Musical evolution in the lab exhibits rhythmic universals

Musical evolution in the lab exhibits rhythmic universals

Andrea Ravignani[1,2,3,*], Tania Delgado[1], Simon Kirby[1]

Author affiliation

[1] Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, UK

[2] Artificial Intelligence Lab, Vrije Universiteit Brussel, 1050 Brussels, Belgium

[3] Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, 1090 Vienna, Austria

Contact information

[*]To whom correspondence should be addressed: andrea.ravignani@gmail.com

Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Tel: 0049 15752407179

Keywords. human evolution, cultural transmission; rhythm perception; transmission chains; iterated learning; evolution of music; biology of music; rhythm production; music cognition.

Word count (excluding abstract, methods and captions): 2399 words

Word count methods: 1055

Number of references, main text: 29; **Methods-only references**: 20

Figure count: 4

Article type: **Letter**

Main text

Music, despite its variety across the world, exhibits some cross-cultural similarities. Evidence from a broad range of human cultures suggests the existence of musical universals [1], here defined as strong regularities emerging across cultures above chance. In particular, humans demonstrate a general proclivity for rhythm [2], though little is known about why music is particularly rhythmic, and why the same structural regularities are present in rhythms around the world. Here we empirically investigate the mechanisms underlying musical universals for rhythm, showing how music can evolve culturally from randomness. Human participants are asked to imitate sets of randomly generated drumming sequences, and their imitation attempts become the training set for the next participants in independent transmission chains. By perceiving and imitating drumming sequences from each other, participants turn initially random sequences into rhythmically structured patterns. Drumming patterns develop into rhythms which are more structured, easier to learn, distinctive for each experimental cultural tradition and characterized by all six statistical universals found among world music [1], appearing adapted to human learning, memory, and cognition. We conclude that musical rhythm partially arises from the influence of human cognitive and biological biases on the process of cultural evolution [3].

Percussion instruments may have provided the first form of musical expression in human evolution. Great apes - our closest living relatives - show drumming behaviour [4], which they can learn socially [5], producing some human-like rhythmic sequences [6]. Hence, percussive behaviour may have already been present in our ancestors some million years ago before the split between human and Pan lineages [2]. Archaeological findings also suggest that the first human musical instrument might have been percussive, as also attested in modern hunter-gatherer societies around the world [7]. This makes rhythm a particularly apt musical dimension to reconstruct crucial steps in the evolution of music.

Six rhythmic features can be considered human universals, showing a greater than chance frequency overall and in all geographic regions of the world. These statistical universals are:

- a regularly-spaced (isochronous) underlying beat, akin to an implicit metronome;
- hierarchical organization of beats of unequal strength, so that some events in time are marked with respect to others;
- grouping of beats in 2 (e.g. marches) or 3 (e.g. waltzes);
- with a preference for binary (2-beat) groupings;
- clustering of beat durations around few values distributed in less than five durational categories;
- use of durations from different categories to construct riffs, i.e. rhythmic motifs.

Until now, research on musical universals has focused either on individual psychological processes [8], investigating rhythm perception/production in meticulously controlled environments [9,10], or large-scale phenomena, performing cross-cultural analyses of world musical traditions [11,12]. Combining these approaches, here we show that basic psychological mechanisms (working memory, perceptual primitives, categorical perception, etc.) can lead to large-scale musical universals via cultural transmission. Our experiment aims at reconstructing in the lab (Figure 1a) how initially unstructured sounds might have been shaped into complex musical systems by early humans perceiving and imitating them [7,12,13]. We test experimentally controlled human micro-societies and show that indeed cultural transmission accounts for the emergence of both structural regularities and all predicted rhythmic universals. Our method builds on previous experimental methodologies, which showed how systematic structure may emerge from weak learning biases [14].

Similarly to the *vertical transmission* shaping the complexity and variety of musical cultures [3,12], in our experiment each participant hears and has to imitate drumming patterns received from a previous participant, who himself has copied them from someone else and thereby potentially introduced errors. In measuring the changes that occur to the drum patterns, we can observe how cognitive biases for rhythm are magnified and mirrored in musical structure, and how initially independently reproduced sequences come to pattern together as part of an overall rhythmical system [15]. As predicted, after several experimental generations, initially random sequences transform into increasingly structured and learnable music-like patterns. In addition, these patterns show convergence towards all the six rhythmic universals found in human musical cultures [1].

First, sequences acquire systematic structure. Systematicity is a measure of mutual predictability among the elements of a system, quantifying how much structural information about a whole system is provided by each constituent element. In musical harmony for instance, rock-n-roll is very systematic, because knowing a musical excerpt provides a better than chance guess on chord progressions of a broad range of songs, while dodecaphonic music is less systematic. Here we find an increase in structural similarities and combinatorial structure over generations (Page's trend test; $L=1558.0$, $m=6$, $n=9$, $p<0.001$; Figure 1b)

Second, sequences become easier to learn. A system or structure is highly learnable if it can be rapidly acquired with low error by an organism. Reproduction errors (time distance between participants' output) decrease over generations (Page trend test; $L=833$, $m=6$, $n=8$, $p<0.0001$; Figure 1c). Learners in later generations found the rhythms easier to imitate accurately, indicating that patterns increasingly fit participants' cognitive biases.

Third, timing patterns converge to durational categories. The frequency distributions of inter-onset intervals (IOIs i.e. time between consecutive drum hits) of all chains show a similar pattern across experimental generations: Initial uniform distributions (the random patterns presented to the first generation) converge on chain-specific clusters of IOIs by the final generation (Figure 2). A K-means cluster algorithm shows that rhythmic patterns converge to 3 durational categories (Table 1S in supplement), matching the statistical universal across world musical cultures which predicts less than five categories [1,11]. The range of durations produced by our participants is consistent with musical rhythms, as used in rhythm experiments [9]. The first cluster in all chains has a median of 203 ms (Table S1), close to 200 ms, a recurrent durational value in musical rhythm and meter [16]. Moreover, the resulting clusters' centroids are related by ratios close to integer ratios (Table S1).

Fourth, the increase in systematicity and learnability maps to the emergence of repeating structures (phase-space plots of IOIs in Figure 3a). Specifically: (a) rhythmic patterns acquire motivic structure, another musical universal [1], i.e. rhythmic "riffs" emerge corresponding to polygons in phase-space coordinates, where the number of vertices equals the length of the repeating riff within a pattern [17]; (b) riffs are used multiple times by each participant across separate drum patterns, shown by similar polygons overlapping in one state-space plot; (c) motivic patterns evolve gradually as they are passed from earlier to later generations (Figure 3a, similar polygons in different plots of one chain); (d) riffs partly differ between chains (different polygons in different chains).

Fifth, sequences become more metronomic (isochronous), hierarchically structured (metrical), and composed by durations related by small-integer ratios. Isochrony and meter in perceived music are usually probed by asking participants to tap along, testing whether their taps occur at simple multiples or at divisors of the occurring musical intervals. As our task involves musical production, we reversed the above logic: participants creating a metrical grid with binary and ternary subdivisions and an underlying regular beat [18] would produce: (a) adjacent IOIs related by small integer ratios, (b) with many values close to 1:1 (equal-length IOIs), (c) or ratios of 2 and 3 (showing binary and ternary subdivisions) [18], and, (d) strongest beats occurring at IOIs twice or three times multiple of each other, suggesting musical meter. We find that distributions of ratios in the last generation (Figure 4a-b) significantly differ from a simulated uniform ratio distribution, predicted under null hypothesis of no pairwise structure between IOIs (2-sample Kolmogorov Smirnov test; all $D>0.08$, all $p<0.01$, see SI). This holds for both distributions of adjacent IOIs and of IOIs between high-intensity hits, suggesting the existence of structural relationship between IOIs. We then tested whether peaks in the ratio distributions (Figures 4a-b) correspond to specific constant relations between IOIs (see Methods). The highest peak in Figure 4c occurs at 1.015, and the median of the distribution is .968. Both values are close to 1:1, providing moderate evidence for isochrony, another universal. We then test whether the highest peaks in Figures 4a-b coincide beyond chance with those expected theoretically in actual music. For adjacent ratios, we find four peaks, namely at: 1:2, 1:4, 3:2 and 3:4. The match between ratios expected in music and experimental ratios is not attributable to chance. (The corresponding Jaccard index, measuring overlap [19], is J=0.222. A randomization test returned an average Jaccard's index J=.064, pseudo p-value: p'=.029, see Methods.) A similar analysis on the distribution of ratios of IOIs between strong beats (median=0.947), found support for the hypothesis that meter is exclusively binary (J=.028, p'=.045), with strong and weak beats alternating, but not exclusively ternary (J=.028, p'=1.0). Strong beats occur above chance in intervals that are half or double each other in length (i.e. related by 1:2 and 2:1 ratios). Notes of ternary length exist, but do not always coincide with the metrical grid (e.g. a binary meter with many notes of length 1/4 and 3/4). This suggests the presence of (a) an underlying regular beat, which is (b) composed of alternating weak-strong beats, and (c) used as a reference duration to generate other notes' duration (by multiplying and dividing it by 2 or 3), providing evidence for the remaining universals.

Sixth, chains evolve independently. We calculated the Kolmogorov-Smirnov D statistic for each generation and pairs of participants using their distribution of IOIs to quantify the degree of cultural divergence. Chains significantly diverge over generations towards separate lineages with different timing structure (L=1586.0, m=6, n=9, $p<0.001$; Figure 3b). Moreover, all IOIs distributions of the final generations are significantly different between chain pairs (Kolmogorov-Smirnov tests, all $D<0.3$, all p-values <0.01, Table S2 in supplement). Hence, the drum patterns within the same lineage participate in a *system* of rhythmic patterns sharing similar characteristics or motifs. As in actual music [12], chains gain more structure over generations, though each transmission chain develops its own set of structural features.

It has been debated whether some human biological traits evolved under selective pressures to specifically hear and perform music [2,7,20-22]. Our data supports an alternative hypothesis: musical structure appears to evolve out of, and get shaped by, more general constraints on learning and memory. In this experiment, rhythmic features evolve cumulatively and gradually from randomness. We obtain divergent musical cultures, where each "musical culture", corresponding to an experimental chain, constitutes a system by itself. The transmission process we re-created in the lab leads to the appearance of design: the patterns evolve in such a way that they appear well adapted to the challenge of being learnable. Generation after generation, learners introduce errors in their efforts to replicate the sequences they hear. The process eventually results in the emergence of rhythmic patterns that are easier to reproduce. Systematic similarities between patterns emerge within a chain: Patterns that no longer act independently may facilitate learning over generations, as it is easier to remember a small number of motifs rather than thirty-two totally independent patterns. Participants were chosen to be non-musicians, so no previous skills in music performance can account for the quick generation of musical patterns we observe. They were instructed to recreate each sequence as closely as possible, neither to innovate, nor to treat the sequences as being related. Crucially, as in human music, our laboratory experiment leads to emergence of commonalities, but also diversity. This experiment provides evidence for the universality of musical features emerging through cultural transmission [1,3].

Similarly to previous results on the evolution of linguistic structure [15,23,24], we hypothesize that a few perceptual, learning and production biases may be responsible for the regularities evolving in our drumming patterns. Formation of durational categories and small integer ratios between intervals might be partially amenable to categorical perception of rhythmic sequences. In fact, small ratios function as attractors when musicians are asked to categorize notes of varying durations not related by integer ratios [25]. The proximity, although not equality, to integer ratios dovetails with previous findings in music psychology [26]. Emergence of few durational categories and motifs may instead be a by-product of the human tendency to compress sensory stimuli, possibly dictated by working memory constraints and limited capacity for processing information [27]. Conversely, motor biases seem to only moderately influence the structures obtained: humans' preferred tapping rate of 600 msec [17] is rarely found in our IOI distributions and clusters (Table 1S and Figure 2). However, our experiment cannot disentangle which human biases generating musical features are basic and which are acquired, and at least two alternative hypotheses can account for our results. In other words, the fact that our participants have already been exposed to a musical culture may be shaping the results. Two points speak against this interpretation, however. First we see clear divergence between chains, suggesting that there is no single culturally acquired attractor that is driving the evolution of the systems. Secondly, there are striking parallels in the evolution of systematic structure between this experiment and another sequence learning experiment in the non-musical domain [15]. Ultimately, cross-cultural replications of this experiment will be needed to accurately gauge the influence of acquired biases in this task.

Music, language, and dance all involve copying to some extent, though imitation/copying is only one of many factors in their evolution [3,21,22,29]. Although the motivations to copy are likely to differ, the outcomes seem to be similar. We believe the assumption that early humans might have had a motivation to copy music-like sequences is quite realistic. Several hypotheses on the origins of the biological capacity for musical rhythm involve some motivation to copy or imitate. These hypotheses often suggest imitation, learning or synchronization of audio-motor behaviours as a necessary step to achieve inter-individual coordination, group cohesiveness, mating success or territorial defence, providing in turn evolutionary pressures on the development of modern humans' rhythmic abilities [20].

Human music is inherently structured, showing a few structural similarities across musical cultures and traditions. Why do these similarities arise? How do different musical traditions end up with similar features? We have addressed these questions empirically: In the laboratory, we set the conditions for random percussion patterns to be transmitted, similarly to real musical traditions. As a result we are able to witness the evolution of musical rhythmic structure in real time as it responds to human constraints and converges towards all six statistical universals found among world rhythms. Musical rhythmic universals arise because human behaviour and cognition slightly transform what is copied [13,23,28,29]. These transformations, amplified by the process of cultural transmission, lead to diverse musical traditions, containing nonetheless a few universals: traces of the biology of the organisms who created them.

Methods

Participants. Forty-eight participants (mean age 23y, 4m; females = 37) were recruited from the University of Edinburgh's graduate employment service "to participate in a 30-minute drumming experiment". Each received £5 for participating. Musicians (having formal musical training or regularly practiced a musical instrument) were excluded from participation. Sample size was established a priori based on a meta-analysis of previous iterated learning experiments [14,24,30-35].

This experiment is modelled on a simple transmission chain paradigm, in which learners receive training input from the output of the previous learners [36]. Participants were randomly assigned to six different lineages (transmission chains: C1,C2,...,C6), each containing eight "generations" of learners (1,2,...,8). The first generation of participants in all chains heard different randomly generated patterns as training input (first column in Figure 2 and Figure 3a).

Stimuli. Participants in each generation were presented with 32 drum patterns. These patterns were random drum sequences (Random/Generation 0) or sequences produced by a previous participant (Generation 1-8). The 32 initial and independent drum patterns were each composed of 12 MIDI snare drum hits (see supplementary). Each chain had its own unique set of 32 initial random patterns. Each snare drum hit in the initial sequences had random velocity (force and speed used to play an instrument) and IOI (duration between the start of one note and the start of the next note). An additional cymbal sound, always presented 1.5 seconds after the last snare hit, signalled the end of a sequence. The cymbal timing was neither counted as part of the pattern nor included in the analyses. Participants heard and reproduced two blocks of the same thirty-two drum patterns, with the order of drum patterns within each block randomized. The first block of patterns was intended for the participant to practice drumming and copying. Patterns reproduced in the second block, recorded on a laptop, were used as training stimuli for the next learner in the chain.

Procedure. Participants were given headphones, a single drumstick and an Alesis SamplePad, connected to a Macbook Pro laptop via a Duo-Capture EX USB-MIDI interface. The Python code recording drumming patterns rounded temporal information to the nearest millisecond (although the theoretical maximum resolution of MIDI is slightly better than 1 ms). The interface had four independent drum pads: Three produced the snare drum sound, while one produced the cymbal sound participants struck to conclude a pattern. Participants were instructed to reproduce each pattern immediately after hearing it to the best of their ability. Each sequence was recorded and given to the next participant in the chain.

Participants were unaware that they would be listening to stimuli produced by a previous learner. After the behavioural task, participants completed a questionnaire (see supplement).

Analyses. The output patterns were analysed to determine if they would evolve to become easier to learn over generations and if the initially independent sequences evolve in such a way to form rhythmic-like systems with structural regularities. Data analysis was performed in R, Stata 11.0 and using custom-written Python scripts. All analyses were performed on the inter-onset intervals (IOIs) between contiguous drum hits within a pattern. In fact, experiments in human perception of musical rhythms have shown that the IOI is usually more important than the length of the notes themselves [37]. Several quantitative measures were adapted in order to assess the learnability and structure of the patterns [24,27,38-46].

Ratios were taken to normalize with respect to tempo and to compare structures (rather than absolute durations) across patterns. For each ratio distribution, we found the location of the maxima by taking the 2nd derivative of the KDE function. We then tested whether these fixed IOIs relations (i.e. the peaks in Figure 2) coincide beyond chance with those expected theoretically. The most parsimonious way of generating a musical duration from another is to multiply or divide it by 2, 3 or 4. Hence we predicted to find with high frequency ratios of 1:1 (equal duration IOIs), 1:2, 1:3, 1:4, 2:3, 3:4, and their reciprocals, giving a total of 11 expected theoretical ratios. As the predicted ratios spanned 11 possible values, we extracted the 11 most frequent ratios from our empirical distributions. We then matched expected with empirical ratios (with a 0.01 tolerance on ratio differences), and quantify the match using the Jaccard index, a measure of the overlap between two sets [19]. Given two sets, the Jaccard index is calculated as the ratio between their union and their intersection, i.e. the number of elements in common divided by the number of overall elements. Finally, we performed a Monte Carlo simulation with 1 million iterations to test whether the matching of predicted and found peaks was attributable to chance. This provided a pseudo p-value, calculated as the relative number of randomizations with an average Jaccard index greater than or equal to the empirical Jaccard index, i.e. the relative number of cases for which a list of 11 random ratios has equal number or more matches with predicted ratios than the 11 empirical ratios.

Increase in structure/systematicity measure G. Unlike previous cultural transmission research, the transmitted behaviour in this experiment is continuous (i.e. time intervals) rather than discrete. We discretized the intervals into 3 categories using a K-means clustering algorithm (Table S1), mapping each duration to the tercile it belonged to (e.g. three durations like {0.1, 0.8, 0.4} would map to {short, long, medium}). The number of categories in the K-means algorithm was established using the 'Elbow' method [47], with 3 categories emerging as the most parsimonious clustering for each chain (see Supplement). We then calculated a grammatical structure index G (a modified measure for entropy comparable with previous studies [41]) for each participant.

Decrease in imitation errors E, equivalent to an increase in learnability/imitation fidelity. This is calculated as the (edit) time distance between two drum patterns: the total cost of the minimal cost set of substitutions, insertions, or deletions among IOIs necessary to transform the pattern of durations a participant has heard into the pattern she has reproduced, where edit costs are taken to be the absolute difference in time between durations [43]. The time distance between identical patterns equals zero. Notice that, unlike other metrics in musicology assuming beat induction or metrical hierarchies [48,49], this edit distance minimizes assumptions about metrical, top-down processing.

Data Availability. The data that support the findings of this study are available for download as supplementary material.

Competing financial interests. The authors declare no competing financial interests.

Correspondence & Materials Requests. Correspondence and requests for materials should be addressed to Andrea Ravignani (andrea.ravignani@gmail.com)

Acknowledgements. AR was supported by FWO grant V439315N and KWA Uni Wien grant 000245, and ERC grants 283435 ABACUS (to Bart de Boer) and 230604 SOMACCA (to W. Tecumseh Fitch). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank P. Filippi, B. Thompson, B. de Boer, H. Little, S. van der Ham, N. Chr. Hansen, J. Iversen, D. Bowling, T. Grossi, A.C. Miralles, P. Norton, V. Spinosa, Y-H. Su, P. Tinits, K. Smith, all members of the Centre for Language Evolution (Edinburgh), AI-Lab (VUB Brussels), Biolinguistics (Barcelona), and attendants of Evolang XI, IBAC XXV, Statistical Learning 2015, DZG Graduate Meeting 2015, for comments and advice.

Author contributions. A.R. and S.K. conceived the study, A.R., T.D. and S.K. designed research, T.D. performed research, A.R. and S.K. wrote Python scripts for data analysis and experimental testing, A.R., T.D. and S.K. analysed data and wrote the paper.

References

1 Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences* **112**, 8987-8992 (2015).

2 Fitch, W. T. in *Language and Music as Cognitive Systems* (Oxford University Press. Oxford, UK, 2009).

3 Trehub, S. E. Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences* **112**, 8809-8810 (2015).

4 Ravignani, A. *et al.* Primate Drum Kit: A System for Studying Acoustic Pattern Production by Non-Human Primates Using Acceleration and Strain Sensors. *Sensors* **13**, 9790-9820 (2013).

5 Fuhrmann, D., Ravignani, A., Marshall-Pescini, S. & Whiten, A. Synchrony and motor mimicking in chimpanzee observational learning. *Scientific Reports* **4** (2014).

6 Dufour, V., Poulin, N., Curé, C. & Sterck, E. H. Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Scientific Reports* **5** (2015).

7 Morley, I. *The evolutionary origins and archaeology of music*. Darwin College, Cambridge University Cambridge, (2003).

8 Fritz, T. *et al.* Universal recognition of three basic emotions in music. *Current Biology* **19**, 573-576 (2009).

9 Repp, B. H., Iversen, J. R. & Patel, A. D. Tracking an imposed beat within a metrical grid. *Music Perception* **26**, 1-18 (2008).

10 Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences* **106**, 2468-2471 (2009).

11 Brown, S. & Jordania, J. Universals in the world's musics. *Psychology of Music* **41**, 229-248 (2013).

12 Le Bomin, S., Lecointre, G. & Heyer, E. The Evolution of Musical Diversity: The Key Role of Vertical Transmission. *PloS one* **11**, e0151570 (2016).

13 Morgan, T. *et al.* Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature communications* **6** (2015).

14 Kirby, S., Dowman, M. & Griffiths, T. L. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* **104**, 5241-5245 (2007).

15 Cornish, H., Smith, K. & Kirby, S. Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. 340-345.

16 London, J. *Hearing in time: Psychological aspects of musical meter*. (Oxford University Press, 2012).

17 Ravignani, A. Visualizing and interpreting rhythmic patterns using phase space plots. *Music Perception* (in review).

18 Celma-Miralles, A., de Menezes, R. F. & Toro, J. M. Look at the beat, feel the meter: Top-down effects of meter induction on auditory and visual modalities. *Frontiers in Human Neuroscience* **10** (2016).

19 Jaccard, P. The distribution of the flora in the alpine zone. *New phytologist* **11**, 37-50 (1912).

20 Iversen, J. R. in *The Cambridge Companion to Percussion* (ed Russel Hartenberger) (Cambridge University Press, 2016).

21 Laland, K., Wilkins, C. & Clayton, N. The evolution of dance. *Current Biology* **26**, R5-R9 (2016).

22 Ravignani, A. & Cook, P. The evolutionary biology of dance, without frills. *Current Biology* (in press).

23 Kirby, S., Cornish, H. & Smith, K. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* **105**, 10681-10686 (2008).

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

25 Desain, P. & Honing, H. The formation of rhythmic categories and metric priming. *Perception* **32**, 341-365 (2002).

26 Repp, B. H., London, J. & Keller, P. E. Systematic distortions in musicians: reproduction of cyclic three-interval rhythms. *Music Perception* **30**, 291-305 (2013).

27 Kirby, S., Tamariz, M., Cornish, H. & Smith, K. Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141**, 87-102 (2015).

28 Kirby, S. & Hurford, J. R. in *Simulating the evolution of language* 121-147 (Springer, 2002).

29 Tamariz, M. & Kirby, S. Culture: copying, compression, and conventionality. *Cognitive science* **39**, 171-183 (2015).

30 Smith, K. & Wonnacott, E. Eliminating unpredictable variation through iterated learning. *Cognition* **116**, 444-449 (2010).

31 Kalish, M. L., Griffiths, T. L. & Lewandowsky, S. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic bulletin & review* **14**, 288-294 (2007).

32 Verhoef, T. The origins of duality of patterning in artificial whistled languages. *Language and Cognition* **4**, 357–380 (2012).

33 Garrod, S., Fay, N., Rogers, S., Walker, B. & Swoboda, N. Can iterated learning explain the emergence of graphical symbols? *Interaction Studies* **11**, 33-50 (2010).

34 Kempe, V., Gauvrit, N. & Forsyth, D. Structure emerges faster during cultural transmission in children than in adults. *Cognition* **136**, 247-254 (2015).

35 Silvey, C., Kirby, S. & Smith, K. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science* **39**, 212-226 (2015).

36 Mesoudi, A. & Whiten, A. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3489-3501 (2008).

37 Parncutt, R. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 409-464 (1994).

38 Boon, J. P., & Decroly, O. Dynamical systems theory for music dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(3), 501-508 (1995).

39 Caldwell, C. A. & Millen, A. E. Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3529-3539 (2008).

40 Clayton, M. R. Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, *11*(1), 27-59 (2007).

41 Jamieson, R. K. & Mewhort, D. Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *The Quarterly Journal of Experimental Psychology* **62**, 550-575 (2009).

42 Kirby, S., Griffiths, T. & Smith, K. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* **28**, 108-114 (2014).

43 Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals in *Soviet physics doklady*. 707-710.

44 Matthews, C., Roberts, G. & Caldwell, C. A. Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior* **33**, 759-770 (2012).

45 Ravignani, A. & Sonnweber, R. Measuring teaching through hormones and time series analysis: Towards a comparative framework. *Behavioral and Brain Sciences* **38**, 40-41 (2015).

46 Ravignani, A. & Norton, P. Measuring the evolution of rhythmic complexity: A statistical primer to quantify and compare temporal structure in human phonology and animal vocalizations. *Journal of Language Evolution* (under review).

47 Thorndike, R. L. Who belongs in the family? *Psychometrika* **18**, 267-276 (1953).

48 Longuet-Higgins, H. C. & Lee, C. S. The perception of musical rhythms. *Perception* **11**, 115-128 (1982).

49 Longuet-Higgins, H. C. & Lee, C. S. The rhythmic interpretation of monophonic music. *Music Perception*, 424-441 (1984).

Figure Legends

Figure 1: Cultural transmission over generations by iterated learning. Iterated learning refers to a process by which the individual learns a new behaviour by observing another individual who acquired the behaviour in the same way [28]. This method directly taps into the dynamics of cultural transmission, thereby enabling an empirical approach to human cultural evolution [36]. Iterated learning of artificial sounds [24], visual representations [29] and language-like systems [23,28] can lead to a large range of outcomes. However, two characteristics seem to emerge in most experiments: random patterns evolve into sequences which exhibit increasing learnability and structure over generations of learners [23,27]. (a) The first two transmission steps in a chain of drummers. We generated sequences of drumming patterns with random velocity (hit strength) and time between hits. These random sequences (Random generation 0, leftmost note sequences) sound completely arrhythmic and mimic incidental occurrences of sound sequences, either naturally produced or human-generated, that an early music-less hominid might have attended to. We then present 32 of these random sequences to an experimental participant (Generation 1), who is asked to faithfully copy the rhythm on a drum set immediately after each of the 32 presentations. The sequences thereby produced, with all their copying errors, form the set of drum patterns presented to the next participant in random order (Generation 2). This process is repeated until the chain reaches 8 participants (rightmost rhythmic patterns). Also, to control for the effects of the initial random patterns or particular participants in a chain, the experiment is repeated in 6 independent chains (not shown), totalling 48 participants. (b) Increase in structure/systematicity measure G, corresponding to a modified measure for entropy (c) Decrease in imitation errors E, equivalent to an increase in

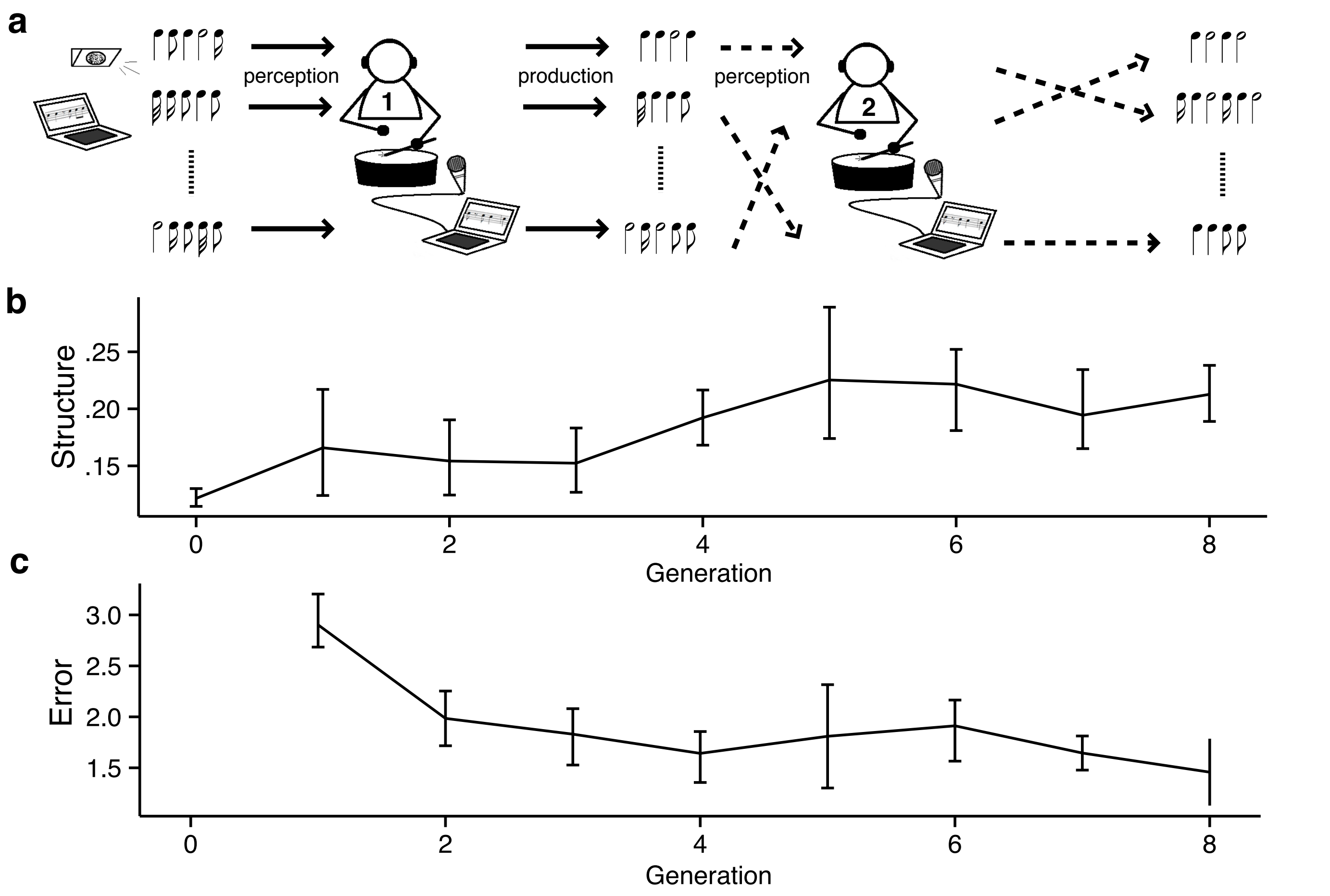
learnability/imitation fidelity, calculated as the (edit) time distance between two drum patterns. Error-bars represent bootstrapped 95% confidence intervals across chains.

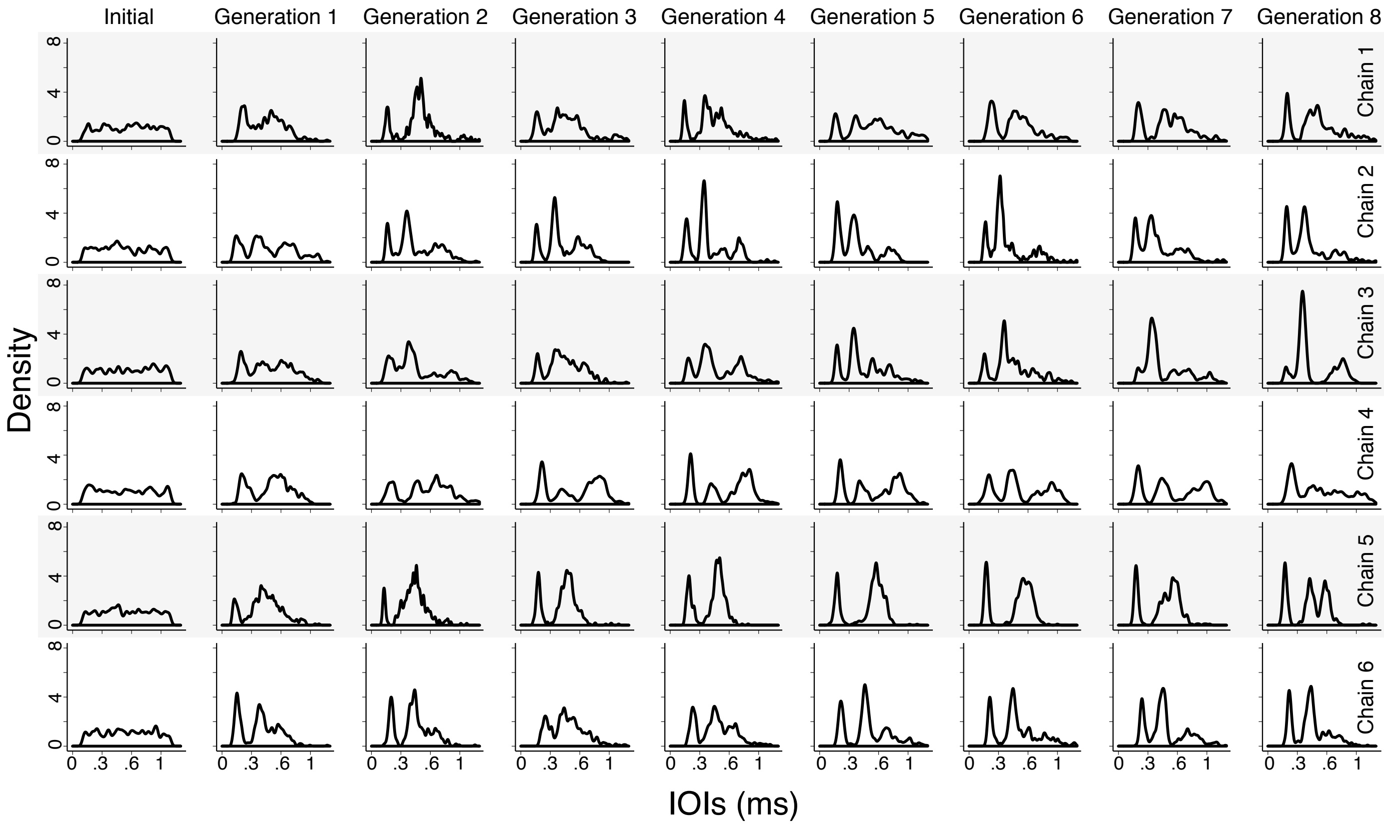
Figure 2: Frequency distributions of inter-onset intervals (IOIs) in drumming sequences for each chain (rows) and generation (columns). IOIs are pooled across all 32 patterns and plotted using Kernel density estimates (KDEs). IOI distributions reflect the timings between the start of one drum hit and the start of the next drum hit played by the participant. Random generation 0 (leftmost column) corresponds to six uniform distributions of the randomly generated patterns: the chains did not start with any structural patterning with respect to time. Over the course of generations there is a gradual development of interval durations, becoming more categorical (corresponding to more peaked distributions) towards the last generation (rightmost column). From left to right, the figure shows how each chain slowly converges towards a different distribution of IOIs from the other chains upon the final generation. Centroids for each last generation's distribution can be found in supplement. Extreme data points, corresponding to values >1.1s, and representing <5% of all data, are not shown.

Figure 3: Emergence of rhythmic riffs and cultural specificity. (a) State-space diagrams of all chains (top to bottom) and generations (left to right). Each state-space diagram depicts one participant's output (all 32 patterns). In a state-space diagram, the duration of each note (x-axis) is plotted against the duration of the next note (y-axis) as a dot; consecutive dots are joined by a line. This is repeated for all 32 patterns. Each state-space diagram depicts one participant's output. The state-space plots here show the evolution of patterns of length≥3, with increasing regularities over generations [17]. Closed polygons represent repeating drumming patterns. For instance, chain 5 shows a clear emergence (already by the third generation) of a repeating pattern of length 3, illustrated by a triangle. Chain 3 converges instead towards drumming patterns containing a combination of two similar ternary patterns, inferred by the two non-overlapping triangles. Chain 2 converges towards patterns including a

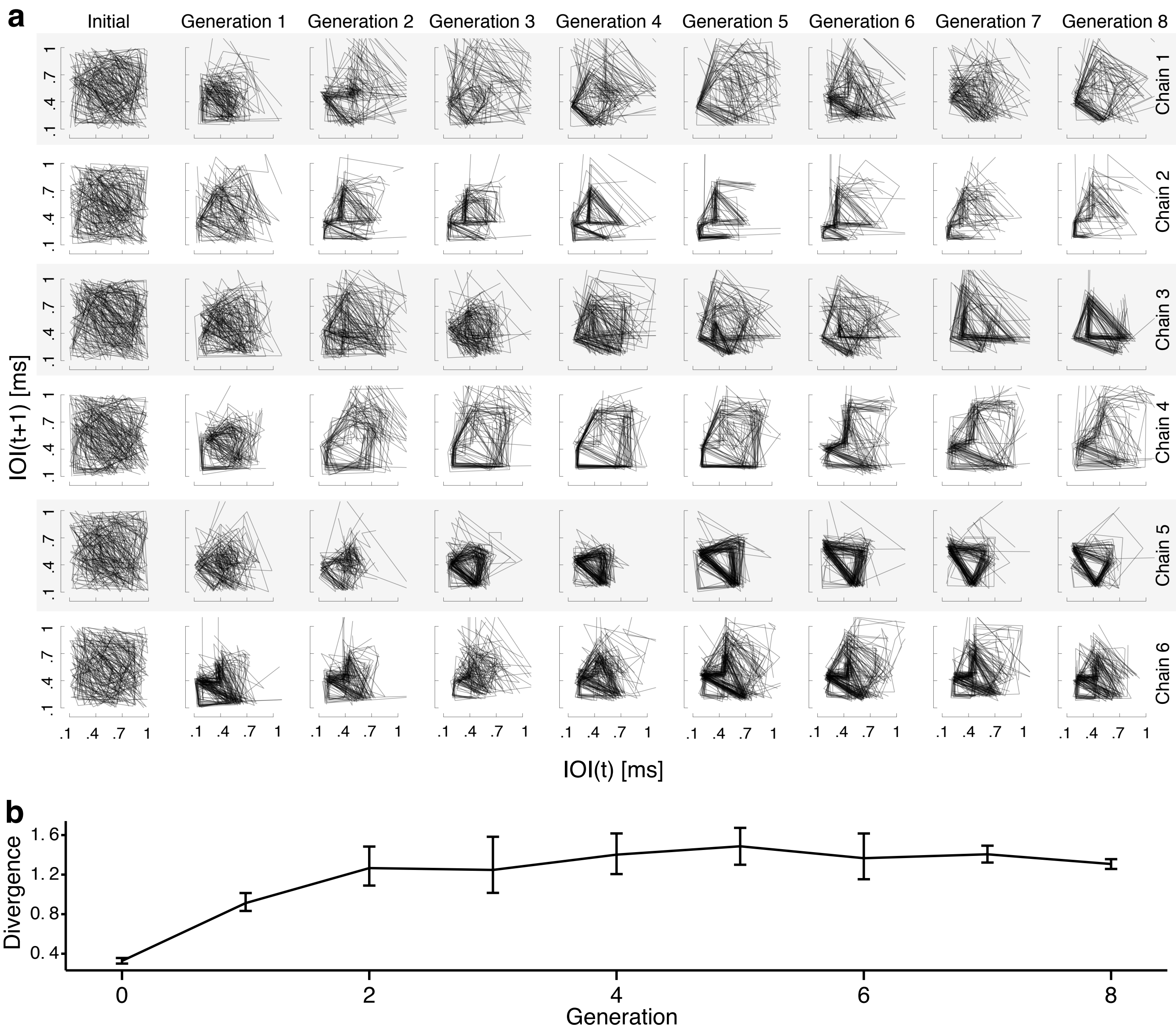
non-repeating sequence of length 5, deduced by the 5-edged segmented line. Also, notice how the vertices of the polygons map to the centroids found with the K-means clustering algorithm (in supplement): e.g. chain 5's centroids are at 177, 436 and 665 msec, while chain 3 centroids' are at 202, 355, and 764. These approximate values can be found when examining and comparing last generations' phase state plots of chains 3 and 5. (b) Rise of divergence measure D across chains. Over generations, variability between chains increases. This, together with the increase in G, suggests that a distinct yet systematic musical "culture" emerges in each chain.

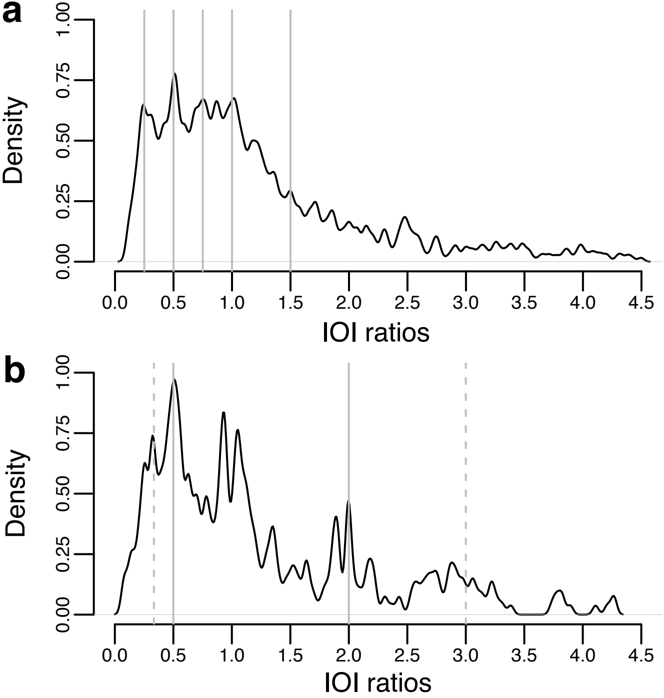
Figure 4: Statistical universals in durational patterns. (a) Frequency distributions of *ratios* between adjacent IOIs pooled across all last-generation participants, calculated using Kernel density estimates (KDEs). For each pattern, we calculated the ratios between all adjacent IOIs in that pattern (INI_1/INI_2, INI_2/INI_3,...). Here, we show the pooled frequency distributions across all 32 patterns produced by the 6 participants in the last generation. The distribution shows peaks (local maxima) centred at 1:2, 1:1, 1:3, 2:3 and 5:2 (solid lines); (b) Frequency distribution of ratios of durations between the 50% strongest beats (all drum hits above the median hit strength) within a pattern, pooled across all last-generation participants. The solid lines represent the 1:2 and 2:1 (binary) ratios, while the dashed lines represent the 1:3 and 3:1 (ternary) ratios. While several hypothesized ratios emerge as peaks in the distribution (e.g. 1:2 and 2:1), there are also peaks that do not map to precise integer ratios, attributable to a number of potential factors (cultural, experimental, etc.). Extreme data points, corresponding to values >4.5, and representing <5% of all data, are not shown.


a
perception
production
perception
1
2
b
Structure
.25
.20
.15
0
2
4
6
8
Generation
c
Error
3.0
2.5
2.0
1.5
0
2
4
6
8
Generation



Initial
Generation 1
Generation 2
Generation 3
Generation 4
Generation 5
Generation 6
Generation 7
Generation 8
Chain 1
Chain 2
Chain 3
Chain 4
Chain 5
Chain 6
Density
0
4
8
0
.3
.6
1
IOIs (ms)


a
Initial
Generation 1
Generation 2
Generation 3
Generation 4
Generation 5
Generation 6
Generation 7
Generation 8
Chain 1
Chain 2
Chain 3
Chain 4
Chain 5
Chain 6
IOI(t+1) [ms]
IOI(t) [ms]
.1
.4
.7
1
b
Divergence
Generation
0.4
0.8
1.2
1.6
0
2
4
6
8


a
Density
1.00
0.75
0.50
0.25
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
IOI ratios
b
Density
1.00
0.75
0.50
0.25
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
IOI ratios