**Article**

# Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction

## Highlights

- Perceptual priors on rhythm are revealed by iterated reproduction of random sequences

- Priors in US listeners show peaks at integer ratio rhythms prevalent in Western music

- Priors in native Amazonians are distinct but also favor some integer ratio rhythms

- Musicians have similar priors as non-musicians, implicating passive exposure to music

## Authors

Nori Jacoby, Josh H. McDermott

## Correspondence

kj2413@columbia.edu

## In Brief

Jacoby and McDermott introduce a method to estimate priors on rhythm in human listeners. Both US participants and members of an Amazonian society exhibited priors that give high probability to rhythms whose time intervals form integer ratios. However, the priors otherwise differed across cultures, suggesting strong influences of musical experience.

CellPress

# Article

CellPress

# Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction

Nori Jacoby[1,2,3,*] and Josh H. McDermott[1]
[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]The Center for Science and Society, Columbia University, New York, NY 10027, USA
[3]Lead Contact
*Correspondence: kj2413@columbia.edu
http://dx.doi.org/10.1016/j.cub.2016.12.031

## SUMMARY

**Probability distributions over external states (priors) are essential to the interpretation of sensory signals. Priors for cultural artifacts such as music and language remain largely uncharacterized, but likely constrain cultural transmission, because only those signals with high probability under the prior can be reliably reproduced and communicated. We developed a method to estimate priors for simple rhythms via iterated reproduction of random temporal sequences. Listeners were asked to reproduce random "seed" rhythms; their reproductions were fed back as the stimulus and over time became dominated by internal biases, such that the prior could be estimated by applying the procedure multiple times. We validated that the measured prior was consistent across the modality of reproduction and that it correctly predicted perceptual discrimination. We then measured listeners' priors over the entire space of two- and three-interval rhythms. Priors in US participants showed peaks at rhythms with simple integer ratios and were similar for musicians and non-musicians. An analogous procedure produced qualitatively different results for spoken phrases, indicating some specificity to music. Priors measured in members of a native Amazonian society were distinct from those in US participants but also featured integer ratio peaks. The results do not preclude biological constraints favoring integer ratios, but they suggest that priors on musical rhythm are substantially modulated by experience and may simply reflect the empirical distribution of rhythm that listeners encounter. The proposed method can efficiently map out a high-resolution view of biases that shape transmission and stability of simple reproducible patterns within a culture.**

## INTRODUCTION

Perception operates on noisy data. Optimal inference thus requires integrating incoming sensory information with prior knowledge. In many areas of perception and cognition, humans appear to combine current observations with internal beliefs about the environment (the prior) in a process approximating statistical inference [1]. Priors have now been characterized in vision [2–6], time perception [7], audition [8], and motor control [9]. The effect of a prior on an observer's estimate of the world is usually to bias the estimate toward states that have high probability under the prior [4]. As a result, priors for domains in which signals are reproduced by humans, such as music and speech, directly constrain cultural transmission: only signals with high probability under the prior are reliably reproduced and communicated [10]. However, priors for cultural artifacts have only recently begun to be characterized for language [6] and speech [11] and are unexplored for music.

Priors on music are presumably essential for extracting musical structure from sound [12]. Because of their likely role in cultural transmission, it also seemed plausible that music priors could provide insight into the origins of salient characteristics of music. Although music is highly variable across societies, it is usually defined in part by the structured arrangement of events in time to form rhythm. Particular types of rhythms—those comprised of time intervals related by ratios of small integers ("simple" integer ratios)—are believed to be common, if not universal [13]. What underlies their prevalence? One possibility is that integer ratio rhythms are more easily produced given our motor constraints [14]. Alternatively, representational constraints could affect what can be accurately perceived [15, 16]. To explore this latter possibility, we examined whether integer ratios are reflected in the prior that constrains human perception and transmission of rhythms and whether this prior varies with musical experience and cultural exposure.

Previous probabilistic models of musical structure and/or perception assume priors [17, 18], but to our knowledge priors on musical structure have never been measured, in part because of methodological limitations. Evidence for perceptual priors is traditionally obtained via biases in discrimination or categorization judgments [4]. For music, such tasks are typically difficult, and most results have been obtained with expert musicians [18–20]. Moreover, discrimination experiments are limited in dimensionality because they require large numbers of judgments to establish discriminability at each point in a stimulus space. Rhythm discrimination experiments are thus typically designed to test particular hypotheses derived from Western music theory and cannot probe the large space of possible priors, particularly when considering non-Western cultures for which scientists may not have hypotheses. Experiments in which participants tap to a beat reveal biases that may reflect perception [16, 21, 22] but

have not explicitly characterized a prior. Other approaches to studying rhythm representation require judgments using Western musical notation [18, 20] and are also unsuitable for non-experts or non-Western listeners.

We developed a method to measure perceptual priors on rhythm that is efficient and implicit. Our paradigm is an extension of the game of "broken telephone" previously used to reveal priors in other domains [6, 10, 23]. Listeners are presented with a random "seed" sequence (clicks or other sounds separated by random time intervals) and reproduce it by tapping. In practice, the reproduction is usually performed synchronous with the seed while it is repeatedly presented (because this is easiest for participants), but it can also be performed after seed presentation. The listener's reproduction is then substituted for the seed, and the process is iterated. If the listener's response depends on the Bayesian integration of the currently heard rhythm with a prior, and if production biases are small, this process converges to a sample from the prior for the domain, and the prior can be estimated by applying the procedure multiple times [24].

Our method is hypothesis-neutral (allowing any possible pattern of results to be detected), can be applied irrespective of the participant's musical or cultural background, and estimates the prior over the entire space of simple rhythms. Although related methods are sometimes used to simulate cultural evolution [25, 26], here our goal is strictly to measure the prior within groups of listeners with a common cultural background, in whom the prior might promote cultural stability.

## RESULTS

Our paper is divided into three parts. First, we show the results of applying our method to simple rhythms, to illustrate the approach. Second, we present a series of control experiments to substantiate the claim that the results reflect a perceptual prior, showing that they do not reflect productive biases, that synchronization and replication from memory yield similar results, and that the measured distribution predicts discriminability. Third, we use the method to test the effect of culture and musical experience on the perceptual prior for rhythm.

### Experiment 1: Priors on Rhythm via Iterated Reproduction

Three-interval rhythms are those formed by a repeating sequence of three time intervals. The intervals can be described in terms of their ratios, which for Western musical rhythms can be approximated with integers (Figure 1A) but which in general could be arbitrary non-integer proportions (Figure 1B). The three intervals are produced by four events (e.g., clicks). When the rhythm is presented repeatedly, the fourth event forms the end of the last interval and the beginning of the first interval (Figure 1C).

Western musical notation can only be used to describe rhythms with simple integer ratios (Figure 1D), and so to describe rhythm more generally we will use the ratios between the intervals, plotted in a "rhythm space" [20]. Three-interval rhythms can be represented in three dimensions corresponding to the three intervals (Figure 1E). In this paper, we focus on rhythms whose overall duration (the sum of the intervals) is fixed (typically 2,000 ms, but 1,000 ms for experiments 3, S1, and S2). Because

the overall duration is fixed, the intervals can be presented on a two-dimensional triangular simplex (Figure 1F) known as a "chronotopological map" [20, 27]; for simplicity, here we refer to it as the "rhythm simplex." Every three-interval rhythm with a fixed duration can be uniquely described by a point on the simplex.

In order to avoid presenting rhythms with intervals too short to reproduce [28], we constrained intervals in the seeds to a smaller simplex (inner triangle in Figure 1E) such that all intervals were larger than 15% of the total duration. All results figures display responses on this inner simplex.

Each trial started with a random seed sampled uniformly from the rhythm simplex (Figure 1F). Participants were presented with ten repetitions of the seed and instructed to synchronize their taps to the repeated rhythm. The inter-tap intervals were averaged across the ten repetitions, and the resulting mean reproduction was then substituted for the seed (Figure 1G). The process was iterated five times.

Figures 2A and 2B show aggregated responses from all participants (n = 26; 20–25 seeds per individual; 609 total distinct seeds; participants were a mix of musicians and non-musicians) shown as points on the rhythm simplex for each iteration. The distribution of responses evolves over time and appears to converge to a multi-modal distribution. Copying accuracy (the distance between stimulus and reproduction) improved across iterations (mean stimulus-reproduction distance for iterations 1–5 was 61, 45, 38, 34, and 32 ms, respectively) but was not significantly different for the last two iterations (t(25) = −1.85 p = 0.08), suggesting that convergence is reached or nearly reached by the fifth iteration. To estimate the continuous distribution underlying the responses, we applied kernel density estimation to each iteration's data (shown in Figure 2C for the fifth and final iteration).

As expected, the initial distribution of seeds was not significantly different from uniform (measured by the Jensen-Shannon divergence; JSD = 0.039, p = 0.82). However, the distributions in all subsequent iterations differed significantly from uniform (p < 0.001 in all cases). To clarify the structure of the final distribution, we superimposed symbols (crosses) at rhythms whose intervals are related by simple integer ratios (Figure 2C). We plot all 22 simple ratios ($i : j : k$) where $1 \leq i, j, k \leq 3$.

The modes of the distribution obtained in the last iteration closely correspond to integer ratio points, though not all integer ratio rhythms correspond to modes. We defined an "integer score" as the distance (JSD) between the data distribution and a uniform distribution over the 22 integer ratio points. To evaluate statistical significance, we compared the score to that of random distributions containing the same number of modes, revealing a significant tendency toward probability mass at integer ratios (p < 0.001). To assess how closely the modes were centered on integer ratios, we fit a 22-component Gaussian mixture model to the data from the last iteration and measured whether the component means aligned with integer ratios. Two of the modes were significantly biased away from integer ratios (those corresponding to 2:3:1, p < 0.001; those corresponding to 3:2:1, p = 0.002), consistent with previous evidence for human rhythm categories centered close to, but not exactly on, integers [22].

We performed a similar experiment for two-interval rhythms, the advantage being that the lower dimensionality permitted
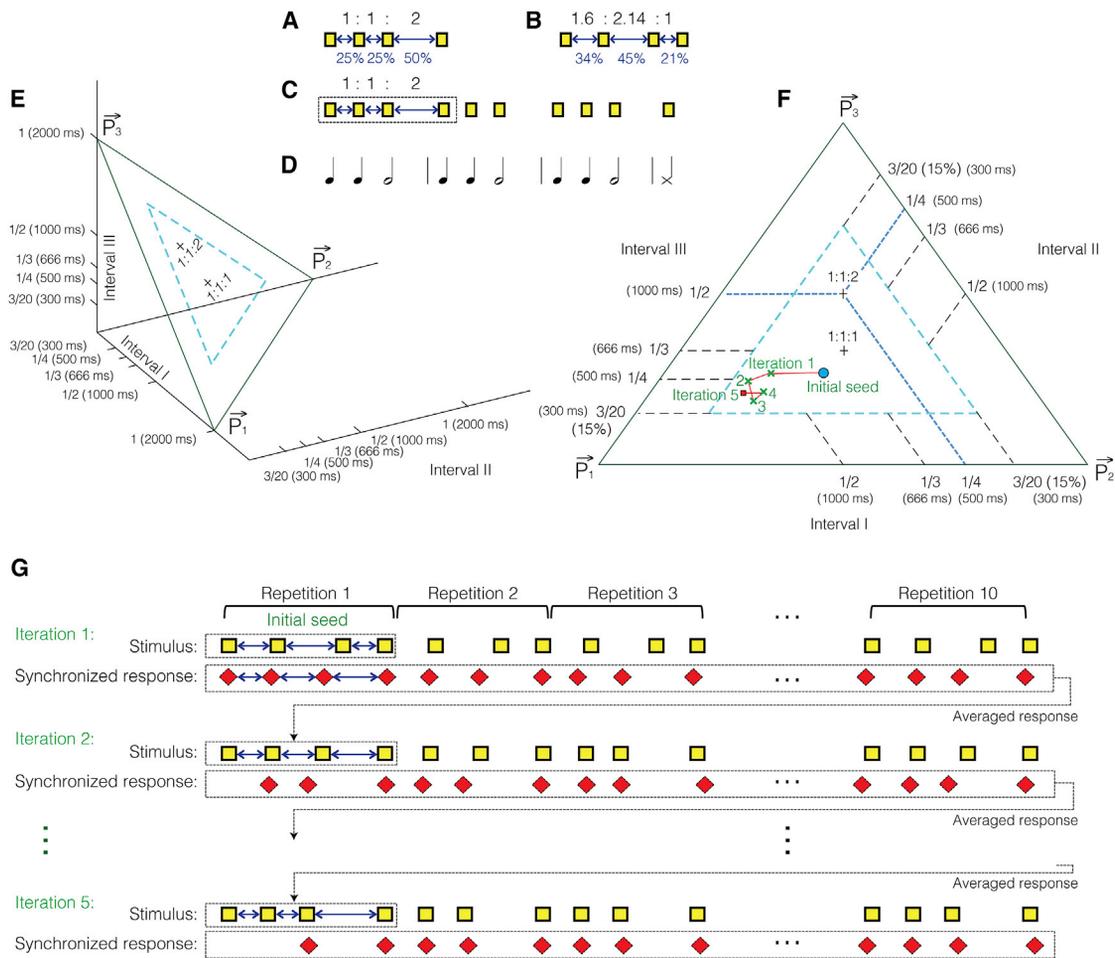
**Figure 1. Illustration of Stimuli and Experimental Paradigm**

(A) One repetition of a three-interval rhythm (four clicks) with integer ratio proportions (1:1:2), resulting in a rhythm with intervals of 25%, 25%, and 50% of the total pattern duration.

(B) One repetition of a three-interval rhythm with non-integer ratio proportions (1.6:2.14:1).

(C) Three repetitions of the 1:1:2 three-interval rhythm (ten clicks).

(D) The repeated pattern 1:1:2 written in Western musical notation.

(E) As first described in [20], three-interval rhythms can be presented as a point in a three-dimensional space, where the x, y, and z coordinates correspond to the first, second, and third interval, respectively. The points with an overall fixed duration (2,000 ms in this case) form a triangle (simplex) that can be projected to a two-dimensional plane, creating a "rhythm simplex," also known as a "chronotopological map," or "rhythm chart" [20, 27]. Here and in (F), the dashed triangle denotes the region from which stimuli were drawn (defined by inter-onset intervals that are larger than 15% of the total duration, e.g., 300 ms for the 2,000 ms duration used in most experiments), with two example integer ratio rhythms indicated by crosses.

(F) The rhythm simplex. The red line plots an example stimulus/response trajectory across the five iterations of an example trial.

(G) Schematic of trial structure.

ten iterations per seed, which we used to further examine convergence (Figure S1). As with three-interval rhythms, reproductions converged to a distribution whose modes were approximately simple integer ratios (1:1, 1:2, 1:3…), and copying accuracy converged by iteration 5 (see Figure S2 for convergence behavior for all experiments). It thus seems that the iterative process for simple rhythms converges after about five iterations (see also Figures S7C and S7D).

## Controlling for Productive Biases

Our method has the advantage of being efficient, comprehensive, and hypothesis neutral, but it relies on a complex task involving perception and synchronized reproduction, and the re-

sults could in principle reflect a combination of perceptual and productive constraints. To interpret the results in terms of a perceptual prior, we must ensure that production biases do not substantially contribute. We thus conducted a series of control experiments. To test whether the results depended on production modality, we asked participants to produce a vocal syllable rather than tap (experiment 2). To test whether the results depended on entrainment, we asked participants to replicate a pattern from memory rather than tap in synchrony (experiment 3). To test the role of successive iterations within participants, we temporally intermixed reproductions from different seeds and mixed iterative chains across participants (experiment 4). To verify that the predicted prior would be evident without
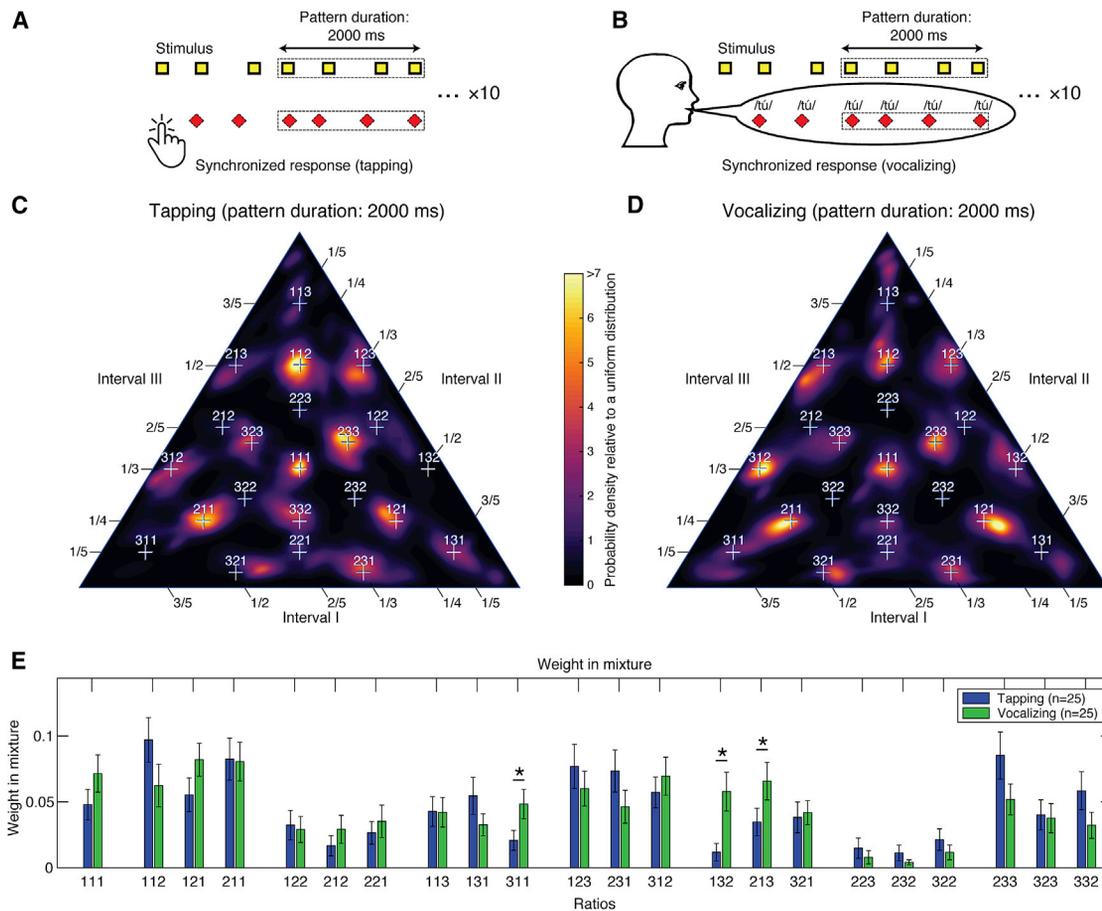
**Figure 2. Results of Experiment 1: Iterated Reproduction by Synchronous Tapping**
(A) All seeds of experiment 1 displayed on the rhythm simplex.
(B) All responses from each of the five iterations of experiment 1.
(C) Kernel density estimate of the continuous distribution underlying the data from iteration 5 of experiment 1. Crosses plot simple integer ratio rhythms.
See also Figures S1, S2, and S4.

reproduction demands, we employed a standard discrimination task (experiment 5) and compared the results to a Bayesian ideal observer using the distribution obtained in experiment 1. Finally, to test whether integer ratios would result from reproducing any type of temporal pattern, we asked listeners to repeat the syllables of meaningful English phrases (experiment 6).

**Experiment 2: Tapping versus Vocalizing**
In separate, counterbalanced blocks, participants reproduced sequences of clicks either by tapping as in experiment 1 (Fig-

ure 3A) or by vocalizing the syllable /tú/ (Figure 3B). The motivation was that production biases for speech and finger tapping might be expected to be different—speech rhythms are often produced aperiodically [29], whereas finger tapping is often periodic due to finger movement kinematics [14]. One might thus expect the results to differ with reproduction modality if production biases play a large role.

Figures 3C and 3D show that tapping and vocalizing produce qualitatively similar distributions, with prominent integer ratio peaks in both cases (both integer scores were significant,

**Figure 3. Results of Experiment 2: Tapping versus Vocalizing**

(A) Schematic of the tapping task.

(B) Schematic of the vocalization task.

(C) Response distribution from the fifth iteration of the tapping task.

(D) Response distribution from the fifth iteration of the vocalization task.

(E) Weights of Gaussian mixture components assigned to each integer ratio rhythm in the two distributions (see also Figure S7). Error bars plot confidence intervals on the weights (SDs of the weight distributions derived from bootstrapping), with significance here and elsewhere evaluated after Bonferroni correction for multiple comparisons.

$p < 0.001$). To compare the detailed form of the distributions, we fit a Gaussian mixture model to the final (fifth) iteration for each condition, constraining the component means to the vicinity of a corresponding integer ratio point (Figures S7E–S7H demonstrate the resulting quality of fit). The weights of the mixture components for the two conditions were significantly correlated (Figure 3E; r = 0.58, p = 0.005, r = 0.83 after correcting for test-retest reliability of the weights), and only three of the 22 integer ratio points showed a significant difference between conditions (3:1:1, 1:3:2, and 2:1:3; p = 0.01, 0.0002, and 0.04, respectively). These small differences were reflected in a significant JSD (JSD = 0.082, p = 0.001). Overall, however, the results indicate that reproduction modality has little effect on the obtained distributions.

## Experiment 3: Replication from Memory versus Synchronization

To test whether the results were independent of the requirement to synchronize the reproduction with the heard signal,

we asked participants to listen to a rhythm and replicate it from memory (Figures 4A and 4B). To reduce working memory load, we reduced the rhythm duration to 1 s, doubling the tempo from that in experiments 1 and 2. Participants completed a separate block in which they tapped synchronously, as in experiment 1.

We observed a slightly different distribution compared to experiment 1, presumably due to the shorter duration (see Figure S3 for a separate experiment manipulating tempo), but both synchronization and replication showed significant integer scores (p < 0.001). Moreover, integer ratio weights (Figure 4E) for the two conditions were correlated (r = 0.68, p = 0.001, r = 0.79 after correcting for test-retest reliability of the weights). The differences between distributions could reflect the number of stimulus repetitions per iteration (ten for synchronization; three for replication, to reduce memory load). These results are consistent with findings of similar tapping biases for synchronization and replication from memory [21] and are consistent with the idea that our results reflect a perceptual prior.
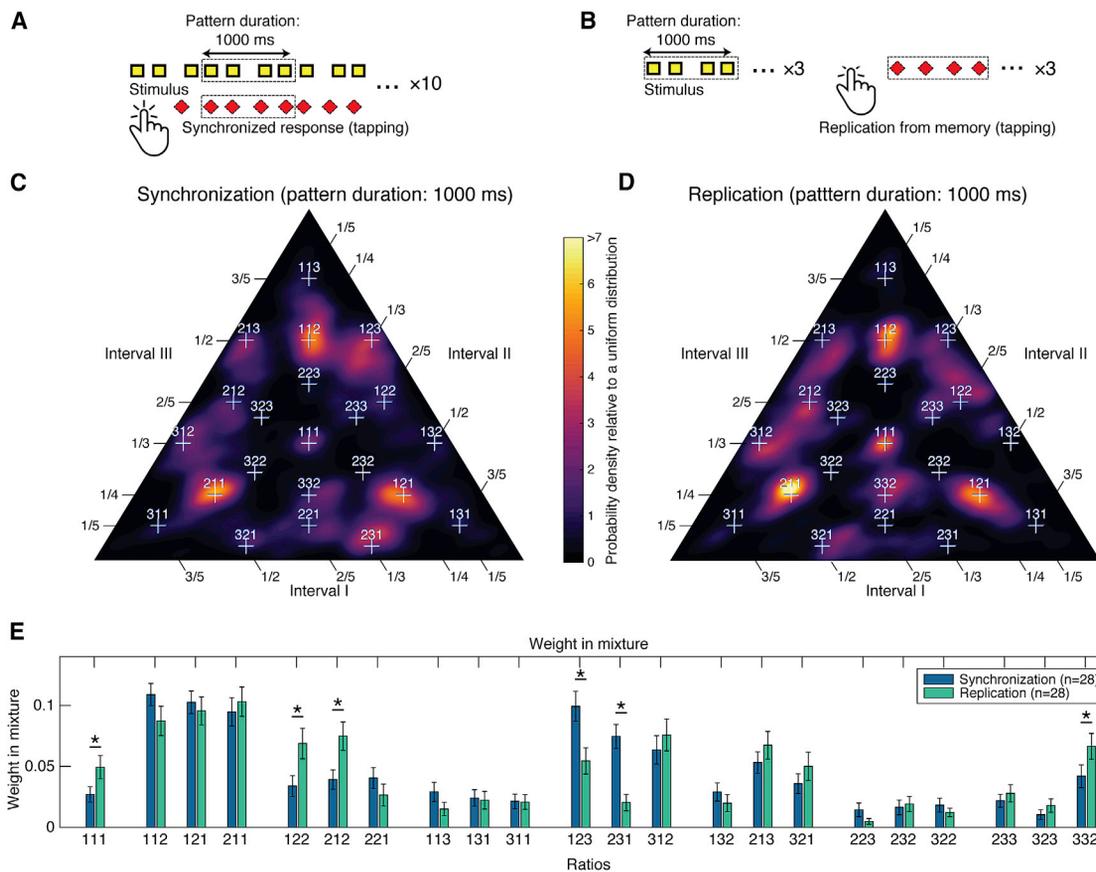
**Figure 4. Results of Experiment 3: Synchronization and Replication**
(A) Schematic of the synchronization task.
(B) Schematic of the replication task.
(C) Response distribution from the fifth iteration of the synchronization task.
(D) Response distribution from the fifth iteration of the replication task.
(E) Weights of Gaussian mixture components assigned to each integer ratio rhythm in the two distributions. Error bars plot confidence intervals on the weights (SDs of the weight distributions derived from bootstrapping).
See also Figure S3.

## Experiment 4: Eliminating Dependencies across Iterations

In order for iterated reproduction to converge to the prior, the reproductions in principle must be influenced only by the prior and the stimulus (the seed, or the previous iteration's reproduction), and not by previous history [30]. Our main paradigm could violate this assumption of Markovian behavior because the same participants complete all iterations for a particular seed in succession, and reproductions could be affected by the memory of previous iterations. To test the effect of such factors, we repeated experiment 1 with two alternative procedures. In the first condition ("random order within participants"), participants again completed five iterations per seed, but the iterations for different seeds were intermingled. In the second condition ("across participants"), each participant completed one iteration for a large number of seeds, and the reproductions of one participant formed the seed for the next participant (Figure S4).

The results of both conditions were similar and did not differ substantially from those of experiment 1 (Figure S4). Integer ratio weights were significantly correlated with those of experiment 1

in both conditions (random order within participant: r = 0.83, p < 0.001, r = 0.96 after correcting for reliability; across participants: r = 0.76, p < 0.001, r = 0.85 after correction), and in only two cases were individual weights significantly different. It thus appears that any effects of across-iteration memory are modest and that our basic paradigm is reasonable to employ in practice (the advantage being that it is logistically simpler to perform the experiment separately for each subject). These results also provide evidence that the priors being measured are similar across participants—if there was substantial variation, the two procedures would yield different results [31].

## Experiment 5: Rhythm Discrimination

To further test whether the results reflect perceptual priors, we used a discrimination task [19, 32] as is traditionally employed to probe priors. The disadvantage of this approach is that it is in practice not feasible to explore the entire space of rhythms. However, a classical discrimination task removes all production constraints and provides a useful consistency check. We measured discrimination of stimuli drawn from a
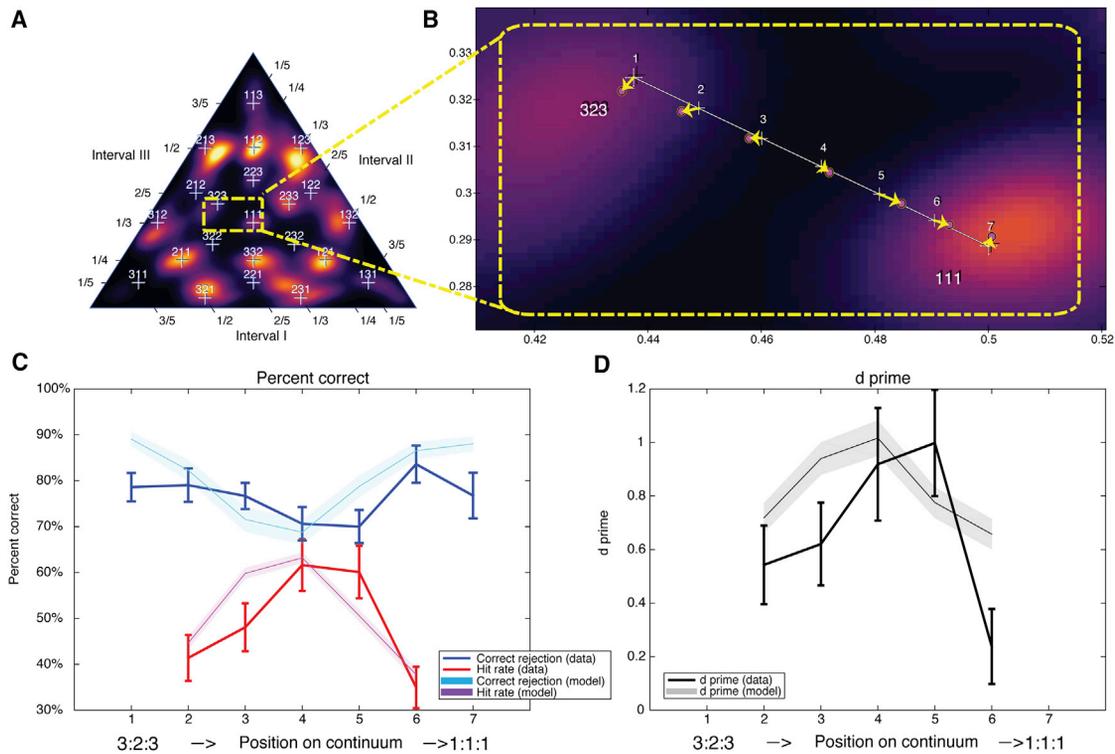
**Figure 5. Results of Experiment 5: Rhythm Discrimination, with Predictions from the Measured Prior**

(A) The response distribution from experiment S2 (with a pattern duration of 1,000 ms).

(B) Close-up view of a slice through rhythm space occupied by experimental stimuli. Stimuli are depicted with white crosses, and perceptual biases predicted by the prior are marked as magenta diamonds; yellow arrows show direction of bias. Stimuli are "pulled" toward the modes, such that estimates of points straddling the trough are further apart than those of points on one side of the trough, improving discrimination.

(C) Hit and correct rejection rates for human listeners and for an ideal observer using the measured prior. Error bars plot the SEM.

(D) Perceptual sensitivity for each position along the continuum, along with the predicted sensitivity of the ideal observer from (C). The value plotted for position $k$ is the discriminability of the rhythms at positions $k - 1$ and $k + 1$. Values are thus available only for positions 2 through 6. Error bars plot the SEM.

See also Figure S7.

one-dimensional slice through the three-interval rhythm simplex. If our results indeed reflect a prior, discrimination should be best at the troughs of the distribution and worst at the modes, as in classic categorical perception (because under Bayesian inference, estimates of stimuli are typically "pulled" toward the closest mode of the prior, rendering pairs of stimuli close to a mode more similar than pairs of stimuli between modes [11]).

Participants discriminated pairs of three-interval rhythms straddling positions along a continuum between 1:1:1 and 3:2:3 (Figures 5A and 5B). Consistent with previous discrimination results [19, 32], and as predicted by a Bayesian interpretation of our iterated reproduction results, discriminability increased between the modes of the distribution measured in experiment 1 (Figures 5C and 5D; F(4,18) = 10.77, p < 0.001). A planned comparison between the average discriminability of stimuli closest to integer ratio rhythms (points 2 and 6 in Figure 5D) and of stimuli in between (points 3–5) showed better discrimination for the former, as predicted (t(18) = 6.27, p < 0.001). Moreover, the discrimination performance of a Bayesian ideal observer using the prior obtained in experiment 1 qualitatively replicated human performance (Figures 5C and 5D). These results provide further evidence that our method produces the perceptual prior for rhythm.

## Experiment 6: Tapping and Speaking

To test whether integer ratio biases extend to the perception of all sound signals, we asked participants to repeat spoken utterances in an analogous iterated reproduction paradigm (see also [33]). The syllables of five four-syllable phrases were segmented from recordings and substituted for the clicks used in our main paradigm (i.e., presented with initially random time intervals between syllables); participants were asked to repeat them back (Figure 6A). The onsets of the replicated syllables were detected and substituted for the initial onsets times as in experiments 1–4 (Figures 6C and 6D). We chose phrases with distinctive rhythms and short syllables with salient onsets. To facilitate syllable pronunciation without overlap or coarticulation, we presented phrases at a relatively slow tempo. Participants also completed the analogous task with clicks and taps (Figure 6B) in a separate block.

In contrast to click rhythms, speech rhythms converged to distributions whose peaks did not overlap with integer points. None of the five spoken patterns showed a significant integer score (p = 0.07, 0.1, 0.1, 0.85, and 0.14 for phrases 1–5, respectively). In contrast, the same participants produced significant integer scores when asked to replicate click rhythms by tapping (Figure 6D; p < 0.001 for all five equal-sized subsets of tapping data, analyzed in subsets to equate in power to individual speech
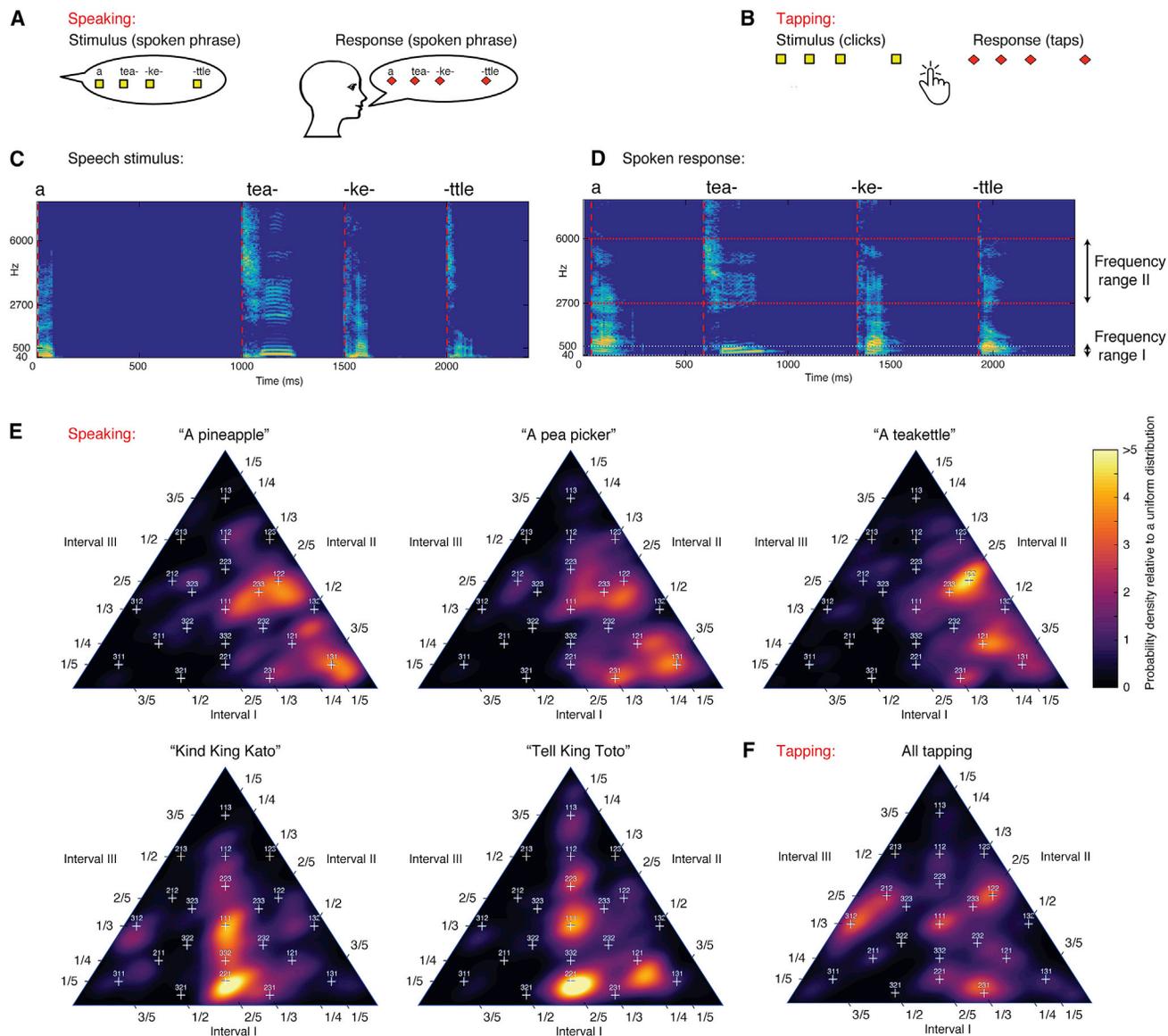
**Figure 6. Results of Experiment 6: Speaking versus Tapping**

(A) Schematic of the speaking task. Listeners heard a spoken phrase, segmented into syllables that were assigned random onset times, and repeated it.

(B) Schematic of the tapping task.

(C) Spectrogram of an example speaking stimulus. Silences of variable durations were inserted between syllables such that the intervals between voicing onsets (dashed red lines) had the desired intervals—in this example, $r_1 = 1000$, $r_2 = 500$, and $r_3 = 500$ ms (for illustrative purposes).

(D) Spectrogram of example recorded response. Dotted horizontal lines mark the two ranges used for the onset detection process (range I, white: 40–500 Hz; range II, red: 2700–6000 Hz). Vertical dashed lines mark the detected onsets.

(E) Response distribution from the fifth iteration of the speaking task, for each of five phrases.

(F) Response distribution from the fifth iteration of the tapping task.

See also Figures S5 and S6.

conditions). These spoken phrases were a lone exception among our experimental conditions; all others produced statistically significant integer scores (Figure S6). The results indicate that the primacy of integer ratio rhythms is at least somewhat specific to music (consistent with well-known differences between speech and music rhythm [29, 34]) and underscore that integer ratios do not always emerge from iterated reproduction.

The results for both tasks are asymmetric; this is most likely because the stimulus was only presented once per iteration (all

other experiments used cyclically repeated stimuli). This effect is consistent with previous findings that tapping biases are altered by cyclic repetition [21]. Because cyclic permutations of patterns (e.g., 1:1:2 and 1:2:1) produce stimulus sequences that are identical but for the beginning and end, repetition could cause these patterns to be similarly perceived, reducing the differences in the measured prior for distinct cyclic permutations. However, we found that adding an accent to the first click of the repeating stimulus (to remove this ambiguity) had little effect
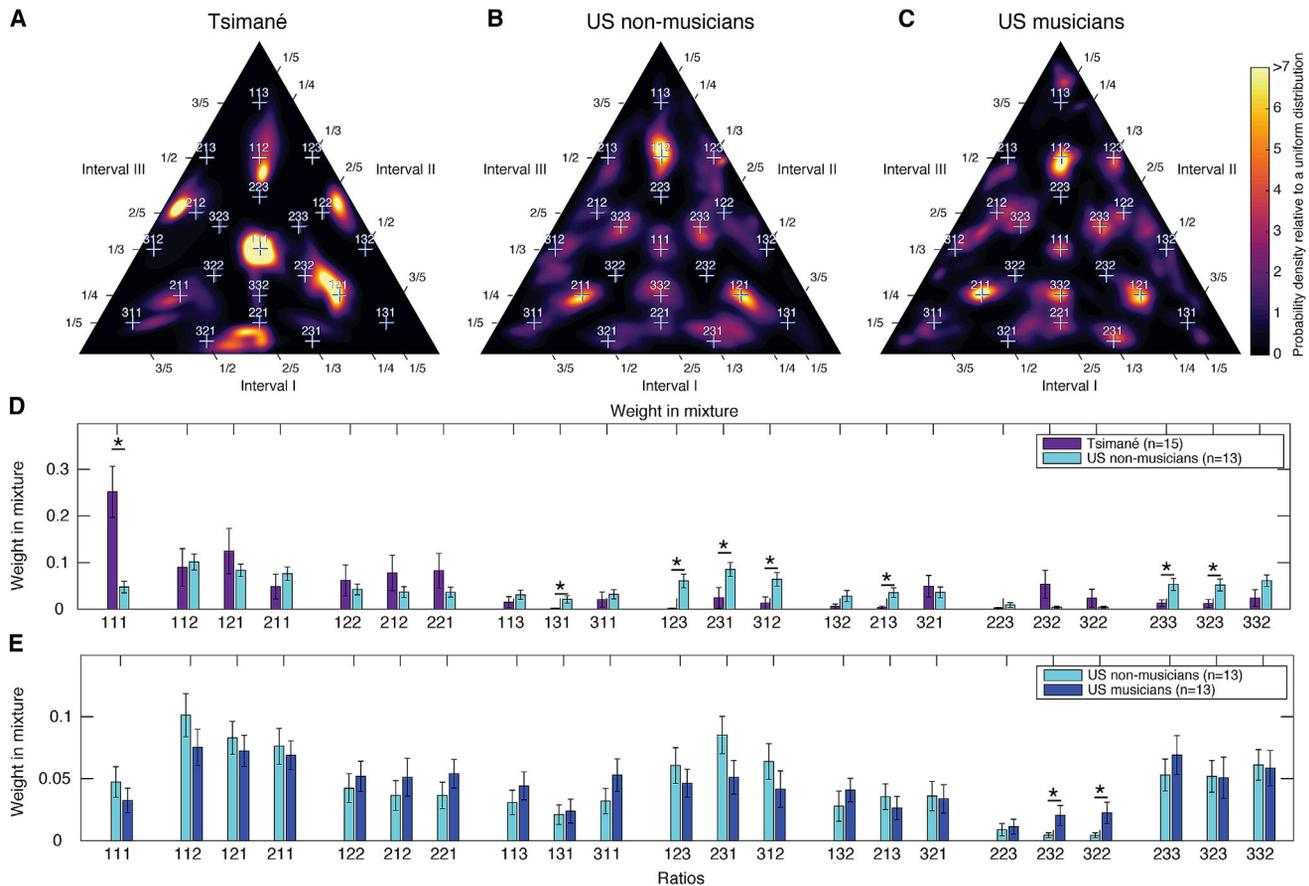
**Figure 7. Priors in Amazonians and US Participants with and without Musical Training**

(A–C) Response distribution from fifth iteration of tapping task for Tsimané participants (A), US non-musicians (B), and US musicians (C).

(D and E) Weights of Gaussian mixture components assigned to each integer ratio rhythm for Tsimané and age-matched US participants with little musical exposure (D) and US musicians and non-musicians (E). Error bars plot confidence intervals on the weights (SDs of the weight distributions derived from bootstrapping).

on the results (Figure S5), suggesting that cyclic ambiguity is not the main reason for the symmetry of many repetitions compared to one. Instead, it seems that repetition itself may change the prior probability of rhythmic patterns, consistent with its importance in music [35].

**Experiment 7: Priors in Native Amazonians**

To explore the possible dependence of rhythmic biases on exposure to music, we tested members of the Tsimané, an indigenous society of farmer-foragers residing in remote areas of the Bolivian Amazon basin. The Tsimané remain relatively isolated from Western music and exhibit dramatically different harmony preferences than US participants [36], raising the possibility that they might exhibit distinct rhythmic priors as well. None of the Tsimané participants regularly played a musical instrument. We therefore compared them to an age-matched group of US non-musicians. Despite their profoundly different cultural exposure, the Tsimané also displayed a significant integer score (p = 0.005). However, the prior derived from their reproductions differed substantially from that of US participants (Figures 7A and 7B; JSD = 0.23, p < 0.001). Although both cultures assigned high probability to some integer ratio rhythms (e.g., 1:1:2 and

1:2:2), eight of the 22 rhythms showed significant differences between cultures, and the integer ratio weights of the two populations were not significantly correlated (r = 0.19, p = 0.43; Figure 7D).

Several uninteresting explanations for the cross-cultural differences are unlikely given the constraints of our method and the results. It is not the case that the Tsimané simply cannot entrain to rhythms, because the method discards taps that do not closely correspond to stimulus events. It is also not the case that the differences reflect an inability of the Tsimané to reproduce short intervals, because three of the modes of their distribution lie near the boundary of the rhythm simplex, and as such they contain intervals close to the shortest allowable in our paradigm. And although Tsimané tapping was more variable than that of US participants, evident in the broader modes of their response distribution, it is visually apparent that there is "room" for additional modes on the rhythm simplex that nonetheless did not arise.

The Tsimané prior is consistent with previous ethnographic descriptions of their music ([37], pp. 345–368), which reported patterns of 1:1:1 and 1:1:2 (and their cyclic permutations). In contrast, patterns prominent in US participants but not in

**Cell**Press

Tsimané, e.g., 2:3:3 and 1:2:3, are not evident in descriptions of Tsimané music. Like that of the US participants, the Tsimané distribution showed significant deviations from integer ratios in some cases (1:2:2, p = 0.005; 2:1:2, p < 0.001).

**Effect of Musical Expertise**

The apparent effects of experience found in experiment 7 raise the question of whether priors primarily reflect passive experience with music. To explore this issue, we compared US musicians and non-musicians (Figures 7B and 7C). Although the two resulting distributions were significantly different (JSD = 0.08, p < 0.001), it is visually apparent that the distributions are qualitatively similar. Both groups exhibited a significant integer score (p < 0.001), and their integer ratio weights were correlated (r = 0.79, p < 0.001, r = 0.91 after correcting for reliability; Figure 7E). Out of the 22 simple integer ratio rhythms, only two were more prominent in musicians than non-musicians (2:3:2 and 3:2:2). These patterns occur in Balkan music but are not accurately perceived by Western adults and children [38]. The difference in weights for these rhythms could greater familiarity of musicians with music patterns from other cultures.

It is also apparent that the effect of culture was more prominent than the effect of musical training; the JSD between Tsimané and US non-musicians (JSD = 0.23) was larger than the JSD between US musicians and non-musicians (JSD = 0.08, p = 0.02). Overall, these results indicate that rhythm priors are not strongly affected by musical expertise, in US participants at least, but that culture has a significant role in shaping musical priors.

**DISCUSSION**

We have introduced a method to measure internal biases in rhythm reproduction and have presented evidence that the method reveals perceptual priors on rhythm that constrain what listeners can accurately perceive and reproduce. Our method permits the measurement of the prior over the entire space of simple rhythms, and as such it could reveal priors no matter what form they take. We found that priors on rhythm in US participants favored integer ratios, but only those that are prevalent in Western music. Priors in a native Amazonian society also exhibited peaks at integer ratios, but they were otherwise qualitatively different from priors in US participants, in ways that are consistent with what we know of their music. The results were similar for several different modes of reproduction but did not extend to the reproduction of spoken phrases, indicating that integer ratio priors are at least somewhat specific to music. In addition, the results were similar for musicians and non-musicians, suggesting that priors are shaped by passive exposure to music.

**Relation to Other Iterative Methods**

Our general methodology is a version of the serial reproduction paradigm first introduced by Bartlett [23] to demonstrate cultural biases in the perception of short stories, jokes, and pictures. Serial reproduction was subsequently used as a model for the evolution of language through cultural transmission ("iterated learning"; [10, 26]). Interest in iterated paradigms has grown recently due to their mathematical similarity to sampling techniques in modern machine learning [10]. Serial reproduction has since been used to study visual biases on color naming [6], the emergence of linguistic structure during transmission of language [26], biases in speech intonation and whistling [25, 33], and emergent structure in symbolic sequences [31] and non-verbal gestural cues [39]. These previous approaches either have used discrete language-mediated judgments or did not systematically map a perceptual space. To our knowledge, our method is the first application of serial reproduction to map an entire analog perceptual space.

**Previous Evidence for Rhythmic Priors**

Although previous research on musical rhythm has generally not made reference to a prior or to Bayesian inference (but see [18]), several previous findings are consistent with our results. Discrimination of simple rhythmic stimuli has been found to be better near boundaries between simple integer rhythms [19, 32], a finding interpreted as evidence for categorical perception. Our discrimination results are qualitatively similar to (and were inspired by) these prior findings, but they are predicted by the prior measured using iterative reproduction, suggesting that categorical perception can result from rhythm priors. Our contribution relative to this discrimination literature lies in providing a method that is more efficient in mapping perceptual biases (enabling high-resolution estimates of the prior). We quantified the efficiency benefit by estimating the number of discrimination trials required to estimate the three-interval prior with the fidelity obtained by our iterative method. We estimate that our iterative procedure recovers the prior 20–40 times faster than traditional discrimination experiments (see the Supplemental Experimental Procedures). Obtaining results equivalent to those of experiment 1 using discrimination would require more than 520 participants (assuming the same session duration).

Another approach to understanding mental representations of rhythm has been to ask trained musicians to categorize rhythms using musical notation. Desain and Honing [20] applied this approach to three-interval rhythms; the results were later used to infer the prior probability of particular integer ratios [18]. Our results are consistent with their findings—rhythms associated with musically notated categories were close to the modes found by our implicit paradigm. The benefit of our method is that it is independent of Western musical notation and can thus be used to test non-musicians or non-Western listeners.

Our results can be viewed as an extension of tapping reproduction experiments, in which reproductions of two- or three-interval rhythms are biased toward integer ratios [16, 21, 22]. However, because the production biases measured in a single reproduction are small, these experiments occasionally provided inconsistent results [22] and did not provide a way to reveal the prior. The first iteration of our basic paradigm is equivalent to such an experiment, but the reproduction biases are enhanced with iteration (Figures 2 and S2), enabling direct measurement of the prior. Our experiments also reveal that the underlying effects are largely independent of production constraints—tapping is merely a way to access a listener's representation of rhythm.

**Factors Influencing Rhythmic Priors**

Although we have argued that our method reveals a listener's prior over rhythms, the detailed form of the obtained distribution

CellPress

varied across several experiments. Task demands required across-experiment differences in tempo, as well as the number of stimulus repetitions per iteration. It seems plausible that both factors affect the prior probability of rhythms. Some patterns could be more common in particular tempo ranges, consistent with evidence that rhythm perception is modulated by tempo [28, 40]. And repetition could affect the tendency for patterns to be interpreted as musical [35, 41], which might amplify the bias for integer ratio rhythms.

### Origins and Functions of Priors on Rhythm

One reason to be interested in priors on rhythm (apart from the role they likely play in extracting musical structure from sound, particularly from multi-instrument pieces [12]) is that they constrain what can be accurately perceived and transmitted. As such, if priors were themselves tightly constrained, these constraints could in principle explain universal features of rhythm. Previous proposals for such biological and/or cognitive constraints have included oscillatory circuits [15, 42, 43], recurrent attractor networks [44], and rhythm representations built on an isochronous pulse [16], all of which should bias perception toward integer ratio rhythms. Alternatively (or additionally), the widespread presence of integer ratio rhythms could reflect constraints on production [14, 45].

Our results do not resolve these issues, but the differences between cultures that we observed indicate that rhythm priors are not fully fixed by biological constraints. The results leave open the possibility that priors themselves are largely free to reflect whatever distribution occurs in the world [5], with cross-cultural similarities instead being driven by productive or other constraints on human musical systems. One informative avenue will be to document the prior over development, perhaps by applying our methods to children.

Further insights into these issues could also be gained by studying additional cultures, particularly those whose musical rhythms deviate substantially from integer ratio proportions [46–48]. Our paradigm is ideal for this purpose. Corpus analysis is no substitute: notated music does not exist for many non-Western cultures, and when available, notations usually reflect what a Westerner hears when listening to the other culture's music. Moreover, it remains possible that priors could deviate from the empirical distribution of a culture's music. Our paradigm instead directly probes the mental representation of the listener. The results here provide another example of the perceptual diversity that can be revealed by experiments in non-Western cultures [36, 49, 50].

### EXPERIMENTAL PROCEDURES

See the Supplemental Experimental Procedures for a full description.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.12.031.

### AUTHOR CONTRIBUTIONS

N.J and J.H.M designed the experiments and wrote the paper. N.J collected the data.

### REFERENCES

1. Knill, D.C., and Richards, W., eds. (1996). Perception as Bayesian Inference (Cambridge University Press).

2. Adams, W.J., Graf, E.W., and Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. Nat. Neurosci. 7, 1057–1058.

3. Weiss, Y., Simoncelli, E.P., and Adelson, E.H. (2002). Motion illusions as optimal percepts. Nat. Neurosci. 5, 598–604.

4. Stocker, A.A., and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. 9, 578–585.

5. Girshick, A.R., Landy, M.S., and Simoncelli, E.P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat. Neurosci. 14, 926–932.

6. Xu, J., Dowman, M., and Griffiths, T.L. (2013). Cultural transmission results in convergence towards colour term universals. Proc. Biol. Sci. 280, 20123073.

7. Jazayeri, M., and Shadlen, M.N. (2010). Temporal context calibrates interval timing. Nat. Neurosci. 13, 1020–1026.

8. Fischer, B.J., and Peña, J.L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. Nat. Neurosci. 14, 1061–1066.

9. Körding, K.P., and Wolpert, D.M. (2006). Bayesian decision theory in sensorimotor control. Trends Cogn. Sci. 10, 319–326.

10. Griffiths, T.L., and Kalish, M.L. (2007). Language evolution by iterated learning with bayesian agents. Cogn. Sci. 31, 441–480.

11. Feldman, N.H., Griffiths, T.L., and Morgan, J.L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. Psychol. Rev. 116, 752–782.

12. McDermott, J.H. (2009). The cocktail party problem. Curr. Biol. 19, R1024–R1027.

13. Savage, P.E., Brown, S., Sakai, E., and Currie, T.E. (2015). Statistical universals reveal the structures and functions of human music. Proc. Natl. Acad. Sci. USA 112, 8987–8992.

14. Maes, P.J., Leman, M., Palmer, C., and Wanderley, M.M. (2014). Action-based effects on music perception. Front. Psychol. 4, 1008.

15. Large, E.W., and Snyder, J.S. (2009). Pulse and meter as neural resonance. Ann. N Y Acad. Sci. 1169, 46–57.

16. Povel, D.J. (1981). Internal representation of simple temporal patterns. J. Exp. Psychol. Hum. Percept. Perform. 7, 3–18.

17. Temperley, D. (2007). Music and Probability (The MIT Press).

18. Sadakata, M., Desain, P., and Honing, H.J. (2006). The Bayesian way to relate rhythm perception and production. Music Percept. 23, 269–288.

19. Clarke, E.F. (1987). Categorical rhythm perception: an ecological perspective. In Action and Perception in Rhythm and Music, Volume 55, A. Gabrielsson, ed. (Royal Swedish Academy of Music), pp. 19–33.

20. Desain, P., and Honing, H. (2003). The formation of rhythmic categories and metric priming. Perception 32, 341–365.

21. Repp, B.H., London, J., and Keller, P.E. (2013). Systematic distortions in musicians' reproduction of cyclic three-interval rhythms. Music Percept. *30*, 291–305.

22. Repp, B.H., London, J., and Keller, P.E. (2012). Distortions in reproduction of two-interval rhythms: when the "attractor ratio" is not exactly 1:2. Music Percept. *30*, 205–223.

23. Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology (Cambridge University Press).

24. Xu, J., and Griffiths, T.L. (2010). A rational analysis of the effects of memory biases on serial reproduction. Cognit. Psychol. *60*, 107–126.

25. Verhoef, T., Kirby, S., and de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. J. Phonetics *43*, 57–68.

26. Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. Proc. Natl. Acad. Sci. USA *105*, 10681–10686.

27. Honing, H. (2013). The structure and interpretation of rhythm in music. In The Psychology of Music, Third Edition, D. Deutsch, ed. (Academic Press), pp. 369–404.

28. London, J. (2012). Hearing in Time: Psychological Aspects of Musical Meter (Oxford University Press).

29. Turk, A., and Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled? Philos. Trans. R. Soc. Lond. B Biol. Sci. *369*, 20130395.

30. Griffiths, T.L., Christian, B.R., and Kalish, M.L. (2008). Using category structures to test iterated learning as a method for identifying inductive biases. Cogn. Sci. *32*, 68–107.

31. Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. Proc. Biol. Sci. *281*, 20141541.

32. James, S., and Marvin, E. (2015). Categorical perception of irregular rhythms. In 2015 Annual Conference of the Society for Music Perception and Cognition (Vanderbilt).

33. Braun, B., Kochanski, G., Grabe, E., and Rosner, B.S. (2006). Evidence for attractors in English intonation. J. Acoust. Soc. Am. *119*, 4006–4015.

34. Patel, A.D. (2010). Music, Language, and the Brain (Oxford University Press).

35. Margulis, E.H. (2014). On Repeat: How Music Plays the Mind (Oxford University Press).

36. McDermott, J.H., Schultz, A.F., Undurraga, E.A., and Godoy, R.A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. Nature *535*, 547–550.

37. Riester, J. (1978). Canción y Producción en la Vida de un Pueblo Indígena: Los Chimane del Oriente Boliviano (Editorial Los Amigos del Libro).

38. Hannon, E.E., and Trehub, S.E. (2005). Metrical categories in infancy and adulthood. Psychol. Sci. *16*, 48–55.

39. Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. R. Soc. Open Sci. *2*, 150152.

40. Repp, B.H., London, J., and Keller, P.E. (2005). Production and synchronization of uneven rhythms at fast tempi. Music Percept. *23*, 61–78.

41. Tierney, A., Dick, F., Deutsch, D., and Sereno, M. (2013). Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. Cereb. Cortex *23*, 249–254.

42. Bååth, R., Lagerstedt, E., and Gärdenfors, P. (2013). An oscillator model of categorical rhythm perception. In Proceedings of the 35th Annual Conference of the Cognitive Science Society (Cognitive Science Society), pp. 1803–1808.

43. Peper, C., Beek, P.J., and van Wieringen, P.C. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. J. Exp. Psychol. Hum. Percept. Perform. *21*, 1117.

44. Desain, P., and Honing, H., eds. (1989). Quantization of Musical Time: A Connectionist Approach, *Volume 1* (MIT Press).

45. Merchant, H., Zarco, W., Pérez, O., Prado, L., and Bartolo, R. (2011). Measuring time with different neural chronometers during a synchronization-continuation task. Proc. Natl. Acad. Sci. USA *108*, 19784–19789.

46. Polak, R. (2010). Rhythmic feel as meter: non-isochronous beat subdivision in jembe music from Mali. Music Theory Online *16*. http://mto.societymusictheory.org/issues/mto.10.16.4/mto.10.16.4.polak.html.

47. Friberg, A., and Sundströöm, A. (2002). Swing ratios and ensemble timing in jazz performance: evidence for a common rhythmic pattern. Music Percept. *19*, 333–349.

48. Haugen, M.R. (2014). Studying rhythmical structures in Norwegian folk music and dance using motion capture technology: a case study of Norwegian telespringar. Musikk og Tradisjon *28*, 27–52.

49. Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., and Koelsch, S. (2009). Universal recognition of three basic emotions in music. Curr. Biol. *19*, 573–576.

50. Cameron, D.J., Bentley, J., and Grahn, J.A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. Front. Psychol. *6*, 366.

# Supplemental Information

# Integer Ratio Priors on Musical Rhythm

# Revealed Cross-culturally by Iterated Reproduction
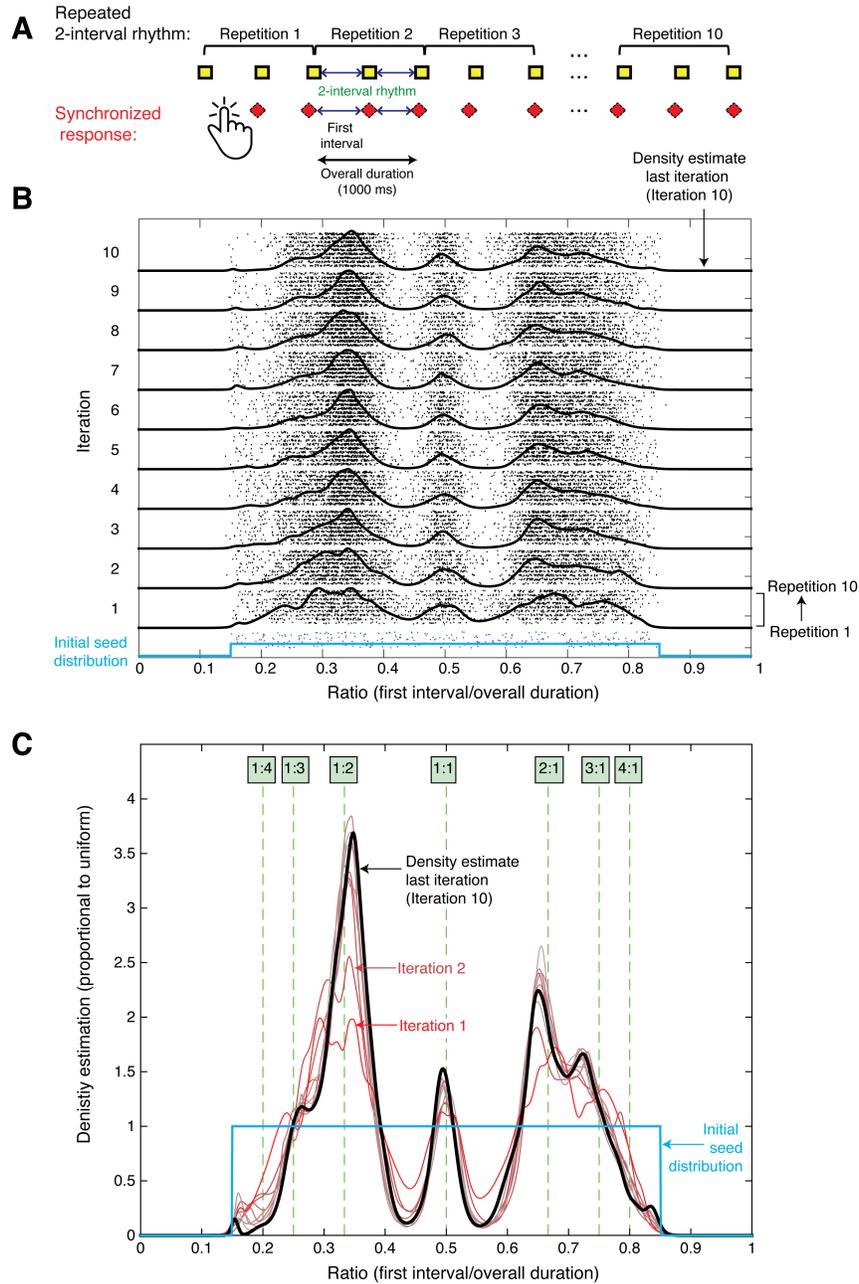
Nori Jacoby and Josh H. McDermott

**Figure S1. Results of Iterated Reproduction of Two-Interval Rhythms (Related to Figure 2)**
(A) Schematic of procedure. The experiment was identical to experiment 1 except that the rhythms were composed of only two time intervals, the overall pattern duration (the sum of the two intervals) was 1000 ms, 10 iterations were collected per trial, and the shortest interval allowed for the initial distribution was 150 ms.
(B) Aggregated results from all trials from 15 subjects. Initial seeds are presented in the lowest row of the diagram. The x coordinate of each dot represents the ratio (the duration of the first interval divided by the pattern duration) produced by a response. The y coordinate represents the serial order (iteration and repetition within the iteration) within the trial. Thick black lines show kernel density estimates for each iteration, pooling across repetitions.
(C) Comparison of density estimates for all 10 iterations. Integer ratios are indicated by dashed green lines.
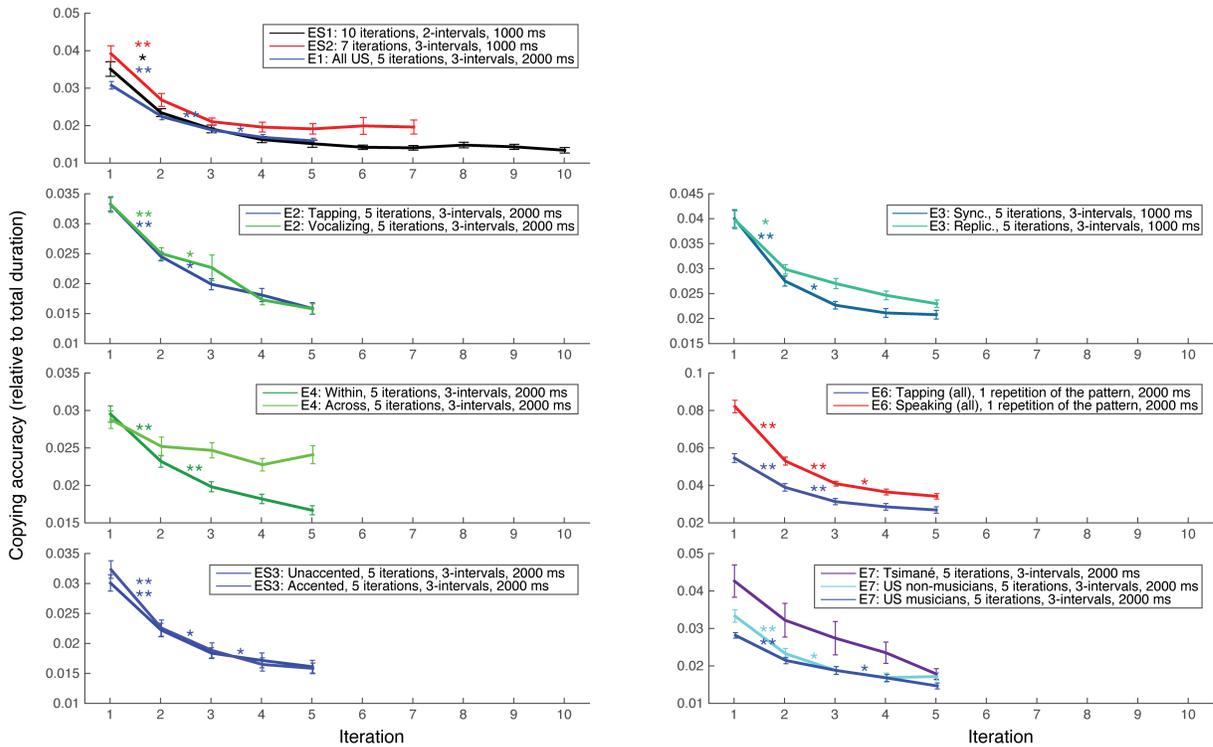
**Figure S2. Copying Accuracy for All Experiments (Related to Figure 2)**

Related to Figure 2. Copying accuracy is plotted vs. iteration number. All significant pairwise comparisons between adjacent iterations are marked by * ($p < 0.05$) and ** ($p < 0.001$). In all cases, convergence seems to occur by about 5 iterations. Error bars plot standard error of the mean. For experiment S1 we also computed the Jensen-Shannon Divergence (JSD). distance between the distributions of the last iteration and of iterations 5–9; this distance was not statistically significant ($p = 0.25$, 0.11, 0.76, 0.43, and 0.23 for iterations 5–9, respectively; significance was computed by bootstrapping), providing further evidence of convergence after 5 iterations.
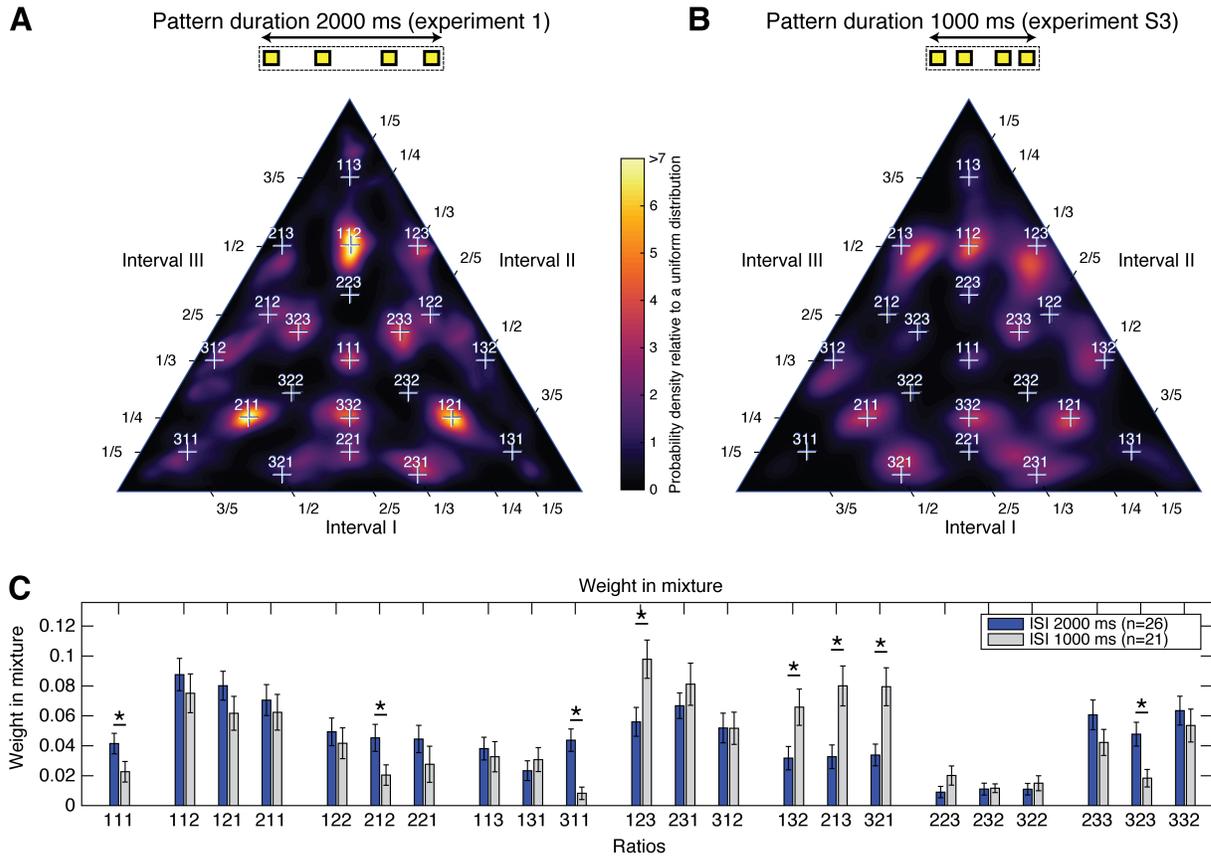
**Figure S3. Results of Experiment S2: Effect of Tempo (Related to Figure 4)**
(A) Response distribution from fifth iteration of the tapping task (experiment 1, re-plotted) with overall pattern duration of 2000 ms.
(B) Response distribution from fifth iteration of tapping task with overall pattern duration of 1000 ms.
(C) Weights of Gaussian mixture components assigned to each integer ratio rhythm in the two distributions. Error bars plot confidence intervals on the weights (standard deviations of the weight distributions derived from bootstrapping), with significance here and elsewhere evaluated after Bonferroni correction for multiple comparisons.
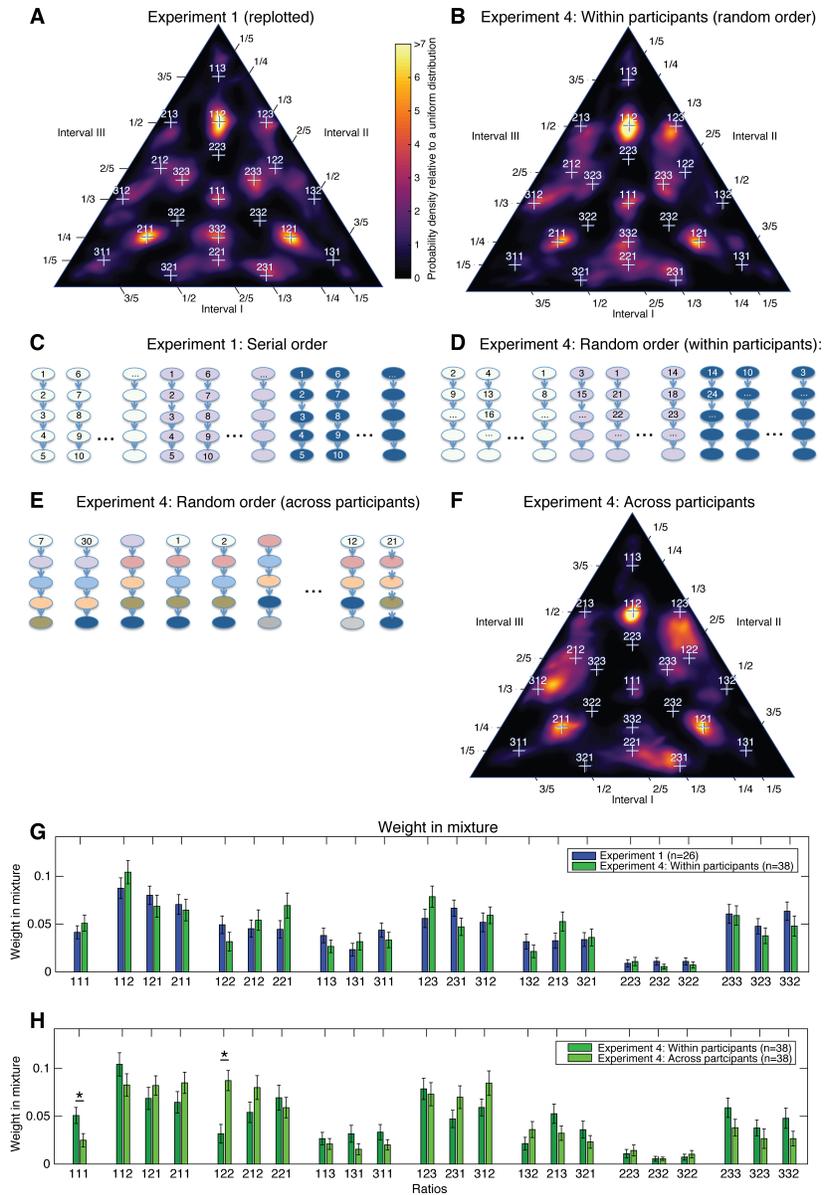
**Figure S4. Results of Experiment 4: Controlling for Memory Biases (Related to Figure 4)**
(A) Results of experiment 1 (re-plotted).
(B) Response distribution from fifth iteration of experiment 4 (within participant condition, randomized order).
(C) Schematic of Experiment 1 format. Ellipses represent iterations, columns represent reproductions of particular seeds, colors represent participants, and numbers indicate the order in which iterations were completed.
(D) Schematic of first condition of experiment 4 (within participant condition, randomized order).
(E) Schematic of second condition of experiment 4 (across participant conditions, with each subject participating once in each chain).
(F) Response distribution from fifth iteration of experiment 4 (across participant conditions)
(G) Weights of Gaussian mixture components assigned to each integer ratio rhythm in experiment 1 and for the within participant condition of experiment 4. Error bars here and in H plot confidence intervals on the weights (standard deviations of the weight distributions derived from bootstrapping).
(H) Weights of Gaussian mixture components assigned to each integer ratio rhythm in the two conditions of experiment 4.
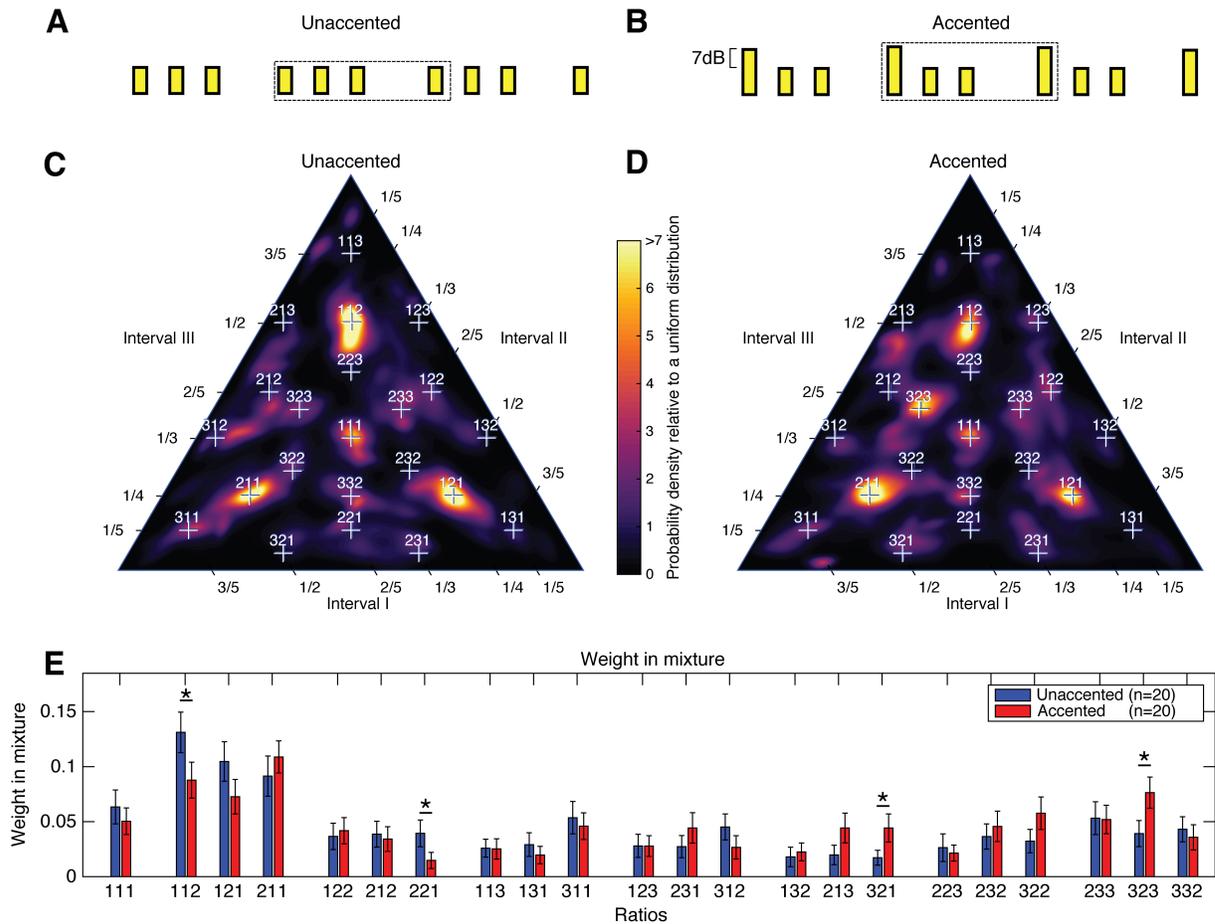
**Figure S5. Results of Experiments S3: Accented Patterns (Related to Figure 6)**

To test whether the ambiguity between cyclic permutations of a rhythm underlies some of the symmetry evident in the response distributions we report, we conducted a follow-up tapping experiment in which the first click of the repeating stimulus was given a 7 dB level increment to render the permutations distinct.

(A) Schematic of unaccented tapping task in which all beats have the same amplitude.

(B) Schematic of the accented condition, in which the first beat of every cycle is 7 dB higher in level.

(C and D) Response distributions for unaccented and accented rhythms, respectively. The 7 dB accent had relatively little effect; both conditions produced results that were similar to those of experiment 1 and that did not exhibit the asymmetry present in Figure 6.

(E) Weights of Gaussian mixture components assigned to each integer ratio rhythm in the two distributions. Although there were significant differences in the fitted GMM weights for 4 of the 22 integer ratio points, the weights for accented and unaccented rhythms were significantly correlated (r = 0.73; p < 0.001; r = 0.90 after correcting for the test-retest reliability of the weights). Error bars plot confidence intervals on the weights (standard deviations of the weight distributions derived from bootstrapping). The results suggest that cyclic ambiguity has only a modest effect on the results of our paradigm.
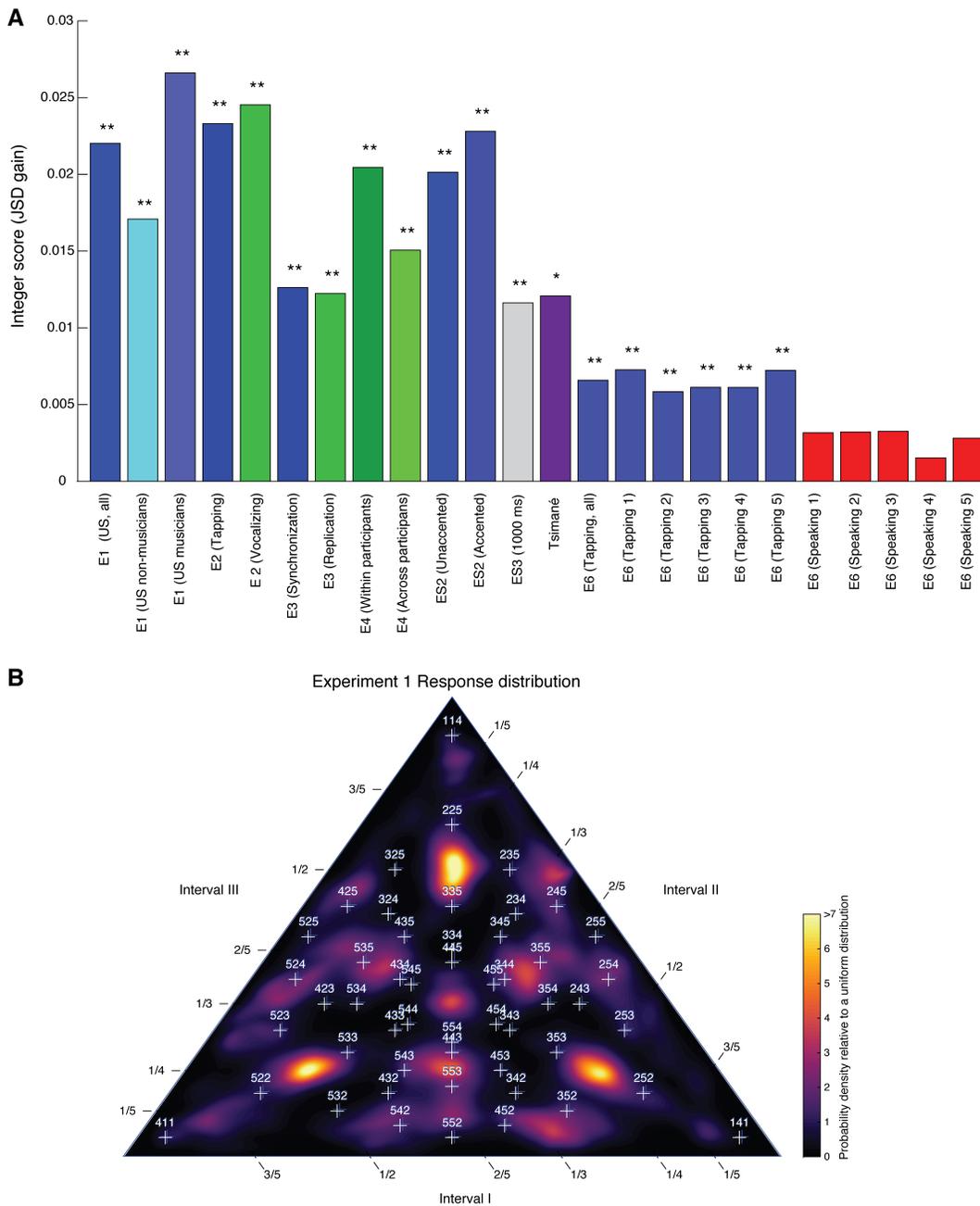
**Figure S6. Integer Ratios (Related to Figure 6)**
(A) Integer scores for all conditions of all experiments. The integer score measures the distance (JSD) between an experimental distribution and a uniform distribution over the 22 simple integer ratio points. To evaluate statistical significance we compared the score produced by the data with that for a mixture of Gaussians with 22 randomly located components. With the exception of those involving spoken phrases, all conditions produced statistically significant integer scores ($p < 0.05$).
(B) Response distribution from experiment 1 with more complex integer ratio rhythms superimposed. The rhythms plotted are those composed of integers between 1 and 5 that do not specify a rhythm identical to one of the simple integer rhythms (the 22 rhythms used throughout this paper composed of integers between 1 and 3). These additional integer ratio rhythms did not significantly overlap with the distribution of experiment 1: the integer score computed using these additional rhythms instead of the 22 simple integer rhythms was not significant ($p = 0.71$).
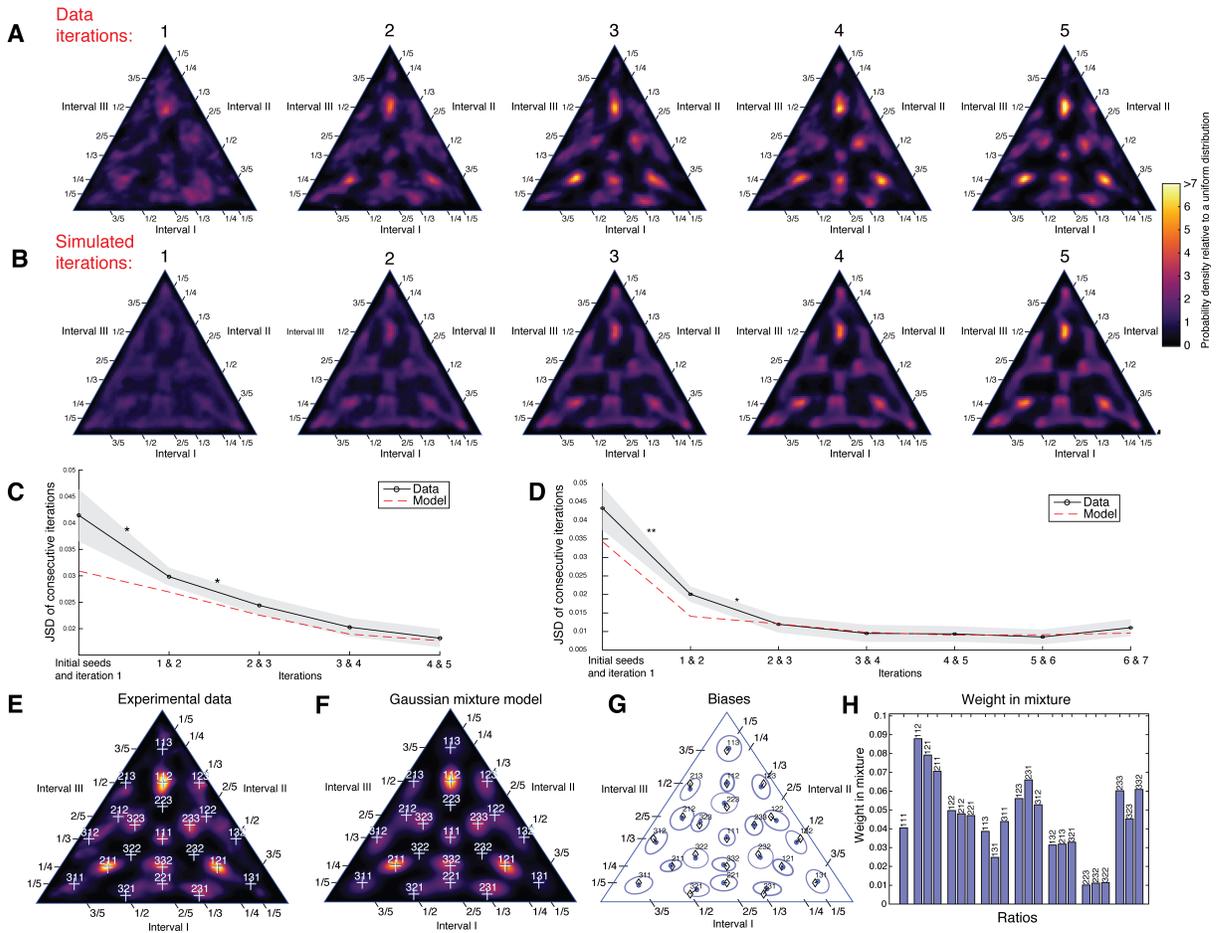
**Figure S7: Dynamics of Convergence in Experimental and Simulated Data (Related to Figures 3, 5, and the Supplemental Experimental Procedures).**

A. The distribution estimated from each iteration of data from experiment 1, normalized relative to a uniform distribution on the simplex: $P(x) = P(x)/U(x)$. B. Simulation results of the rational model (see Methods), using the parameters $\sigma_S = 0.011$ (sensory noise) and $\sigma_R = 0.008$ (production noise), and with the prior taken from iteration 5 of the experimental data. All distributions were computed numerically as described in the methods section.

(C) JSD distance between consecutive distributions in the data and model showing that convergence is reached or nearly reached by the fifth iteration in both the model and the data. Gray region here and in D plots confidence intervals on the JSD (standard deviations of the JSD distribution derived from bootstrapping).

(D) Same as C, but for experiment S3, which had more iterations (seven).

(E) Distribution estimated from data of experiment 1 (using kernel density estimation).

(F) Gaussian Mixture Model distribution fit to the same experimental data, demonstrating the quality of the GMM fits. Note the close similarity to (A).

(G) The weights within of the fitted mixture. H. The means and covariances of the GMM components. Light blue dots and ellipsoids represent the components' mean and covariance, respectively.

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

### Participants

US participants were recruited from the Boston area. All were fluent English speakers. Musical experience was assessed via self-reported number of years spent playing a musical instrument.

Sample sizes were determined by a power analysis in which we estimated the number of participants necessary to produce stable estimates of the response distribution. We used the data from experiment 1 to assess the stability of the response distribution as a function of the number of participants and the number of trials per participant. We used the split-half reliability of the weights of the GMM fitted to the response distribution as a measure of stability. We found that this measure of stability was primarily a function of the total number of trials (doubling the amount of data per participant had a similar effect to doubling the number of participants). We therefore collected as much data as we could per condition and participant in a two-hour session, and recruited sample sizes sufficient to yield an expected test-retest reliability of 0.65 (entailing reliability exceeding 0.5 with 90% probability). The exceptions were experiments 3 and 4, both of which involved novel design elements (faster tempo and replication from memory, and across-participant iterative chains, respectively) that raised the possibility that the results would be more variable than other experiments. Accordingly, sample sizes for these two experiments targeted a mean predicted reliability of ~0.75 given the data properties of experiment 1. Actual sample sizes varied somewhat from targeted numbers due to participant no-shows; the actual number of trials completed by each participant also varied somewhat due to variation in session length (as when participants arrived late).

*Experiment 1 Priors on rhythm in US participants via iterated reproduction:* (n = 26, 16 females; mean age = 26.5, range = 18–43; mean musical experience = 8.3 years, range = 0–23).
The group contained two subgroups (analyzed separately in Figure 7):
   n = 13 Non-musicians: 0–2 years of musical experience (mean 0.5): 8 females, mean age = 26.1, range 18–43.
   n = 13 Musicians: 9–23 years of musical experience (mean musical experience 16.1), 11 were professional musicians or music conservatory students, and 2 were pianists with 10 years of musical experience: 8 females, mean age = 26.9, range 19–40.

*Experiment 2 Tapping and vocalizing*: (n = 25, 8 females; mean age = 39.5, range = 19–62; mean musical experience = 4.1 years, range 0–23).

*Experiment 3 Replication from memory and synchronization*: (n = 28, 20 females; mean age = 27.2, range = 18–63; mean musical experience = 4.1 years, range 0–14).

*Experiment 4 Eliminating dependencies across iteration*: (n = 38, 25 females; mean age = 30.1, range = 18–60; mean musical experience = 11.6 years, range 0–49).

*Experiment 5 Rhythm discrimination*: (n = 19, 9 females; mean age = 40.2 range = 18–63; mean musical experience = 9.4 years, range 0–51).

*Experiment 6 Tapping and speaking:* (n = 39, 21 females; mean age = 36.0, range = 18–62; mean musical experience 5.7, range 0–51)

*Experiment S1 Two-interval rhythms*: (n = 15, 8 females; mean age = 32.9 range = 20–58; mean musical experience = 9.2 years, range 0–51)

*Experiment S2 Shorter (1000 ms) overall duration*: (n = 21, 9 females; mean age = 32.0 range = 18–63; mean musical experience = 6.1 years, range 0–51)

*Experiment S3 Accented patterns*: (n = 20, 11 females; mean age = 38.1 range = 18–63; mean musical experience = 8.3 years, range 0–51).

Participant overlap between experiments: There was no overlap between the participants of experiments 1–4. Due to the finite pool of participants at MIT, some participants who took part in experiments 1–3 participated in

follow up sessions in experiments 5–6 and S1–S4. There was at least a month between sessions, and each session comprised a single experiment. The order of conducted experiments was experiment 1, 2, 3, S1, 6, S2, and S3. Experiment 4 was conducted with a completely separate pool of participants. The following table summarizes the overlap in participants among the 9 experiments:

| Overlap | 2 | 3 | 4 | 5 | 6 | S1 | S2 | S3 |
|---|---|---|---|---|---|---|---|---|
| Experiment 1 | 0 | 0 | 0 | 0 | 3 | 2 | 2 | 0 |
| Experiment 2 | | 0 | 0 | 5 | 2 | 1 | 1 | 2 |
| Experiment 3 | | | 0 | 2 | 2 | 0 | 4 | 2 |
| Experiment 4 | | | | 0 | 0 | 0 | 0 | 0 |
| Experiment 5 | | | | | 5 | 1 | 4 | 8 |
| Experiment 6 | | | | | | 3 | 1 | 3 |
| Experiment S1 | | | | | | | 2 | 1 |
| Experiment S2 | | | | | | | | 3 |

*Experiment 7*: Tsimané participants (n = 15, 9 females; mean age = 24.4, range = 17–43) were recruited from the Tsimané villages of Arenales and Tacural. 4 Tsimané participants self-reported a basic command of Spanish. Tsimané participants generally had little musical experience, and none of them played regularly a musical instrument. 9 participants self-reported playing a musical instrument at least once.

All participants provided informed consent in accordance with the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects.

## Apparatus
All stimuli were presented through Sennheiser HD 280 Pro headphones at a comfortable level chosen by the participant. Finger tapping responses (experiments 1–4, 6, 7, and S1–S3) were recorded with a tapping sensor constructed by the experimenters. The sensor was made from soft material so as to provide minimal auditory feedback. A microphone was attached to the surface of the sensor in order to be sensitive to the lightest touch. Data were acquired with a Focusrite Scarlett 2i2 USB sound card, which simultaneously recorded the microphone output and a split of the headphone signal. Stimulus and response onsets were extracted from the stereo audio signal using a Matlab script (see below). The overall latency and jitter obtained in this way was about 1 ms, measured separately using calibration hardware.

Vocalization and speaking responses (experiments 2 and 5) were recorded with a similar setup, but instead of the tapping sensor we used a Shure SM58 microphone mounted on a microphone stand. Syllable onsets, defined as the very beginning of the voicing onsets, were extracted using a custom Matlab script. Psychophysical measurements (experiment 5) were obtained using the same audio hardware to present the stimulus, but the participants typed their responses directly into the computer.

## Stimuli
Rhythmic patterns consisted of short percussive sounds ("clicks") lasting 55 ms with an attack time of 5 ms Stimuli consisted of a click stream originating from a two- (experiment S1) or three-interval pattern (experiments 1–7, S2, and S3). In each trial the base pattern was repeated continuously; the number of repetitions depended on the condition and experiment. Speech stimuli (experiment 6) consisted of five four-syllable sentences ("A pineapple", "A pea picker", "A teakettle", "Kind King Kato", 'Tell King Toto"). Each sentence was recorded once by one of four native English speakers and then segmented into syllables. One speaker recorded 2 sentences, and 3 other participants (1 female) recorded the other sentences (one each). Timing of syllable onset was defined based on the voicing onsets of each syllable. To realize a given three-interval rhythm we inserted silence of variable durations between each syllable (see Figure 6C).

## Onset extraction
*Tapping*: Tapping stimuli were recorded with a high signal to noise ratio and without crosstalk between channels. This allowed for a simple onset extraction algorithm: we scanned the audio file in 15 second windows, detecting all samples exceeding a relative threshold of 2.25% of the maximal power of the sound waveform in the window. We discarded supra-threshold samples that were not separated from the one preceding them by at least 80 ms (to

eliminate events that corresponded to the same tap), yielding a set of tap onsets. The main analysis challenge was to deal with responses that reflected attentional lapses or motor errors, and that did not correspond to a click. We assumed that biases induced by a prior would be modest (less than 300 ms) and corrected for the fact that humans tend to tap slightly before the beat [S1]. As is standard in the tapping literature [S2], we estimated the mean tapping asynchrony for a participant by matching onsets to the most proximal stimulus. From these pairings we computed the mean asynchrony ($m = mean\ (R_i - S_i)$, where $R_i$ is the time of a response onset corresponding to a stimulus onset $S_i$). We then removed from further analysis all events such that $|(R_i - m) - S_i| > 300$ ms. Note that this procedure is only applied to compute exclusion criteria for taps that are likely to have been made in error, and has no effect on inter-stimulus intervals, because the inter-stimulus interval $R_{i+1} - R_i$ is invariant to a constant time-shift: $(R_{i+1} - m) - (R_i - m) = R_{i+1} - R_i$.

*Speech/Vocalization:* Spoken responses were syllables beginning with a vocal stop or a vowel. This response constraint allowed us to rely on simple heuristics for onset extraction using the power in two different frequency bands. We filtered the signal with two bandpass filters (raised cosine frequency responses, spanning 40–500 Hz and 2700–6000 Hz, respectively). We then computed the amplitude envelopes of both filter outputs (the magnitude of the analytic signal) and found the locations $t_i$ of local maxima in the envelope of the first filter using the Matlab *findpeaks* function (this served to locate the peak of each syllable). We then localized the beginning of the syllable as the time point preceding $t_i$ where the envelope of the second filter first exceeded 4% of the global maximum amplitude of the subband envelope (approximating the point where the amplitude began to ramp up). The resulting set of onsets was trimmed by ignoring events that were not separated from the preceding onset by less than 80 ms. The onset extraction procedure is shown in Figure 6D.

## Simplex definition and initial seed distribution

Following [S8], we mapped a three-interval rhythm with inter-onset intervals $(r_1, r_2, r_3)$ to a point on the simplex: $\frac{r_1}{s}\vec{P}_1 + \frac{r_2}{s}\vec{P}_2 + \frac{r_3}{s}\vec{P}_3 = (\tilde{r}_1, \tilde{r}_2)$ where $s = r_1 + r_2 + r_3$ and "+" is the vector sum of the coordinates on the two dimensional plane of the points $\vec{P}_1, \vec{P}_2$, and $\vec{P}_3$ (Figure 1E-1F). Overall durations ($s = r_1 + r_2 + r_3$) were fixed within each experiment, yielding a one-to-one mapping between all three-interval rhythms possible in each experiment and points on the simplex. Points that are very close to the vertices represent rhythms with at least one very short (near 0 ms) interval. We therefore constrained all intervals in the initial seeds to exceed the perception / production limit (about 100–200 ms; [S3, S4]. Specifically, we limited the initial seed to a smaller simplex (Figure 1E, inner simplex) such that all inter-onset intervals were *larger* than 0.15 of the overall duration (a threshold of 150 and 300 ms. for $s = 1000$ and $2000$, respectively). This results in an inner simplex with vertices: $(f, f), (1 - 2f, f)$ and $\left(\frac{1}{2}, \frac{\sqrt{3}}{2} - f\right)$, where $f = 0.15$. In addition to this visualization, we also describe patterns in terms of ratios $(i_1 : i_2 : i_3)$.

## Integer ratio rhythms

We considered only those integer ratio rhythms composed of the integers 1, 2, and 3. This choice is consistent with findings that more "complex" rhythms are very rare in corpora of classical and folk music [S5, S6]. One explanation for the absence of more complex rhythms is that rhythm representation is dependent on an isochronous common pulse [S7], with the representation limited by the range of isochronous tempi for which beats can be perceived. For example, a rhythm of 5:5:4 with an overall duration of 1000 ms would require a common pulse of $5 + 5 + 4 = 14$ beats, corresponding to a very fast isochronous pulse rate of about 14Hz  (ISI = 71.4 ms), which lies beyond the limits of beat perception (about 10Hz; [S4, S6]).

The set of integer ratios we considered is further supported by the post hoc observation that there are no modes evident in the response distribution of Westerners that deviate substantially from these rhythm patterns. We considered the modes of the empirical prior relative to these exact ratios, and indeed found small but significant biases away from them. However, these biases do not appear to correspond to higher-order ratios. To assess this issue we examined rhythms composed of integers between 1 and 5 in relation to the empirical prior measured in experiment 1. We included only rhythms that fall inside the rhythm simplex (i.e. that do not contain intervals shorter than 15% of the rhythm duration) and that are not equivalent to a simpler ratio. The resulting 51 additional rhythms are shown in Figure S6. These additional integer ratio rhythms did not significantly overlap with the distribution of experiment 1: the integer score computed just for the additional rhythms was not

significant (p = 0.74). It thus appears that the simple integer ratio rhythms considered in the main paper are an appropriate choice for our purposes.

## Psychophysical Procedures

### Experiment 1: Priors on Rhythm in Westerners via Iterated Reproduction

Participants were instructed to synchronize their finger tapping to a stimulus. On each trial we set the seed to a point on the inner simplex (uniformly distributed) and presented the corresponding three-interval rhythm $(s_1, s_2, s_3)$. We then generated a sequence of clicks by repeating the three-interval seed pattern 10 times. After a few clicks (typically 3 or 4), participants began to synchronize to the click stream. A Matlab script extracted response onsets (see above). We averaged the inter-response interval across the 10 repetitions, obtaining an averaged three-interval pattern $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$. We defined a valid iteration as one which had 7 or more repetitions with three taps each *and* whose average response $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$ was not located far beyond the inner simplex, i.e., did not contain an interval shorter than $0.95f$ of the overall duration (where $f$ determines the boundary of the inner simplex as described above; here this constraint excludes intervals shorter than 285 ms). If the iteration satisfied the two criteria above, we set the new seed pattern to the response pattern: $(s_1, s_2, s_3) \leftarrow (\bar{r}_1, \bar{r}_2, \bar{r}_3)$. If an iteration was invalid, the seed remained unchanged, and the data from that iteration was omitted from analysis. We then repeated this process 5 times. Participants completed 20–25 such trials (where each trial is defined as a set of 5 iterations derived from a single seed), yielding 609 trials (seeds) in total.

### Experiment 2: Tapping and Vocalizing

The experiment involved two conditions, tapping and vocalizing. The procedure of the tapping condition was identical to that of experiment 1, while the vocalization condition differed only in that the responses were vocalizations of the syllable /tú/ instead of finger tapping. Participants completed 10–20 trials (seeds) per condition in separate blocks, yielding 327 and 330 trials in total for tapping and vocalizing, respectively. Condition order was counterbalanced across participants.

### Experiment 3: Replication from Memory and Synchronization

The experiment involved two conditions: synchronization and replication from memory. The order of conditions was counterbalanced across participants. The synchronization condition was identical to that of experiment 1 except that the stimulus duration was shorter (1000 ms instead of 2000 ms, to reduce the memory load in the replication condition). To compensate for the different length of the stimuli in the two conditions we restricted analysis to 10 responses from each iteration of the synchronization condition. Participants tapped in time with all 10 repetitions, but the stimulus for the next iteration was determined only from the taps for clicks 16–25.

In the replication-from-memory condition, the seed pattern was repeated 3 times in each iteration. The entire click stream thus lasted 3 s and contained 10 clicks. Participants replicated the entire click stream from memory (10 taps). Participants were notified to start tapping by the appearance of a white bar on the computer screen following the stimulus presentation. Iterations in which the number of taps was not equal to the number of clicks were thrown out, and the iteration was repeated using the same stimulus. To make sure that we used an identical exclusion criterion for the continuation and synchronization conditions, we aligned the responses and stimuli by shifting the responses by a constant amount of time so that the first beat of the response is perfectly aligned with the first beat of the stimulus. We then used the same response exclusion procedure as in experiment 1, and averaged the inter-response interval across the 3 repetitions, obtaining an averaged three-interval pattern $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$. The pattern was then scaled so the overall duration remained 1000 ms. An iteration was thrown out if the average response $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$ was located far beyond the inner simplex, i.e., contained an interval shorter than 142.5 ms. If an iteration was valid, we set the new seed pattern to the response pattern: $(s_1, s_2, s_3) \leftarrow (\bar{r}_1, \bar{r}_2, \bar{r}_3)$, but if the iteration was invalid, participants were notified by a red bar on the computer screen. Participants then had two additional opportunities to hear and replicate the pattern. If a participant failed on the third attempt, the iteration was omitted from analysis, and the seed in the next iteration was set to the output of the previous iteration.

In order to encourage participants to listen to each pattern rather than rely on a memory of a previous version of the pattern, we modified the sequencing of the experiment, intermixing the first iterations for each of a set of 5 seed patterns. After these first iterations were completed, we did the same for the second iterations. This process continued until participants completed 5 iterations of each of the 5 patterns. We then repeated the process with 5

new seeds. Most participants performed 4–6 sets of 5 seeds (5 iterations each). The same ordering procedure was used for the synchronization condition. Participants completed 10–30 trials (seeds) per condition in separate blocks, yielding 584 and 561 trials in total for synchronization and replication, respectively. Condition order was counterbalanced across participants.

### Experiment 4: Within and Across Participants

Each participant completed two conditions: within-participant and across-participant (order was counter balanced). In the first condition (within-participant), each participant performed five iterations of 15 iterated reproduction chains, each starting from a different seed, with iterations completed in random rather than serial order (see Figure S4D). In the second condition (across-participant), each participant completed one iteration of up to 75 chains, randomly selected from a set of 100. Each participant contributed to only a subset of the 100 chains in order to increase the diversity of the participant orderings within chains. After all five iterations of the 100 chains were completed (this typically required 6–7 participants, depending on the random order in which iteration "slots" were filled) we started a new set of 100 chains. Five sets of 100 chains were run to completion in this way (using 38 participants). As a result of this design, the last few participants in each set of chains contributed to fewer than 75 chains.

### Experiment 5: Rhythm Discrimination

We selected seven points on the inner rhythm simplex between the patterns 3:2:3 and 1:1:1 (Figure 5B). We refer to these patterns as $p_i$, where $p_1 = 3:2:3$, $p_2=3:2.166:3$, $p_3=3:2.33:3$, … , $p_7 = 3:3:3$ (= $1:1:1$). On each trial listeners were presented with two three-interval patterns $(p_a, p_b)$. The first pattern $p_a$ was presented two times with no intermission (7 clicks, 6 inter-onset intervals), followed by 1000 ms of silence, followed by two repetitions of the second pattern $(p_b)$. After hearing the stimulus, listeners judged whether they were the same or different. Each listener performed 112 trials in random order, half of which featured two identical patterns $(a = b)$, and half of which featured different patterns. The latter were chosen so that the patterns were two steps apart $(|a - b| = 2)$. Hit rate $h_i$ corresponding to the pattern $i$, was defined as the proportion of times a listener made a different judgment when patterns were either $(p_{i-1}, p_{i+1})$ or $(p_{i+1}, p_{i-1})$. A correct rejection rate $f_i$ was defined as the proportion of times a listener made a same judgment when the patterns were either $(p_{i-1}, p_{i-1})$ or $(p_{i+1}, p_{i+1})$. We computed sensitivity ($d$ prime) for pattern $i$ from these hit and correct rejection rates.

### Experiment 6: Tapping and Speaking

The experiment had two conditions, speaking and tapping. In the speaking condition, we synthesized a phrase for each seed pattern $(s_1, s_2, s_3)$ by adding silence between the syllables of one of the 5 phrases, generating a phrase with inter-syllable intervals of $(s_1, s_2, s_3)$. We selected the original speech phrases so that each syllable's overall duration was less than 300 ms in order to ensure that every possible rhythmic pattern within the inner simplex was realizable without producing overlap between syllables. Participants were instructed to replicate the content, prosody, and rhythm of the phrase to the best of their abilities. After each sentence presentation, we ran the onset extraction algorithm to extract the onset of each of the spoken syllables.

A priori we thought that completing iterations for a particular phrase in succession might decrease the chances that participants would represent the phrases as speech. We thus randomly intermixed the iterations for different seeds, unlike in most of the other experiments. A single trial thus consisted of 5 iterations of each of the 5 phrases; the mth iteration presented each phrase in a random order. Seven such trials were completed by each participant. Critically, the same procedure was employed for the tapping condition. We note that the results of experiment 4, which also randomly intermixed iterations for different seeds, suggests that the random intermixing has little effect, at least for tapping. The procedure was otherwise identical to the replication condition of experiment 3, with the exception that we did not average the responses across repetitions, since the pattern was presented a single time. The tapping condition was identical in procedure to the speech condition, but stimuli were click streams instead of speech.

### Experiment 7: Priors in Native Amazonians

The procedure was identical to experiment 1. Participants completed 5–10 trials (seeds) each, yielding 121 trials in total. To ensure that the smaller amount of total data in Tsimané compared to US participants did not account for the differences we observed in their distributions, we examined subsets of 100 trials from the US participants. The distributions from these subsets of US data generally appeared similar to the distribution from the entire data

set, and consistently distinct from the Tsimané distribution. We note also that the statistical significance calculations for the Jensen-Shannon Divergence (JSD) and GMM weights appropriately incorporate the number of trials per data set, in that the null distribution inherently has higher variance for smaller data sets.

### Experiment S1: Two-Interval Rhythms
The experiment was identical to experiment 1 except that the rhythms were composed of only two time intervals, the overall pattern duration (the sum of the two intervals) was 1000 ms, 10 iterations were collected per trial, and the shortest interval allowed for the initial distribution was 150 ms. The data inclusion criterion was also slightly different: an iteration was thrown out if the one of the average response intervals was shorter than 127.5 ms. Participants completed 21–27 trials, yielding 379 trials (seeds) in total.

### Experiment  S2: Different Tempo
The procedure was identical to that of experiment 1 except that we performed 7 iterations and that the total duration of the pattern was 1000 ms. The shortest interval of the initial distribution was scaled accordingly (150 ms), allowing us to use the same simplex visualization to describe these two experiments. As in other experiments, an iteration was thrown out if the average response $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$ was located far beyond the inner simplex, in this case containing an interval shorter than 142.5 ms. Participants completed 12–22 trials, yielding 359 trials (seeds) in total.

### Experiment S3: Accented Patterns
This procedure was identical to experiment 1. In the unaccented condition we performed the exact same procedure of experiment 1, and in the accented condition (order counter balanced) the first beat of each of the cyclically repeated rhythm was 7 dB louder than the other beats. Participants completed 11–15 trials, yielding 277 and 281 trials (seeds) trials in total for accented and unaccented conditions, respectively.

### Data Analysis and Statistics
Bonferroni correction for multiple comparisons was performed on all sets of multiple statistical tests.

### Kernel Density Estimation
*Experiments 1, 2, 4,7, S2, and S3*
In the kernel density estimation procedure, we took advantage of the fact that each pattern was repeated $J = 10$ times. We assumed that the response on the $j^{th}$ repetition $\left(r_{i,m}^{j,1}, r_{i,m}^{j,2}, r_{i,m}^{j,3}\right)$ was taken from a Gaussian distribution: $\left(r_{i,m}^{j,1}, r_{i,m}^{j,2}, r_{i,m}^{j,3}\right) \sim N(\mu_{i,m}, C_{i,m})$. Here $r_{i,m}^{j,k}$ is the inter-response interval for click $k$ of repetition $j$ of iteration $m$ of trial $i$. We estimated the mean and covariance of the Gaussian for each iteration and trial and then summed all Gaussians to yield an estimate of the distribution. This Gaussian is intended as a low parameter approximation of the prior in the local neighborhood of the stimulus. This procedure is equivalent to using only the means of all responses from all iterations and performing a standard kernel density estimate with the kernel width computed from the local estimates, obviating the need to choose the kernel width.
We computed the empirical mean vector $\mu_{i,m}$ and covariance matrix $C_{i,m}$:

$$\mu_{i,m}^k = \frac{1}{J} \sum_{j=1,\dots,J} r_{i,m}^{j,k}$$

$$C_{i,m}^{k,l} = \frac{1}{J-1} \sum_{i=1,\dots,J} (r_{i,m}^{j,k} - \mu_{i,m}^k)(r_{i,m}^{j,l} - \mu_{i,m}^l)$$

Since the matrix is computed based on a small number of samples, we added a regularization term: $C'_{i,m} = C_{i,m} + \lambda^2 I$, where $I$ is the identity matrix and $\lambda$ is a small regularization constant ($\lambda = 10$ ms).

We denoted by $M$ the linear transformation mapping between (three-dimensional) normalized inter-onset intervals $\left(\frac{r_{i,m}^{j,1}}{s}, \frac{r_{i,m}^{j,2}}{s}, \frac{r_{i,m}^{j,3}}{s}\right)$ and the corresponding points on the simplex $\left(\tilde{r}_{i,m}^{j,1}, \tilde{r}_{i,m}^{j,2}\right)$, where $s = r_{i,m}^{j,1} + r_{i,m}^{j,2} + r_{i,m}^{j,3}$ is the overall pattern duration. This mapping can be written as: $M\left(\frac{r_{i,m}^{j,1}}{s}, \frac{r_{i,m}^{j,2}}{s}, \frac{r_{i,m}^{j,3}}{s}\right) = \left(\tilde{r}_{i,m}^{j,1}, \tilde{r}_{i,m}^{j,2}\right)$ .

When the vertices of the simplex are: $\{(0,0), (1,0), (1/2, \sqrt{3/4})\}$ (an equilateral triangle), this transformation is represented by the matrix:

$$M = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1/2 & \sqrt{3/4} \end{bmatrix}$$

We can therefore project the multivariate distribution to the 2-dimensional simplex:

$$\tilde{\mu}_{i,m} = M\mu_{i,m}; \; \tilde{C}_{i,m} = M^T C'_{i,m} M$$

Yielding the density estimate for that iteration: $P_{i,m}(x) = \dfrac{1}{2\pi\sqrt{|\tilde{C}_{i,m}|}} \exp\left(-\frac{1}{2}(x - \tilde{\mu}_{i,m})^T (\tilde{C}_{i,m})^{-1}(x - \tilde{\mu}_{i,m})\right)$

To obtain the full kernel density estimate, we aggregate the Gaussians for each of the five iterations across all trials: $P_m(x) = \left(\frac{1}{T}\right) \Sigma_i P_{i,m}(x)$ where $T$ is the total number of trials. We denoted by $U(x)$ the uniform distribution on the inner simplex. To facilitate comparison between distributions, the distribution estimates in Figures 2–7, S1–S7 were normalized relative to the uniform distribution, namely: $P_m(x) = \dfrac{P_m(x)}{U(x)}$. In practice, we computed these densities with bin size of $0.005 \times 0.005$ on the rhythm simplex.

*Experiments S1 (2-interval rhythms)*
We adapted the procedure of the three-interval rhythm to two dimensions, replacing the two-dimensional simplex by a one-dimensional simplex (a line). The procedure was otherwise identical. For density estimates we used a bin size of $0.001$.

*Experiments 3 and 6*
Experiments 3 and 6 did not have enough repetitions in each trial to reliably compute the covariance matrix. We therefore used the same procedure above, but fixed the kernel width so that $C'_{i,m} = \lambda^2 I$ with $\lambda =40$ ms. This value was selected based on the estimation of response error from experiments 1 and 2 and from previous literature [S1].

**Testing Distance between Distributions using the JSD Method**
To compute the distance between distributions we used the JSD. The JSD of two distributions $P$ and $Q$ is defined by: $JSD(P,Q) = \frac{1}{2}D_{KL}(P,M) + \frac{1}{2}D_{KL}(Q,M)$ where: $M = \frac{1}{2}(P + Q)$
and $D_{KL}(P,Q)$ is the Kullback-Leibler divergence $D_{KL}(P,Q) = \int_x P(x)\log_2 \frac{P(x)}{Q(x)} dx$. Note that the JSD is symmetric and bounded between 0 and 1, and is 0 when $P = Q$.

**Permutation Test for the Significance of the Difference between Groups of Participants**
Recall that each distribution estimate for iteration $m$ was computed based on averaging across all trials:

$$P_m = \frac{1}{T_1}\sum_i P_{i,m} \; ; Q_m = \frac{1}{T_2}\sum_i Q_{i,m}$$

To compute the significance of the JSD between two such distributions we compared the actual JSD with one obtained by a permutation test. We reshuffled all trials into two new randomized groups of the same size. Namely, we denote by $Q_{i,m}$ all the $T_1 + T_2$ trials of $P_{i,m}$ and $Q_{i,m}$
We randomize $N = 1000$ permutations $\pi_1, \dots, \pi_n$ of length $T_1 + T_2$ and compute:

$$j_m^n = JSD\left(\frac{1}{T_1}\Sigma_{i=1,\dots,T_1} R_{\pi_n(i),m}, \frac{1}{T_1}\Sigma_{i=T_1,\dots,(T_1+T_2)}R_{\pi_n(i),m}\right)$$

The values reported in the Results section are the probability of the actual JSD under the resulting null distribution.

**Comparing the Role of Culture and the Role of Expertise**
To compare the role of musical expertise to that of culture, we calculated the JSD distance between the distribution of non-musicians ($P_1$) and musicians ($P_2$), and between the distribution of Tsimané ($P_3$) to that of non-musicians $P_2$ (approximately matched to the Tsimané in age and musical expertise). We compared the

difference between JSDs: $JSD(P_3, P_2) - JSD(P_1, P_2)$ to the difference obtained by reshuffling the datasets into three new randomized groups of the same sizes. The values reported in the Results section are the probability of the actual difference in JSDs under the resulting null distribution.

**Testing Deviation from a Uniform Distribution Using JSD Distance**

To quantitatively test whether the distribution of seeds presented in Figure 2 significantly deviated from a uniform distribution over the inner simplex, we computed the JSD distance between the kernel density estimate from all seeds and a uniform distribution. To evaluate statistical significance of this JSD we compared it to a null distribution of JSD distances computed from samples from a uniform distribution. We applied kernel density estimation to these samples and computed the JSD from a uniform distribution as with the experimental sample. This was necessary because the JSD between two distributions is always non-negative, such that any distribution evaluated from a finite number of points would have a non-zero distance from a uniform distribution. For all iterations we used a fixed kernel width of 40 ms.

We also used the classical method suggested by [S9] without relying on bootstrapping. The inner simplex was divided into 900 equilateral triangles, and we computed the histogram of the number of responses in each cell and compared it to the one obtained under the null hypothesis of a uniform distribution using the $\chi^2$ goodness of fit statistics.

Consistent with the JSD results, the initial distribution of experiment 1 was not significantly different from a uniform distribution measured by a $\chi^2$ goodness of fit ($\chi^2(9)=8.41$ p = 0.49). And the distributions in all five subsequent iterations were significantly different from uniform, with progressively increasing $\chi^2$ values, indicating increasing deviation from uniformity ($\chi^2(9)$ = 376.5, 1111.9, 1916.7, 2049.2, 2842.1; p < 0.001 in all cases).

**Copying Accuracy**

We computed copying accuracy as the average distance between response and stimuli. Namely, if the seed $i$ at iteration $m$ is given by $(s_{i,m}^1, s_{i,m}^2, s_{i,m}^3)$ and response averages (that are used as the next seed) are given by $(r_{i,m}^1, r_{i,m}^2, r_{i,m}^3)$ we defined the copying accuracy to be: $d = \sqrt{\left(\frac{1}{3}\sum_{k=1}^3 \left(r_{i,m}^k - s_{i,m}^k\right)^2\right)}$.

In the case of two interval rhythms we used a similar definition: $d = \sqrt{\left(\frac{1}{2}\sum_{k=1}^2 \left(r_{i,m}^k - s_{i,m}^k\right)^2\right)}$.

We then average the responses for all trials for a given participant and compare the results across participants and conditions. To obtain a value that can be compared across different total durations (2000 ms for experiments 1–4, 6, 7, S3 and 1000 ms for experiments S1 and S2) we normalize the value by total duration, obtaining a Weber fraction.

We also used the JSD between distributions from successive iterations to quantify convergence. For experiment 1 this measure showed similar results, progressively decreasing until the fourth iteration (JSD = 0.030, 0.024, 0.020 and 0.018, for iterations 1–2, 2–3, 3–4, and 4–5, respectively; the difference between the two last pairs was not significant, p = 0.20), again suggesting that convergence is reached or nearly reached by the fifth iteration. Statistical significance of this measure was obtained using bootstrapping. We used this measure to compare convergence of the model and the data (Figure S7C and S7D)

**Gaussian Mixture Model Fitting**

A Gaussian Mixture Model (GMM) was used to estimate weights for different integer ratio rhythms in the distributions measured in our experiments. The GMM on the two dimensional plane is the following distribution:

$$Q_{\{w_k, \vec{\mu}_k, C_k\}_{k=1,\ldots,K}} = \sum_{k=1,\ldots,K} \frac{w_k}{2\pi\sqrt{|C_k|}} \exp\left(-\frac{1}{2}\left(x - \vec{\mu}_k\right)^T \left(C_k\right)^{-1}\left(x - \vec{\mu}_k\right)\right)$$

where $\{\vec{\mu}_k\}_{k=1,\ldots,K}$, $\{C_k\}_{k=1,\ldots,K}$, and $\{w_k\}_{k=1,\ldots,K}$ are mean vectors, covariance matrices and weight scalars, respectively. We fit a GMM to a probability density $P$ by minimizing the Kullback–Leibler divergence over the set of all Gaussian mixtures: $Q = \text{argmax}_Q D_{KL}(Q_{\{w_k, \vec{\mu}_k, C_k\}_{k=1,\ldots,K}}, P)$

To compare different distributions with peaks near simple integer ratio points, we used $K = 22$ components (one for each integer ratio), and constrained the means of each component to a square of size $0.025 \times 0.025$ centered around the corresponding integer ratio point. Minimization was performed using Matlab's *fmincon* function. These constraints allowed us to uniquely associate every component in the mixture with a simple integer ratio, facilitating comparison between weights and biases across more than one probability density (Figures 3, 4, 7, and S3–S5). Figures S7E–H demonstrate the quality of the obtained fits.

We used bootstrapping to compute error bars on the weights. Recall that the distribution $P$ is a kernel density estimate obtained by aggregating $T$ responses. We sampled $T$ responses with replacement and computed the GMM fits for each of $N = 100$ bootstrapped datasets. Figures 3, 4, 7, and S3–S5 plot the mean and standard deviation of these bootstrapped weight estimates. The same bootstrapping process provides estimates of the variability of the mean of each component. To estimate the significance of the biases of the distribution modes $\vec{\mu}_k$ from the corresponding integer point $\omega_k$ we evaluated the probability of obtaining a bias greater than or equal to the actual bias $r_0 = ||\vec{\mu}_k - \omega_k||$ under a two-dimensional Gaussian distribution fitted to the means obtained by bootstrapping. To estimate the correlation between the weights for any two experimental conditions we computed the Pearson correlation between the 22-dimensional vectors of all attractor weights. Correction for attenuation was done using Spearman's method: $\frac{\text{corr}(X,Y)}{\sqrt{R_x R_y}}$ where $\text{corr}(X,Y)$ is the correlation coefficient of the two sets of weights, and $R_x, R_y$ are the estimates for test-retest reliability of the weights of each condition based on a split-half procedure.

### Integer Score
As a measure of compatibility between a given distribution $P$ and the set of relevant simple integer ratio points, we examined the JSD between $P$ and the normalized indicating function of the set $\Omega$ of all ratios $(i{:}j{:}k)$ for which $1 \leq i, j, k \leq 3$, namely: $JSD(P, Q_\Omega)$, where: $Q_\Omega = \frac{1}{|\Omega|}\mathbb{I}_\Omega(x) = \frac{1}{|\Omega|}\sum_{\omega \in \Omega} \delta(x - \omega)$,
and $\delta$ is the Dirac delta function. This score is maximum for a distribution that has mass at all of the integer ratios and only at the integer ratios, and is lower for distributions that have mass off of the integer ratios or at fewer of them.

To evaluate statistical significance, we compared the score produced by the data $JSD(P, Q_\Omega)$ with random distributions with the same number of modes. We initially fitted an unconstrained GMM with 22 components to the distribution. We then created 10,000 randomized distributions $P_i$ by randomizing the means of the components of the mixture while keeping the weights and covariance matrices. The means were randomized independently from a uniform distribution over the simplex. The statistical significance was obtained as the probability of the observed JSD in this null distribution. To obtain Figure S6, which compares the results from all experiments we computed the gain in the integer score of the data compared with the mean score obtained by randomization. Namely:

$$s_\Omega(P) = E_{P_i}\big(JSD(P_i, Q_\Omega) - JSD(P, Q_\Omega)\big)$$

### A Rational Model of the Experimental Procedure
We argue in the main text, on empirical grounds, that our iterated reproduction method converges to the prior over simple rhythms. Here we present theoretical evidence that the procedure produces the prior. Our analysis follows that of [S10], who analyzed the relationship between serial reproduction and one-dimensional priors with one or two categories, and [S11], who describe a rational model of iterated learning. Similar to [S10], we assume that participants within a group share the same prior. Our main paradigm (experiment 1) differs from that modeled by [S10] in three respects: (i) the prior contains more than one mode as in [S12, S13]; (ii) the prior is two-dimensional; (iii) our paradigm has an averaging step in which the repeated reproductions within each iteration are averaged.

Here we show that when reproduction noise is small, and the prior distribution is approximately a mixture of Gaussians with relatively small variances, the procedure of experiment 1 converges to a distribution that approximates the true prior.

Let us define the following Markov chain:

- $\ldots \to X_m \to \left[S_m^j\right]_{j=1,\ldots,J} \to \left[S'^j_m\right]_{j=1,\ldots,J} \to \left[R_m^j\right]_{j=1,\ldots,J} \to X_{m+1}$

Where:

- $X_m$ is the stimulus on iteration $m$.
- $S_m^j$ is the perceived stimulus at repetition $j$ where $j = 1, \ldots, J$.
- $S'^j_m$ is the reconstructed stimulus (what the participant tries to reproduce).
- $R_m^j$ are the reproduced responses (with additional production noise).

The paradigm structure determines:

- The averaging procedure: $X_{m+1} = (1/J) \sum_{j=1,\ldots,J} R_m^j$
- The initial (uniform) distribution: $X_0 \sim U$

We will make the following assumptions:

- The sensory noise is Gaussian and unbiased: $S_m^j | X_m \sim N(X_m, \sigma_S^2)$.
- The reproduction noise is Gaussian and unbiased: $R_m^j | S'^j_m \sim N(S'^j_m, \sigma_R^2)$.
- The participant stores a prior distribution $P(X)$.
- The participant uses the "sample-sample" strategy from [S9], i.e. the reproduction is a sample of the posterior distribution $P(X|S)$.

In addition we can use Bayes's rule to obtain: $p(X|S) = \frac{P(S|X)P(X)}{\int P(S|X')P(X')dX'}$

Note that $P(X)$ is the prior and $P(S|X) \sim N(X, \sigma_S^2)$.

As a consistency check, we simulated an experiment assuming a prior distribution identical to that obtained by the final iteration of experiment 1. The parameters $\sigma_R, \sigma_S$ were selected to minimize the sum of JSDs between the distributions in all iteration of the experiment and the iteration of the experimental data, yielding $\sigma_R = 0.011$, $\sigma_S = 0.008$.

Figure S7 shows the close match between the data (Figure S7A) and simulation (Figure S7B). Note that all distributions were normalized relative to the same uniform distribution ($P(X = x)/P(U = x)$). Critically, the simulation results show that the simulated responses converge to the prior. This result provides some justification for the direct estimation of the prior from the final iteration of the experimental process.

## Bayesian Observer Model

This analysis closely follows [S13], who analyzed a similar effect known as the "perceptual magnet effect" for phonetic categories. The analysis is also compatible with the aforementioned rational model of the iterated process. As above, we assume that the participant has a noisy observation $S$ of an external stimulus $X$. To perform the discrimination we assume that the participant estimates the posterior: $P(\hat{X}|S) \propto P(S|\hat{X})P(\hat{X})$, where $\hat{X}$ is the estimated value of the stimulus, and $P(\hat{X})$ is the prior (obtained from the iterated experiment). We assumed that the perceptual noise is Gaussian and unbiased, namely: $P(S|X) = N(X, \sigma_S^2)$. If the participant is presented with two stimuli $X_1$ and $X_2$, we assume that a "different" answer is given if $|\hat{X}_1 - \hat{X}_2| > e$ where $e$ is an internal criteria. This random variable has the following distribution:

$$P\left(|\hat{X}_1 - \hat{X}_2| > e|X_1, X_2\right) = \int P\left(|\hat{X}_1 - \hat{X}_2| > e|X_1, X_2, S_1, S_2\right) p(S_1, S_2|X_1, X_2) dS_1 dS_2$$

This integral can be computed numerically by using the distribution $P(\hat{X}|S)$, and by noting that $X_1$ and $X_2$ are independent so that: $p(S_1, S_2|X_1, X_2) = p(S_1|X_1)p(S_2|X_2)$. The only free parameters of the model given the prior $p(X)$ are the sensory noise $\sigma_S^2$ and the criteria $e$. To obtain Figure 5, we fitted $\sigma_S^2$ and $e$ for each participant by minimizing the squared error between the hit rate of the participant and the model (we did not use the correct rejection nor the model sensitivity to perform fitting, these parameters were predicted from the model using the obtained parameters).

## Comparing the Efficiency of Iterative and Discrimination Paradigms

To compare the time required to achieve equivalent results with discrimination and iterative paradigms, we simulated estimates of the prior from a discrimination experiment and compared them with estimates obtained with the iterative paradigm. The discrimination experiment contains "different" trials, where participants perform a pairwise comparison between two stimuli that differ by a fixed distance, as well as "same" trials, where participants judged identical stimuli. All stimuli were positioned on a two-dimensional grid spaced by d = 0.024 horizontally and vertically. This distance was selected such that we would obtain a percent correct of about 80% given the average parameters of experiment 5 (and thus the performance in the discrimination task would not be at ceiling or floor). To avoid overfitting, we split the data from the iterative procedure (from experiment 1) into two non-overlapping halves. We simulated a discrimination experiment using the model of experiment 5 and the empirical prior estimated from of the first half of the data from experiment 1. We then fit a prior to these results using the same perceptual model such that the squared error between the accuracy of the simulated and the fitted data was minimized. We varied the number of trials per condition, in each case computing the JSD distance between the obtained estimate of the prior and the prior obtained from the second half of the data from experiment 1. We compared these JSD distances to that obtained by subsampling a variable number of trials from the iterative experiment. Here we subsampled the first half of the data, and compared the results to the empirical prior derived from the second half. To estimate the time required in each case we took into account the total duration of one trial (13 s for the discrimination paradigm and 60 s for running 5 iterations of the iterative paradigm).

In both experiments the JSD accuracy increased with the total duration of data collection, namely the total number of participants multiplied by the session duration. However, the required data collection time was much larger for the discrimination paradigm compared with the iterative paradigm. Obtaining a JSD of 181%, 132% and 111% of the asymptotic accuracy (estimated as the JSD between distributions estimated from split-halves of all trials) was 21.1, 36.6, and 44.7 times faster in the iterative paradigm compared with the discrimination paradigm, respectively.

## Assessing the Benefit of Multiple Iterations

With enough data, in principle one ought to be able to infer the prior from the reproductions of just the first iteration of our iterative procedure, by fitting a model that best explains the reproductions. The difficulty is that this approach requires assuming a model by which reproductions are generated, and a parametric form for the prior, the parameters of which (along with parameters for perceptual and production noise) must be optimized to maximize the likelihood of the data. In many domains it may not be obvious how to model the prior or the reproduction (we adopted a mixture of Gaussians only after having observed the results of iterated reproduction). We nonetheless investigated how well we could estimate the prior from first-iteration data if we were to assume a mixture of Gaussians parametric form.

We split all the trials from experiment 1 into two halves, using one half to estimate a "ground truth" prior from the fifth iteration (the "test" data), and the other half ("training" data) to infer the prior from either the first or fifth iteration. We evaluated the accuracy of the inferred prior as the JSD between the ground truth prior estimate and the prior inferred from the first or fifth iteration. We performed this evaluation for multiple splits of the data and for varying amounts of data (i.e. using different fractions of the second half of the data to estimate the prior). When estimating the prior from the fifth iteration, we used kernel density estimation, on the assumption that the data distribution is converged to the prior by this point. When estimating the prior from the first iteration, we modeled the distribution of responses assuming Gaussian sensory noise, Gaussian reproduction noise, and a mixture of Gaussians prior, and estimated the model parameters that minimized the JSD between the model response distribution and the observed response distribution (the latter was estimated from the training data via kernel density estimation). We then compared the resulting prior to the ground truth prior estimated from the fifth iteration of the other half of the data. Because this procedure is computationally intensive, we evaluated only a small patch of the prior, containing two modes.

We found that the accuracy of the inferred prior increased with the amount of data for both methods (using the first and fifth iteration, respectively), but that the asymptotic accuracy was worse (by a factor of 2.21) when using the first iteration. We suspect this is because some aspect of the reproduction model is not fully accurate (potentially the way that reproductions are generated from the posterior —here we assume they are samples from

the posterior). This underscores an advantage of the iterative method, in that the prior can be read more directly from the data, with fewer required assumptions.

## SUPPLEMENTAL REFERENCES

S1.    Repp, B.H. (2005). Sensorimotor synchronization: a review of the tapping literature. Psychon. Bull. Rev. *12*, 969–992.

S2.    Repp, B.H., and Su, Y.H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). Psychon. Bull. Rev. *20*, 403–452.

S3.    Friberg, A., and Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. J. Acoust. Soc. Am. *98*, 2524–2531.

S4.    London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. Music Percept. *19*, 529–550.

S5.    Sadakata, M., Desain, P., and Honing, H.J. (2006). The Bayesian way to relate rhythm perception and production. Music Percept. *23*, 269–288.

S6.    London, J. (2012). Hearing in Time: Psychological Aspects of Musical Meter (Oxford University Press).

S7.    Povel, D.J. (1981). Internal representation of simple temporal patterns. J. of Exp. Psychol. Hum. Percept. Perform., *7*, 3–18.

S8.    Desain, P., and Honing, H. (2003). The formation of rhythmic categories and metric priming. Perception *32*, 341–365.

S9.    Feller, W. (1959). An Introduction to Probability Theory and its Applications (John Wiley & Sons).

S10.   Xu, J., and Griffiths, T.L. (2010). A rational analysis of the effects of memory biases on serial reproduction. Cogn. Psychol. *60*, 107–126.

S11.   Griffiths, T.L., and Kalish, M.L. (2005). A Bayesian view of language evolution by iterated learning. In Proceedings of the 27th Annual Conference of the Cognitive Science Society  (Cognitive Science Society ), pp. 827–832.

S12.   Kronrod, Y., Coppess, E., and Feldman, N.H. (2016). A unified account of categorical effects in phonetic perception. Psychon. Bull. Rev. *23*, 1–32.

S13.   Feldman, N.H., Griffiths, T.L., and Morgan, J.L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. Psychol. Rev. *116*, 752–782.