako	1- 4 (mean = 2.00 std= 0.86)	Bambara (95%), French (65%), (10%), Bambar (5%), Djoula (5%), Kasouka (5%), Mandekan (5%), Maninaka (5%), Peul (5%)	7-42 (mean = 23.75 std= 11.62)	Dundun (95%), jembe (90%), tamalan (25%), ngoni (10%)	Traditional Malian (95%), Malian Pop (40%), Alpha Blondy (5%), Guinean Pop (5%), Hip hop (5%), Malian Reggae (5%)	0-17 (mean = 7.05 std= 4.31)	606	14-22 (mean = 16.76 std= 2.32)
9	MA	DA	Mali	Sagele	1- 2 (mean = 1.08 std= 0.28)	Maninka (100%), French (8%)	0- 3 (mean = 0.12 std= 0.61)	None (50%), ji-dunun (25%), voice (17%), dundun (4%)	Traditional Malian (79%), Guinean Pop (8%), Malian Pop (8%), Malian Rap (8%), American Rap (4%), French Rap (4%), Traditional-Guinea (4%)	0-12 (mean = 2.96 std= 3.83)	604	18-28 (mean = 22.67 std= 2.23)
10	NA	NM	Nambia	Spitzkopp e	1- 3 (mean = 1.75 std= 0.58)	Afrikaans (94%), English (75%), Oshivambo (6%)	0- 8 (mean = 3.91 std= 2.76)	Singing (88%), drums (13%), none (13%)	Gospel (19%), Malgaïsa (19%), Ma/Aisa (13%), Damara (6%), Damaramusic: Damara punch (6%), House (6%), Mal Eisa (6%), No interest in music (6%), Pop (6%), Rap (6%), Slow Jam (6%)	not reported	361	21-34 (mean = 25.68 std= 3.37)
11	NA	LM	Namibia	Spitzkopp e	1- 3 (mean = 2.00 std= 0.55)	English (93%), Afrikaans (79%), German (14%), Afrikaans (7%), Otjiherero (7%)	4-20 (mean = 10.86 std= 4.19)	Singing (100%), drums (21%), trumpet (14%)	R&B (14%), Bob Marley (7%), Damara (7%), Gospel (7%)	not reported	341	22-32 (mean = 25.78 std= 2.97)
12	IN	NM	India	Mumbai	2- 5 (mean = 2.73 std= 0.96)	English (100%), Hindi (80%), Marathi (33%), (13%), Elementary German & Persian (7%), Gujarati (7%), Malay (7%), Marwadi (7%), Odia (7%), Sanskrit (7%), Tamil (7%)	0- 1 (mean = 0.38 std= 0.48)	None (87%), guitar (7%), keyboard (7%)	Arijit Singh (20%), Arrahman (20%), Abida Parveen (7%), Amit Trivedi (7%), Hindi and Urdu movie singers (7%)	14-23 (mean = 17.60 std= 2.97)	376	22-29 (mean = 25.31 std= 2.59)
13	IN	LM	India	Mumbai	2- 4 (mean = 2.75 std= 0.62)	English (100%), Hindi (100%), Marathi (42%), (8%), Kannada (8%), Telugu (8%), Tulu (8%)	7-20 (mean = 12.25 std= 3.74)	Tabla (67%), flute (17%), harmonium (17%)	Ustad Zakir Hussain (34%), Hariprasad Chaurasia (17%), Indian Ocean (17%), Agam (8%)	11-21 (mean = 16.00 std= 3.13)	300	16-23 (mean = 19.86 std= 1.93)

References

1. Patel AD (2008) *Music, Language, and the Brain* (Oxford University Press, Oxford).
2. Sundberg J, Prame E, & Iwarsson J (1996) Replicability and accuracy of pitch patterns in professional singers. *Vocal fold physiology, controlling complexity and chaos*, eds Davis PJ & Fletcher NH (Singular Publishing Group, San Diego, CA).
3. Repp BH (2005) Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review* 12(6):969-992.
4. Vurma A & Ross J (2006) Production and perception of musical intervals. *Music Perception* 23(4):331-344.
5. Larrouy-Maestri P & Morsomme D (2014) Criteria and tools for objectively analysing the vocal accuracy of a popular song. *Logopedics, Phoniatrics, Vocology* 39:11-18.
6. Roeske TC, Tchernichovski O, Poeppel D, & Jacoby N (2020) Categorical rhythms are shared between songbirds and humans. *Current Biology* 30(18):3544-3555.
7. Siegel JA & Siegel W (1977) Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics* 21(5):399-407.
8. Clarke EF (1987) Categorical rhythm perception: An ecological perspective. *Action and Perception in Rhythm and Music*, ed Gabrielsson A (Royal Swedish Academy of Music, Stockholm, Sweden), pp 19-33.
9. Desain P & Honing H (2003) The formation of rhythmic categories and metric priming. *Perception* 32:341-365.
10. Sadakata M, Desain P, & Honing H (2006) The Bayesian way to relate rhythm perception and production. *Music Perception* 23:269-288.
11. Goldstone RL & Hendrickson AT (2010) Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(1):69-78.
12. Lomax A & Berkowitz N (1972) The evolutionary taxonomy of culture. *Science* 177:228-239.
13. Brown S & Jordania J (2013) Universals in the world's musics. *Psychology of Music* 41(2):229-248.
14. Savage PE, Brown S, Sakai E, & Currie TE (2015) Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences* 112(29):8987-8992.
15. Mehr SA, et al. (2019) Universality and diversity in human song. *Science* 366(6468):eaax0868.

16. Savage PE (2020) An overview of cross-cultural music corpus studies. *Oxford Handbook of Music and Corpus Studies*, eds Shanahan D, Burgoyne A, & Quinn I (Oxford University Press, New York).
17. Kessler EJ, Hansen C, & Shepard RN (1984) Tonal schemata in the perception of music in Bali and in the West. *Music Perception* 2(2):131-165.
18. Perlman M & Krumhansl CL (1996) An experimental study of internal interval standards in Javanese and Western musicians. *Music Perception* 14:95-116.
19. Krumhansl CL, *et al.* (2000) Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition* 76(1):13-58.
20. Curtis ME & Bharucha JJ (2009) Memory and musical expectation for tones in cultural context. *Music Perception* 26(4):365-375.
21. Fritz T, *et al.* (2009) Universal recognition of three basic emotions in music. *Current Biology* 19:1-4.
22. Laukka P, Eerola T, Thringujam NS, Yamasaki T, & Beller G (2013) Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion* 13(3):434-449.
23. Egermann H, Fernando N, Chuen L, & McAdams S (2015) Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese pygmies. *Frontiers in Psychology* 5:1341.
24. McDermott JH, Schultz AF, Undurraga EA, & Godoy RA (2016) Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535:547-550.
25. Jacoby N & McDermott JH (2017) Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology* 27:359-370.
26. Polak R, *et al.* (2018) Rhythmic prototypes across cultures. A comparative study of tapping synchronization. *Music Perception* 36(1):1-23.
27. Jacoby N, *et al.* (2019) Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology* 29:3229-3243.
28. McPherson MJ, *et al.* (2020) Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nature Communications* 11:2786.
29. Athanasopoulos G, Eerola T, Lahdelma I, & Kaliakatsos-Papakostas M (2021) Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PLoS ONE* 16:e0244964.

30. Locke S & Kellar L (1973) Categorical perception in a non-linguistic mode. *Cortex* 9(4):355-369.
31. Burns EM & Ward WD (1978) Categorical perception - phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America* 63:456-468.
32. Povel DJ (1981) Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance* 7:3-18.
33. Bartlett FC (1932) *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press).
34. Griffiths TL & Kalish ML (2007) Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31:441-480.
35. Cooper GW & Meyer LB (1960) *The rhythmic structure of music* (University of Chicago Press, Chicago).
36. Feldman NH, Griffiths TL, & Morgan JL (2009) The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116(4):752-782.
37. Henrich J, Heine SJ, & Norenzayan A (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33:61-83.
38. Barrett HC (2020) Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends Cogn. Sci.* 24:620-638.
39. Gabrielsson A, Bengtsson I, & Gabrielsson B (1983) Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology* 24(1):193-213.
40. Mead A (1992) Review of the development of multidimensional scaling. *Journal of the Royal Statistical Society: Series D (The Statistician)* 41(1):27-39.
41. Jacoby N, et al. (2020) Cross-cultural work in music cognition: Challenges, insights and recommendations. *Music Perception* 37(3):185-195.
42. Holzapfel A (2015) Relation between surface rhythm and rhythmic modes in Turkish makam music. *Journal of New Music Research* 44:25-38.
43. England NM (1995) *Music among the Zū'wāsi and related peoples of Namibia, Botswana, and Angola* (Garland, New York).
44. Goldberg D (2015) Timing variations in two Balkan percussion performances. *Empirical Musicology Review* 10:305-328.

45. Hannon EE, Soley G, & Ullal S (2012) Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance* 38:543-548.
46. Hannon EE & Trehub SE (2005) Metrical categories in infancy and adulthood. *Psychological Science* 16(1):48-55.
47. Kubik G (2001) Africa. *Grove Music Online*.
48. Agawu VK (2006) Structural analysis or cultural analysis? Competing perspectives on the “Standard Pattern” of West African rhythm. *Journal of the American Musicological Society* 59(1):1-46.
49. Kubik G (1979) *Angolan Traits in Black Music, Games and Dances of Brazil: A Study of African Cultural Extensions Overseas* (Junta de Investigações Científicas do Ultramar).
50. Washburne C (1997) The clave of jazz: A Caribbean contribution to the rhythmic foundation of an African-American music. *Black Music Research Journal* 17(1):59-80.
51. Toussaint GT (2013) *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?* (CRC Press).
52. Rocamora M (2018) Computational methods for percussion music analysis: The Afro-Uruguayan Candombe drumming as a case study PhD (Universidad de la República Montevideo, Uruguay).
53. Spiegel MF & Watson CS (1984) Performance on frequency-discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America* 76:1690-1695.
54. Kishon-Rabin L, Amir O, Vexler Y, & Zaltz Y (2001) Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic Clinical Physiology and Pharmacology* 12:125-143.
55. Fujioka T, Trainor LJ, Ross B, Kakigi R, & Pantev C (2004) Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience* 16(6):1010-1021.
56. McPherson MJ & McDermott JH (2018) Diversity in pitch perception revealed by task dependence. *Nature Human Behavior* 2:52-66.
57. Getz LM, Barton S, & Kubovy M (2014) The specificity of expertise: For whom is the clave pattern the “key” to salsa music? *Acta Psychologica* 152:56-66.
58. Patel AD & Daniele JR (2003) An empirical comparison of rhythm in language and music. *Cognition* 87:B35-B45.

59. Norman-Haignere S, Kanwisher N, & McDermott JH (2015) Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88:1281-1296.
60. Nettl B (2000) An ethnomusicologist contemplates universals in musical sound and musical culture. *The Origins of Music*, eds Wallin NL, Merker B, & Brown S (MIT Press, Cambridge, MA).
61. Smith K, Kalish ML, Griffiths TL, & Lewandowsky S (2008) Introduction. Cultural transmission and the evolution of human behaviour. *Philosophical Transactions of the Royal Society B* 363:3469-3476.
62. Xu J, Griffiths T, & Dowman M (2010) Replicating color term universals through human iterated learning. in *The Annual Meeting of the Cognitive Science Society*.
63. Ravignani A, Delgado T, & Kirby S (2017) Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behavior* 1:1-7.
64. Zaslavsky N, Kemp C, Regier T, & Tishby N (2018) Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences* 115(31):7937-7942.
65. Carr JW, Smith K, Culbertson J, & Kirby S (2020) Simplicity and informativeness in semantic category systems. *Cognition* 202:104289.
66. Tierney AT, Russo FA, & Patel AD (2011) The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences* 108(37):15510-15515.
67. Tillmann B, Bharucha JJ, & Bigand E (2000) Implicit learning of tonality: a self-organizing approach. *Psychological Review* 107(4):885-913.
68. Berlin B & Kay P (1969) *Basic Color Terms: Their Universality and Evolution* (University of California Press, Berkeley, CA).
69. Repp BH (1984) Categorical perception: Issues, methods, findings. *Speech and Language* 10:243-335.
70. Majid A & Kruspe N (2018) Hunter-gatherer olfaction is special. *Current Biology* 28(3):409-413.
71. Large EW & Snyder JS (2009) Pulse and meter as neural resonance. *Annals of the New York Academy of Science* 1169:46-57.
72. Apicella C, Norenzayan A, & Henrich J (2020) Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior* 41(5):319-329.

73. Amir D & McAuliffe K (2020) Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior* 41(5):430-444.
74. Ravignani A, Thompson B, Grossi T, Delgado T, & Kirby S (2018) Evolving building blocks of rhythm: how human cognition creates music via cultural transmission. *Annals of the New York Academy of Science* 1423:176-187.
75. Boebinger D, Norman-Haignere SV, McDermott JH, & Kanwisher N (2021) Cortical music selectivity does not require musical training. *Journal of Neurophysiology* 125:2237-2263.
76. London J (2012) *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford University Press, Oxford).
77. Repp BH (2003) Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior* 35:355-370.
78. London J, Himberg T, & Cross I (2009) The effect of structural and performance factors in the perception of anacrases. *Music Perception* 27(2):103-120.
79. Tarentum Ao (4th c. BCE) *Elementa rhythmica*
80. Hippo Ao (387-391) *De musica*.
81. Descartes R (1618) *Compendium musicae*.
82. Euler L (1737) *Tentamen novae theoriae musicae, ex certissimis harmoniae principiis dilucide expositae* (St. Petersburg).
83. Parncutt R & Hair G (2018) Psychocultural theory of musical interval: Bye bye Pythagoras. *Music Perception* 35(4):475-501.
84. Serra X (2017) The computational study of a musical culture through its digital traces. *Acta Musicologica* 89(1):24-44.
85. Mehrabi N, Morstatter F, Saxena N, Lerman K, & Galstyan A (2019) A survey on bias and fairness in machine learning. *arXiv:1908.09635*.
86. McAuley JD, Jones MR, Holub S, Johnston HM, & Miller NS (2006) The time of our lives: lifespan development of timing and event tracking. *Journal of Experimental Psychology: General* 135:348-367.
87. Anglada-Tort M, Harrison PMC, & Jacoby N (2021) REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. *bioRxiv:2021.2001.2015.426897*.

88. Bridges D, Pitiot A, MacAskill MR, & Peirce JW (2020) The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ* 8:e9414.
89. Anwyl-Irvine A, Dalmaijer E, Hodges N, & Evershed J (2020) Online timing accuracy and precision: A comparison of platforms, browsers, and participant's devices. *PsyArXiv Preprints*.
90. Posch L, Bleier A, Flöck F, & Strohmaier M (2018) Characterizing the global crowd workforce: A cross-country comparison of crowdworker demographics. *arXiv:1812.05948*.
91. Riester J (1978) *Canción y producción en la vida de un pueblo indígena: los Chimane del oriente Boliviano* (Los Amigos del Libro, La Paz).
92. Huanca T (2008) *Tsimane' oral tradition, landscape, and identity in tropical forest* (Wa-Gui, La Paz, Bolivia).
93. Lima IMdF (2005) *Maracatus-Nação: Ressignificando Velhas Histórias* (Edições Bagaço, Recife).
94. Tinhorão JR (2012) *Os sons dos negros no Brasil. Cantos, danças, folguedos: origens. 3rd edition* (Editora 34, São Paulo).
95. Ferreira L (2007) An Afrocentric approach to musical performance in the black south Atlantic: The candombe drumming in Uruguay. *TRANS-Transcultural Music Review* 11.
96. Rice T (1994) (University of Chicago Press, Chicago).
97. Buchanan DA (2006) *Performing Democracy: Bulgarian Music and Musicians in Transition* (University of Chicago Press, Chicago).
98. Polak R (2000) A musical instrument travels around the world: Jenbe playing in Bamako, West Africa, and beyond. *The World of Music* 42(3):7-46.
99. Hood M (1960) The challenge of “bi-musicality”. *Ethnomusicology* 4(2):55-59.
100. Ingram C (2012) Tradition and divergence in Southwestern China: Kam big song singing in the village and on the stage. *The Asia Pacific Journal of Anthropology* 13(5):434-453.
101. Garner WR (1974) *The Processing of Information and Structure* (Lawrence Erlbaum).
102. Fraisse P (1982) Rhythm and tempo. *The Psychology of Music* 1:149-180.
103. Povel DJ & Essens P (1985) Perception of temporal patterns. *Music Perception* 2(4):411-440.

104. Kolinski M (1973) A cross-cultural approach to metro-rhythmic patterns. *Ethnomusicology* 17:494-506.
105. Locke D (1982) Principles of offbeat timing and cross-rhythm in southern Ewe dance drumming. *Ethnomusicology* 26:217-246.
106. Stobart H & Cross I (2000) The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosí, Bolivia. *British Journal of Ethnomusicology* 9(63-94).
107. London J, Polak R, & Jacoby N (2017) Rhythm histograms and musical meter: A corpus study of Malian percussion music. *Psychonomic Bulletin & Review* 24:474-480.
108. Lerdahl F & Jackendoff RS (1983) *A generative theory of tonal music* (MIT Press, Cambridge, Mass.) pp xiv, 368.
109. Cooper G & Meyer LB (1963) *The rhythmic structure of music* (University of Chicago Press, Chicago).
110. Toiviainen P & Eerola T (2003) Where is the beat? Comparison of Finnish and South African listeners. in *The 5th Triennial ESCOM Conference*, eds Kopiez R, Lehmann AC, I.Wolther, & C.Wolf (Hanover, Germany).
111. Friedman J, Hastie T, & Tibshirani R (2009) *The Elements of Statistical Learning* (Springer, New York).
112. Yates CM, Justus T, Atalay NB, Mert N, & Trehub SE (2017) Effects of musical training and culture on meter perception. *Psychology of Music* 45(2):231-245.
113. London J, Burger B, Thompson M, & Toiviainen P (2016) Speed on the dance floor: Auditory and visual cues for musical tempo. *Acta Psychologica* 164:70-80.
114. Levitin DJ, Grahn JA, & London J (2018) The psychology of music: Rhythm and movement. *Annual Review of Psychology* 69:51-75.
115. Arom S (2004) L'aksak: Principes et typologie. *Cahiers de Musiques Traditionnelles* 17:11-48.
116. Dzhidzhev T (1981) *Problemi na metroritūma i strukturata na pesenniya folklor* (Bulgarskata Akademiya na Naukite, Sofia, Bulgaria).
117. Polak R (2010) Rhythmic feel as meter: Non-isochronous beat subdivision in jembe music from Mali. *Music Theory Online* 16(4).
118. Corcoran C & Frieler K (2021) Playing it straight. *Music Perception* 38(4):372-385.

119. Kirby S, Cornish H, & Smith K (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31):10681-10686.
120. Xu J & Griffiths TL (2010) A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology* 60(2):107-126.