ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Evolving building blocks of rhythm: how human cognition creates music via cultural transmission

Andrea Ravignani,[1,2,3,*] Bill Thompson,[1,2,*] Thomas Grossi,[4] Tania Delgado,[4,5] and Simon Kirby[4]

[1]Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium. [2]Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [3]Research Department, Sealcentre Pieterburen, Pieterburen, the Netherlands. [4]Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK. [5]Department of Cognitive Science, University of California, San Diego, La Jolla, California

Address for correspondence: Andrea Ravignani, Research Department, Sealcentre Pieterburen, Hoofdstraat 94a, 9968 AG Pieterburen, the Netherlands. andrea.ravignani@gmail.com

Why does musical rhythm have the structure it does? Musical rhythm, in all its cross-cultural diversity, exhibits commonalities across world cultures. Traditionally, music research has been split into two fields. Some scientists focused on musicality, namely the human biocognitive predispositions for music, with an emphasis on cross-cultural similarities. Other scholars investigated music, seen as a cultural product, focusing on the variation in world musical cultures. Recent experiments found deep connections between music and musicality, reconciling these opposing views. Here, we address the question of how individual cognitive biases affect the process of cultural evolution of music. Data from two experiments are analyzed using two complementary techniques. In the experiments, participants hear drumming patterns and imitate them. These patterns are then given to the same or another participant to imitate. The structure of these initially random patterns is tracked along experimental "generations." Frequentist statistics show how participants' biases are amplified by cultural transmission, making drumming patterns more structured. Structure is achieved faster in transmission within rather than between participants. A Bayesian model approximates the motif structures participants learned and created. Our data and models suggest that individual biases for musicality may shape the cultural transmission of musical rhythm.

Keywords: evolution of music; rhythm; timing; cultural transmission; biomusicology; Bayesian model

Introduction

The ability to produce music is a human behavior found universally in all human cultures. Though within the scientific study of music, there is a long-standing division between those who are concerned with music as a cultural product versus those who focus primarily on music and the mind.[1] Traditionally, music researchers distinguish *music* from *musicality*:[2] Music is a cultural product, while musicality is the human biological machinery used to produce and process music. At one extreme, cultural anthropologists and field researchers have historically focused on music as a cultural product.[3,4]

According to these scholars, only years of immersion in a musical culture could enable us to truly understand its uniqueness. However, embracing cultural uniqueness leaves little room for cross-cultural comparability. If instead we consider cross-cultural comparisons, illuminating patterns like the so-called *statistical universals* can emerge. These statistical universals are musical features that appear above chance across many or all musical cultures in the world.[5,6] At the other extreme, experimental psychologists and cognitive neuroscientists have traditionally focused on musicality by probing into the biocognitive substrates that enable music processing.[7,8] This neuropsychological approach assumes that these biocognitive substrates are common to all humans. As a result, data from a specific

---

population and culture are assumed to be generalizable to mankind at large. While exceptions exist,[9] the field of music research has largely been polarized into these two apparently segregated dimensions. These separate approaches to music research can be unified and revelatory, especially when asking evolutionary questions.[10,11]

How music changes over time seems to be orthogonal to the evolution of the biocognitive apparatus underlying music making. Recent experiments, however, have shown that music and musicality are intimately connected.[12–16] Focusing on rhythm, individual participants[14] were asked to imitate snare drum sequences on an electronic drum set to the best of their abilities. All participants were nonmusicians and were not told where the patterns came from. A "first generation" of participants was given computer-generated drum patterns, featuring random beat intensity and duration. Successive "generations," however, were asked to imitate the output of the previous participant in the experiment (Fig. 1A, top chain of participants). Unbeknown to each participant, patterns included all the errors and imperfections introduced in the previous generation. In this way, the process of cultural transmission of rhythmic patterns was recreated in the laboratory.[12,17,18] Over time, these sequences became more structured and easier to imitate. In addition, rhythmic sequences converged toward all rhythmic universal features that characterize music around the world.[6,19] This experiment, which aligns with similar findings in nonmusical domains,[20–25] suggests a link among cognition, biology, and culture in human music.[26,27] It shows that features present in almost all musical traditions around the world emerge through basic biocognitive biases and are thereby amplified by the process of cultural transmission. This experiment also raised a number of additional questions. In particular, are the biocognitive biases—responsible for the emergence of structure and universals—specific to (1) all humans, (2) adults, (3) Westerners, (4) a cognitive domain or modality, or (5) particular individuals?

We begin to tackle the question: how do individual biases affect the process of cultural evolution of music?[28] Two alternative hypotheses originally advanced for language can be readily adapted.[27,29] One *nativist hypothesis* sees human cognition as constrained to produce only certain forms of music and musical structure as designed by individual

minds and transmitted down generations. Another *interactive hypothesis* is that the structure in music results from long-term distributed effects of multiple, individually varying minds that together create a kind of structural compromise that is pleasing to and learnable by all.[17,18] These contrasting hypotheses have an exact parallel in the human language and animal communication literatures.[30–33] In the nativist scenario, if all individuals share strongly constraining biases, music would emerge as a result of strengthening few, human-wide tendencies toward musicality (akin to a vote by consensus). In the alternative interactive scenario, if all individuals have weak and potentially variable biases, music would emerge as a result of the interplay among individual-specific tendencies toward musicality, amplifying or averaging each other out (similarly to mixing paint of different colors). To summarize, we ask: do regularities in musical structure result (1) from humans imposing universal structural regularities, or (2) from distributed, weakly biased processing?[34]

We address this question using experimental manipulations and Bayesian modeling techniques. First, we replicate a previous experiment with one variation.[14] In the original experiment, a set of drumming patterns was transmitted across generations of participants in a between-participants design (top row of Fig. 1A). In contrast, the current experimental design features a within-participant structure (bottom two rows of Fig. 1A).[13,30,34,35] Each participant takes part in multiple rounds, instead of one round per participant, as in the previous experiment. Second, we analyze and compare the data from the two experiments using standard inferential statistics. Third, we introduce a probabilistic model for latent structures underpinning rhythmic sequences, alongside a psychologically plausible algorithm for inferring these structures. This allows us to obtain approximate structural descriptions of rhythmic patterns across conditions and generations and to explore how these structures are used and reused. Whereas previous models have focused on inferring cognitive biases for integer-ratio rhythmic categories from experimental data,[13] our model focuses on approximating the process through which individuals combine rhythmic categories into predictable motif-like sequences.

A general prediction is that the new within-participants "chains" will likely produce data
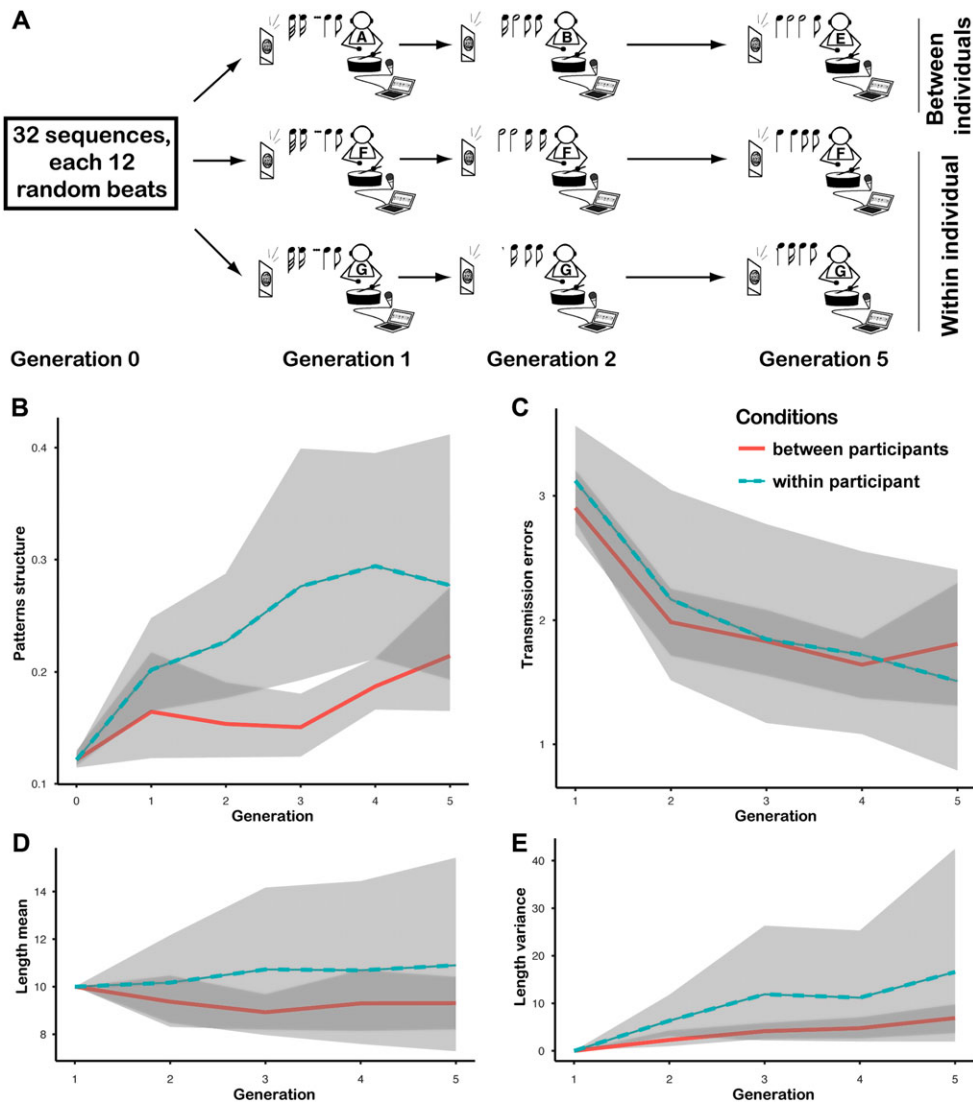
Figure 1. Experimental design and summary statistics tracking the evolution of patterns in the first six generations of both experiments. (A) The experimental design follows a transmission chain paradigm: the output of one experimental generation constitutes the input of the next generation. Generation 0 consists of computer-generated drumming patterns, where drum hits have random velocity and inter-onset intervals (IOIs). Each generation 0 pattern is individually heard and imitated by a generation 1 participant. The resulting imitated patterns can be given to the same participant to imitate once again (bottom two rows of A, within-individual design) or to a different participant (top row of A, between-individuals design). The procedure is repeated over generations (left to right) and additional chains (not shown). (B) Increase in structural complexity quantified over generations using a modified measure of entropy. (C) Decrease in imitation error between adjacent generations, quantified using a modified Levenshtein distance. (D) Pattern length across generations and chains. Length measured in the *r* space.[b] (E) Variance in pattern length across generations and chains. (B–E) Shaded areas depict bootstrapped 95% confidence intervals; lines connect the mean value for each generation and experiment type across chains.

qualitatively comparable to the original between-participant design. In particular, patterns should increase in structure and become easier to learn. If this holds, self-learning will be shown as an effective method for uncovering musical biases

in participants. *Quantitatively*, however, the two experimental designs might show different behaviors. In particular, if participants have individual-specific biases, they will likely impose idiosyncratic structure every generation. This, in turn, will make

structure emerge faster in the within-participant chains and slower in between-participants chains (i.e., some innovations will cancel out). If this is the case, we should observe a significant difference between the variables measuring the evolution of structure in the two experiments. If, instead, participants have homogeneous biases toward rhythmic structures, the evolution of chains will be unaffected by experimentally substituting many participants for one repeated participant. In this case, the two experimental designs will produce similar data, hence no measurable difference of key variables between conditions.

Experimental methods

Data from six experimental chains (30 participants) from a previous experiment were reanalyzed (details in Ref. 14). In brief, six different sets of 32 sequences of 12 random beats were given to first-generation participants to imitate. First-generation output became second-generation input, and so on (Fig. 1). In addition, new data were collected from 12 experimental chains (12 participants) in conditions comparable to the previous experiment,[14] with two key differences (for details, see Supplement, online only). First, each of the six different sets of 32 sequences of 12 random beats used in the first experiment was given to two first-generation participants in this new experiment (as opposed to one first-generation participant in the previous experiment, see Fig. 1). Second, the new experiment featured within-individual transmission, so that the same participant listened and imitated his/her own drum patterns over five experimental generations.[a]

Frequentist statistics: results and discussion

The metrics tracking structure and imitation error behave similarly over generations (Fig. 1B and 1C), confirming our qualitative prediction. Analyses of variance (ANOVAs) tested whether the generation and transmission type (i.e., between or within individual) could account for a possible increase in structure[14,36,37] (Fig. 1B) and decrease in imitation error[14,38] (Fig. 1C). Stepwise model selection suggested that both generation and transmission

type should be entered in the ANOVA as predictors of structure (minimizing Akaike information criterion). Both variables were significant predictors of structure (transmission type: $F = 7.4$, $P < 0.01$; generation: $F = 14.5$, $P < 0.001$). Another stepwise model selection suggested that only generation should be entered in the ANOVA as predictor of imitation error. Generation was a significant predictor of imitation error ($F = 14.8$, $P < 0.001$).

We used further ANOVAs to test that these differences were not due to superficial features. We calculated the length of each drumming pattern (in the r space, see Figs. 1D and 2),[b] and computed mean and variance length for each set of 32 patterns. Stepwise model selection suggested that neither generation nor transmission type should be entered in the ANOVA as a predictor of mean pattern length, and only generation should be entered in the ANOVA as a predictor of variance in pattern length. Generation was a significant predictor of variance in pattern length ($F = 4.0$, $P < 0.05$). In other words, while participants vary the number of beats produced within an experimental session across generations, this does not appear to be affected by the transmission type. Crucially, transmission type affects neither mean nor variance in pattern length, suggesting that simple differences in pattern length cannot account alone for structural variability across experimental groups.

Together, these results suggest that both structure and learnability change over transmission steps. Although transmission type does not affect learnability, it does affect the amount of structure: a within-participant design results in higher levels of structure. This provides preliminary support for the idiosyncrasy of biases hypothesis over the

---

[a]In both experiments, the participant was unaware that she would imitate her own, or someone else's, previous pattern.

[b]Each of the 32 imitated patterns can be described as a time series of interonset intervals (i.e., the time between adjacent drum hits $IOI_1, IOI_2, \ldots, IOI_n$). As in the original study, to account for possible tempo drift within and across patterns, we use the ratio between adjacent beats[41,46] (i.e., $r_i = IOI_{i+1}/IOI_i$). Hence, each pattern of n hits can be represented as a time series $r_1, \ldots, r_{n-2}$. The computational model proceeds by first clustering data points into rhythmic categories. We assume that participants do not perceive and represent the absolute magnitude of the r_i data points, but potentially reduce the variation among data points by assigning each to a rhythmic category.[43]
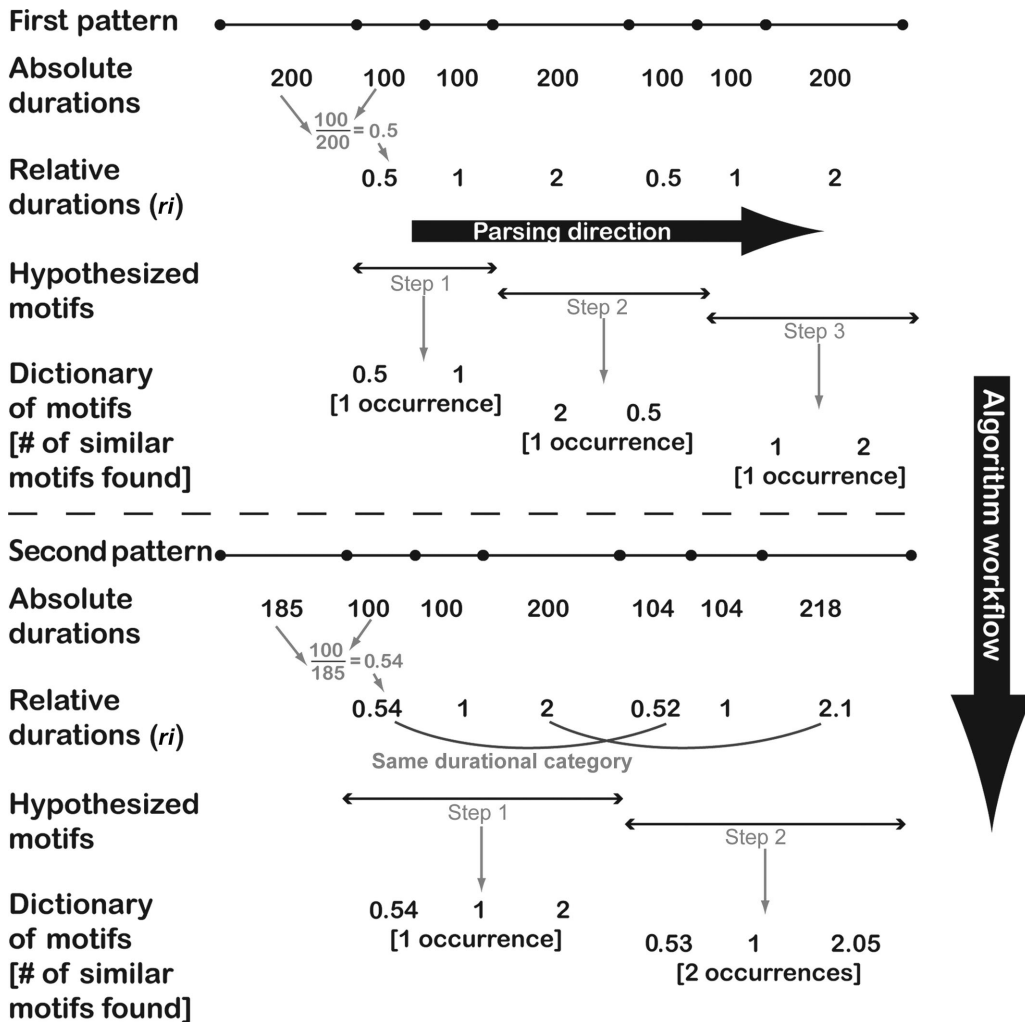
Figure 2. Latent variables Bayesian model. Sketch of how the algorithm processes two drumming patterns (workflow proceeds from top to bottom and from left to right). A drumming pattern (first pattern, topmost row) can be conceptualized as a series of durations marked by drum events (lines broken by circles). The (absolute) durations between drum events are represented as a vector of IOIs. By taking the ratios between adjacent IOIs, one obtains a vector of relative durations $r = (r_1, \ldots, r_{n-2})$. All IOI sequences with the same *r* vector have the same rhythmic pattern up to a tempo multiplicative constant. The algorithm generalizes first over r_i categories (e.g., in the first pattern, 0.5 and 0.5 belong to the same category) and then assigns every hypothesized motif either to its prototypical category or to a new category. When the participant hears a new pattern (second pattern) with more variability, ratios such as 0.54 and 0.52 might be assigned to the same r_i category. Likewise, the algorithm randomly attempts to be "greedier," probing the existence of motifs of length 3 or above, hence finding that sequences like (0.54, 1, 2) and (0.52, 1, 2.1) belong to the same category.

homogeneity of biases hypothesis. These inferential statistics suggest that repeated idiosyncrasies enhance the emergence of structure (within-participant transmission type). Conversely, new minds introduce more variance, slowing down the emergence of structure (between-participants transmission type). Inferential statistics, however,

are unfit to unveil the structures participants infer. Below, we outline a mathematical model approximating what participants perceive and learn.

Bayesian model: methods

Data from both experiments were further analyzed using a computational model (see Supplement,

online only). Our model formalizes the idea that participants may decompose individual drum patterns into sequences of motifs that can be reused across patterns. Given a drum pattern, the model attempts to infer boundaries between latent motifs and to categorize these motifs into coherent groups based on prototypes. Our approach takes inspiration from two related fields. First, since the inferential task is essentially one of joint segmentation and clustering, we can adapt techniques from the machine learning and speech technology literature[39] to specify a probabilistic model of underlying latent variables (e.g., motifs and their boundaries). Second, the literature on statistical learning provides a psychologically plausible algorithm that makes guesses about these unobserved variables.[40]

Our model formalizes two levels of structure within a pattern by grouping adjacent interval ratios[41] into a small number of Gaussian categories via Bayesian inferenceb and by using sequences of these inferred categories to construct an inventory of motifs that can be reused. Our posterior approximation algorithm learns this structure "on the fly" in a probabilistic, sequentially dependent, psychologically plausible fashion: as new elements of a drumming sequence are perceived, the backwards transition probabilities (BTPs) between interval ratio categories are estimated[37] and used to hypothesize boundaries between motifs. The model attempts to assign any hypothesized motif to an existing category of motifs (Fig. 2, top), and creates a new motif category whenever this fails (Fig. 2, bottom).

A candidate subsequence has to go through two criteria to be considered a motif: (1) how often has the model seen this particular subsequence of categories? and (2) how often have the preceding and current element been found together (BTPs)? If the same sequence has been seen before, this provides evidence for the current subsequence to be another occurrence of this motif type. If the first element of a subsequence and the previous element rarely co-occur, this provides evidence that the two elements belong to different motifs, implying a boundary. The model includes parameters that influence, for example, how willing the learner is to invent new motif categories. These parameters and all other details of the model are described in detail in the supplement to this article. In the analyses we present here, we set these parameters such that the model

is weakly biased to prefer reusing existing motifs but able to invent new categories whenever the data dictate. Because the model includes a free parameter that determines this balance of reuse and invention, this assumption could be straightforwardly revisited in future analyses. Crucially, we fix parameters to be identical across analyses of both experimental data sets and examine how the structures inferred by the model vary by experimental condition (rather than examine the specific structures inferred, which can be sensitive to model parameterization).

Bayesian model: results and discussion

We ran the model through both experimental data sets 10 times, which (insofar as our model is a psychologically plausible theory of participants' behavior) provides an approximation to the representation of patterns induced by participants. We examined two principal measures of structure: (1) the number of unique motifs discovered by the model at each generation of each chain and (2) the number of patterns in which each attested motif was discovered at each generation of each chain. Together, these measures quantify the evolution of structural regularity across a set of rhythms, over generations, as a function of the data participants saw and produced.

In both experimental conditions, the number of unique motifs attested within a generation decreased over generations (Fig. 3A, top row). In line with our inferential statistics, this suggests an increased degree of reuse of prototypical building blocks over generations and that these building blocks are discoverable by a simple algorithm making local, sequential decisions. The regression slope plotted in Figure 3A shows that reuse happens faster in the within-subjects chains and that the final generation of within-subjects chains reuses motifs to a slightly greater extent than the final generation of between-subjects chains.

Our analyses also suggest (Fig. 3A, second row) that, in within-subjects chains, each attested motif tends to be present in an increasing number of patterns over generations, suggesting participants are entrenching the motifs they have invented. This pattern is also visible in the between-subjects chains, but to a lesser extent. While these chains do evidence increasing reuse of motifs, they do not appear to evidence the same degree of entrenchment on a small set of widely reused motifs, as is suggested in
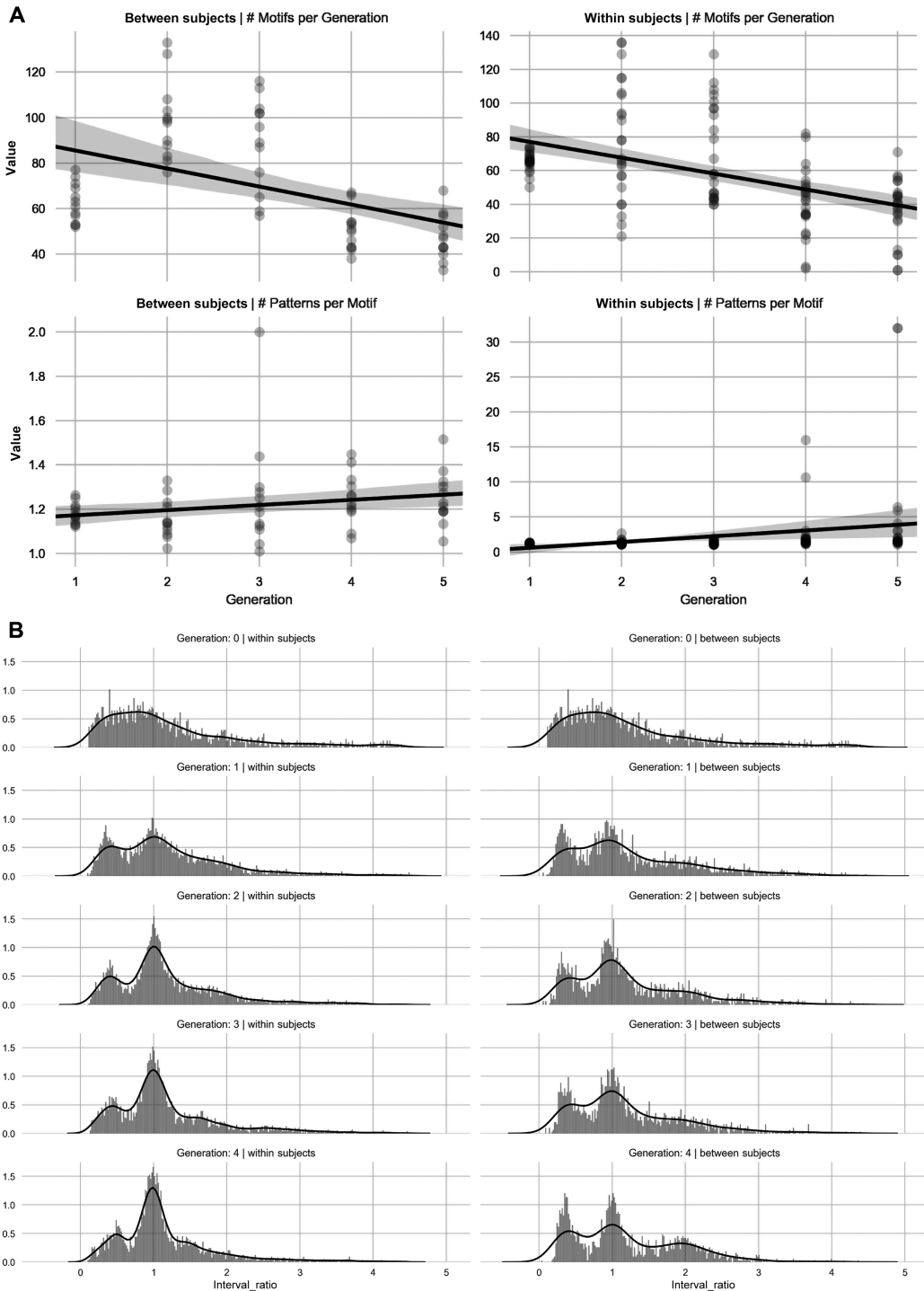
Figure 3. (A) The number of motifs inferred by the model (averaged over 10 independent simulations) at each generation (top row) in the within-subjects (right) and between-subjects (left) chains and the number of independent patterns in which each motif evidenced at least once in a generation was identified in that generation (bottom row). Points show these quantities for individual motifs (simulated 10 times); lines show regression slopes. (B) Normalized histograms for the distribution of interval ratios at each generation in the between-subjects (right) and within-subjects (left) chains. Lines show kernel density estimates of these distributions.

the within-subjects results. We interpret this as a sign that (1) both experimental conditions lead to an increase in structure, but (2) the within-subjects condition allows the idiosyncrasies of individual minds to repeatedly bias the distribution of structures in a chain-specific way that is less probable when new learners are forced to reinterpret the structures invented by previous individuals.

Visualizing integer ratios

We computed the distribution of interval ratios[13,14] at each generation (Fig. 3B). Both experimental conditions evidence a sharp transition from unstructured initial distributions (generation 0, top row) to highly structured categories of intervals. Interestingly, while the trimodal distribution with peaks near interval ratios is clear in the between-subjects data (as previously reported[14]), the within-subjects chains appear to converge on an approximately two-way category distribution with peaks at 1:2 and 1:1 ratios.[13,42] Figure 4 illustrates these differences in more detail, providing an overview of the evolution of interval ratio distributions across generations. The figure shows Gaussian kernel density estimates of the distribution of interval ratios that participants observed (blue lines) and then produced (black lines) at each generation of each chain (in both conditions).

Focusing on the between-participants chains (Fig. 4A), all five chains independently converge on an approximately three-way category distinction with one category of roughly equal-length intervals and one category each for long–short intervals and short–long intervals. In contrast, the within-participant chains (Fig. 4B) do not converge so emphatically toward this musical universal.[6,13,14] Instead, within-participant chains demonstrate idiosyncratic (but nonetheless structured) final distributions. Only two of the within-participant chains (chains 3 and 6) appear to approximate the three-way category distinction found in all between-participant chains, and even here those distinctions are less clear. The other four within-participant chains converge on different solutions: a single primary category of equal-length intervals strongly peaked at zero (chain 1); a four-way distinction with two below-zero categories (chain 2); a two-category distinction (chain 4) with roughly equal-length intervals and long–short intervals; and an approximately five-way distinction covering

the range (chain 5). It is also notable that the majority of chains in both conditions include a well-defined final category at mean zero—roughly adjacent equal-length intervals.

Also note the dynamics of distribution change over generations. In the between-participant chains, change is gradual and constant until the final generations: the distribution of produced interval ratios (black line) marginally deviates from the observed distribution (the blue line). In contrast, in the within-participant chains, we see an initial generation of deviance between observed and produced distributions as in the between-participant chains, but the following generations generally reproduce the observed distribution accurately (black lines trace blue lines). Along with the modeling results above, we take this to imply that within-participant chains do result in structured rhythms, but that this structure is largely idiosyncratic, arguing strongly against the nativist explanation for these universals.

General discussion and conclusions

Here, we investigated the role of individual cognitive biases in the creation of rhythmic patterns. We presented two rhythm-imitation experiments (simulated cultural transmission) in nonmusicians, analyzed using two complementary techniques. We showed that similar regularities emerge when participants are asked to imitate their own or other participants' patterns. However, different designs affect the degree of regularity emerging. When participants imitate their own previous productions, convergence is faster but results in less pronounced universal rhythmic features. We suggest this is due to the presence of weak idiosyncratic biases. When transmission occurs between participants, idiosyncratic biases partially cancel each other out. Instead, when transmission occurs within a participant, biases reinforce each other.[c]

Previous cultural transmission experiments of musical rhythms show that[12,14–16] (1) initially random sequences become structured, and (2) the resulting structures reflect universally observed

---

[c]Our conclusions partly contrast with computer simulations done for language, where a few outlier agents distort the signal transmitted in an otherwise heterogeneous population.[34]
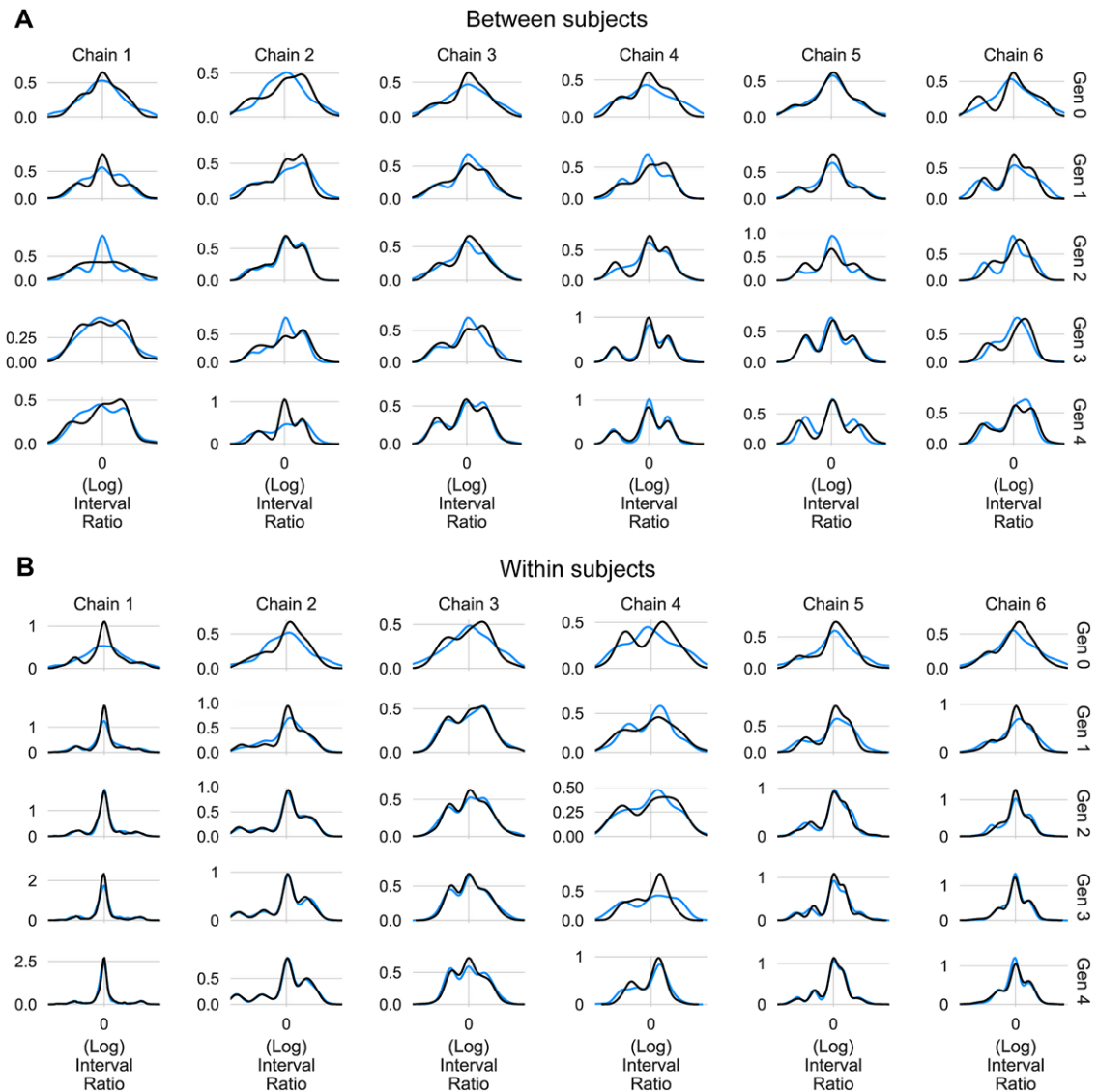
Figure 4. The evolution of interval-ratio distributions over generations, for all chains, in the between-participants (A) and within-participants experiments (B). Lines show Gaussian kernel density estimates inferred from (the logarithms of) the raw interval ratios participants observed (blue lines) and produced (black lines). We plot the distribution of the logarithm of interval ratios, since we found this to be the clearest illustration of categories. An interval ratio of 1 denotes adjacent equal-length intervals, and the logarithm of this ratio is 0. Values below zero in log space indicate that the second interval is shorter than the first (long–short interval pair); values above zero indicate that the second interval is longer (short–long interval pair).

patterns in music, such as small integer ratios of durations between note onsets. Our new experiment probed whether repeated computation by the same individual leads—as in the original experiment[14]—to increased structure and increased proximity to rhythmic universals. Our findings are surprising because (1) structure indeed emerges, and even

faster than in the between-participants chains, but (2) the resulting structure appears to be a poor approximation to the integer ratio universal.[6,13,14] We take this to be supportive of an account in which rhythmic structure results from a balance between individuals imprinting their biases and interactive transmission among listeners, learners, and

performers. Our results support the interactive hypothesis over the nativist hypothesis.[21]

Several limitations of this study could be addressed by future work. First, we adapted and designed the model to have cognitive plausibility and match experimental conditions as closely as possible. However, our Bayesian model still neglects several findings from rhythm perception and production[43,44] and should be understood as a first approximation. Future research should refine the model toward greater psychological and neural plausibility. Likewise, while we focused on psychological plausibility by implementing a sequential posterior approximation algorithm as our model of the learner, future analyses might instead focus on inferring the best possible structural descriptions by implementing more computationally intensive posterior sampling procedures. Second, the intensity information of each beat is unused in our models. Velocity and accents are an integral part of rhythm, so future extensions of this work should go beyond purely durational information. Third, several parameters in the Bayesian model undoubtedly influence our results. Future modeling efforts might aim to find empirical motivation for parameters accounting for human perception and induction of categories of intervals[43,44] or derive simpler models with fewer parameters. Ours is a first attempt, and we will share our data and scripts with interested researchers who would like to perform modifications. Fourth, given a restricted set of assumptions, we have predicted the number and distribution of rhythmic motifs inferred by participants. Ideally, whether participants actually acquire similar motif-like substructures will need to be tested experimentally by asking participants to classify motifs and check how closely these decisions align with the model's predictions.[d]

This report makes a contribution to a number of disciplines. Over the past century, a deep divide has separated cultural anthropology and music psychology.[1,11] Our experiments aim to bridge this divide with a design accounting for both the cultural medium and human biopsychological features. Likewise, our behavioral experiments could be combined with neuroimaging or electrophysiology techniques[45] to tap into the neural basis of human biases for musical rhythm. Within the interdisciplinary field of cultural evolution, we show how within-participant transmission speeds up the process of convergence. Quantitative models abound in music information retrieval, but are still scarce in music cognition. Here, we adapted some recent computational techniques to the interpretation of human data. In brief, we hope that our paper will spur a tighter integration of modeling and empirical research in the study of the psychology, neuroscience, and evolution of music.

Acknowledgments

---

[d]While we would expect less difference between conditions if priors were strong and homogenous, the design of the two experiments differs along one additional dimension. In the within-participant experiment, participants can potentially carry memory over from previous generations. If that were the case, one would predict a slower evolution of structure in the within-participant condition, as the effect of the prior would be relatively weakened from the drag of increasing pile of data. The fact that this does not occur may suggest that—at least in this design—participant memory is not a crucial factor. In fact, subjects in the within-participant condition were not told they were listening to their own data. That, plus the sheer number and length of sequences and the participants' lack of musical training, should explain why memory is not as strong a factor. Future research will be faced with disentangling bias' homogeneity from memory effects. Assigning participants to experimental groups based on their electroencephalographic signature might be a fruitful solution.[45]

Supporting Information

Additional supporting information may be found in the online version of this article.

Supplement. Experimental methods and models.

Competing interests

The authors declare no competing interests.

References

1. Ravignani, A., B. Thompson & P. Filippi. 2018. The evolution of musicality: What can be learned from language evolution research? *Front. Neurosci.* **12:** 1–8.

2. Honing, H., C. ten Cate, I. Peretz, *et al.* 2015. Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370:** 20140088.

3. Arom, S. 1991. *African Polyphony and Polyrhythm: Musical Structure and Methodology.* Cambridge: Cambridge University Press.

4. Nettl, B. 2000. An ethnomusicologist contemplates universals in musical sound and musical culture. In *The Origins of Music.* N.L. Wallin, B. Merker & S. Brown, Eds.: 463–472. Cambridge, MA: The MIT Press.

5. Brown, S. & J. Jordania. 2013. Universals in the world's musics. *Psychol. Music* **41:** 229–248.

6. Savage, P.E., S. Brown, E. Sakai, *et al.* 2015. Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. USA* **112:** 8987–8992.

7. Peretz, I., D. Vuvan, M.-É. Lagrois, *et al.* 2015. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370:** 20140090.

8. Sameiro-Barbosa, C.M. & E. Geiser. 2016. Sensory entrainment mechanisms in auditory perception: neural synchronization cortico-striatal activation. *Front. Neurosci.* **10:** 361.

9. Fritz, T., S. Jentschke, N. Gosselin, *et al.* 2009. Universal recognition of three basic emotions in music. *Curr. Biol.* **19:** 573–576.

10. Honing, H. & A. Ploeger. 2012. Cognition and the evolution of music: pitfalls and prospects. *Top. Cogn. Sci.* **4:** 513–524.

11. Lumaca, M., A. Ravignani & G. Baggio. Music evolution in the laboratory: cultural transmission meets neurophysiology. *Front. Neurosci.* In review.

12. Fitch, W.T. 2017. Cultural evolution: lab-cultured musical universals. *Nat. Hum. Behav.* **1:** 1–2.

13. Jacoby, N. & J.H. McDermott. 2017. Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* **27:** 359–370.

14. Ravignani, A., T. Delgado & S. Kirby. 2016. Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* **1:** 0007.

15. Ravignani, A. & T. Verhoef. Which melodic universals emerge from repeated signaling games? *Artificial Life* 24.

16. Lumaca, M. & G. Baggio. 2017. Cultural transmission and evolution of melodic structures in multi-generational signaling games. *Artif. Life* **23:** 406–423.

17. Le Bomin, S., G. Lecointre & E. Heyer. 2016. The evolution of musical diversity: the key role of vertical transmission. *PLoS One* **11:** e0151570.

18. Trehub, S.E. 2015. Cross-cultural convergence of musical features. *Proc. Natl. Acad. Sci. USA* **112:** 8809–8810.

19. Rzeszutek, T., P.E. Savage & S. Brown. 2012. The structure of cross-cultural musical diversity. *Proc. R. Soc. Lond. B Biol. Sci.* **279:** 1606–1612.

20. Verhoef, T., S. Kirby & B. de Boer. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *J. Phon.* **43:** 57–68.

21. Thompson, B., S. Kirby & K. Smith. 2016. Culture shapes the evolution of cognition. *Proc. Natl. Acad. Sci. USA* **113:** 4530–4535.

22. Miranda, E.R., S. Kirby & P. Todd. 2003. On computational models of the evolution of music: from the origins of musical taste to the emergence of grammars. *Contemp. Music Rev.* **22:** 91–111.

23. Kirby, S., T. Griffiths & K. Smith. 2014. Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* **28:** 108–114.

24. Kirby, S., M. Dowman & T.L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. USA* **104:** 5241–5245.

25. Kirby, S., H. Cornish & K. Smith. 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.

26. Ravignani, A., H. Honing & S.A. Kotz. 2017. The evolution of rhythm cognition: timing in music and speech. *Front. Hum. Neurosci.* **11:** 303.

27. Kirby, S. 2017. Culture and biology in the origins of linguistic structure. *Psychon. Bull. Rev.* **24:** 118–137.

28. Gingras, B., H. Honing, I. Peretz, *et al.* 2015. Defining the biological bases of individual differences in musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370:** 20140092.

29. Berwick, R.C. & N. Chomsky. 2017. Why only us: recent questions and answers. *J. Neurolinguistics* **43:** 166–177.

30. Fehér, O., I. Ljubičić, K. Suzuki, *et al.* 2017. Statistical learning in songbirds: from self-tutoring to song culture. *Philos. Trans. R. Soc. B* **372:** 20160053.

31. Feher, O., H. Wang, S. Saar, *et al.* 2009. *De novo* establishment of wild-type song culture in the zebra finch. *Nature* **459:** 564–568.

32. James, L.S. & J.T. Sakata. 2017. Learning biases underlie "universals" in avian vocal sequencing. *Curr. Biol.* **27:** 3676–3682.e4.

33. Tchernichovski, O. & D. Lipkind. 2017. Animal communication: origins of sequential structure in birdsong. *Curr. Biol.* **27:** R1268–R1269.

34. Navarro, D.J., A. Perfors, A. Kary, *et al.* 2017. When extremists win: on the behavior of iterated learning chains when priors are heterogeneous. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* pp. 847–852.

35. Griffiths, T.L. & M.L. Kalish. 2007. Language evolution by iterated learning with Bayesian agents. *Cogn. Sci.* **31:** 441–480.

36. Jamieson, R.K. & D. Mewhort. 2009. Applying an exemplar model to the artificial-grammar task: inferring grammaticality from similarity. *Q. J. Exp. Psychol.* **62:** 550–575.

37. Ravignani, A. & G. Madison. 2017. The paradox of isochrony in the evolution of human rhythm. *Front. Psychol.* **8:** 1820.

38. Levenshtein, V.I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Phys. Dokl.* **10:** 707–710.

39. Lee, C.-Y. & J. Glass. 2012. A nonparametric Bayesian approach to acoustic model discovery. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers—Vol. 1*: 40–49, Jeju Island, Korea. Association for Computational Linguistics, Stroudsburg, PA.

40. McCauley, S. & M. H. Christiansen. 2011. Learning simple statistics for language comprehension and production: the CAPPUCCINO model. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. L. Carlson, C. Hölscher & T. Shipley, Eds.: 1619–1624. Austin, TX: Cognitive Science Society.

41. Sakai, K., O. Hikosaka, S. Miyauchi, *et al.* 1999. Neural representation of a rhythm depends on its interval ratio. *J. Neurosci.* **19:** 10074–10081.

42. Motz, B.A., M.A. Erickson & W.P. Hetrick. 2013. To the beat of your own drum: cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain Cogn.* **81:** 329–336.

43. Desain, P. & H. Honing. 2002. The formation of rhythmic categories and metric priming. *Perception* **32:** 341–365.

44. Honing, H. 2013. Structure and interpretation of rhythm in music. In *Psychology of Music*. D. Deutsch, Ed.: 369–404. London: Academic Press.

45. Lumaca, M. & G. Baggio. 2016. Brain potentials predict learning, transmission and modification of an artificial symbolic system. *Soc. Cogn. Affect. Neurosci.* **11:** 1970–1979.

46. Grondin, S. 2010. Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten. Percept. Psychophys.* **72:** 561–582.