Supplemental Information

The evolutionary biology of dance, without frills

Andrea Ravignani, Peter F. Cook

Some basic concepts and methodological notes

Borrowing a definition from the physics of dynamical systems [S1-S3], we define *entrainment* as the process of mode locking of coupled oscillators: When two oscillating systems, characterized by different periods when they function independently, start interacting and assume a common period, they are said to entrain. Notice that the oscillators' relative phase does not enter the definition of entrainment.

Isochrony refers to a series of events, where each adjacent pair of events is separated by a constant duration. In other words, isochronous events are those which can be characterized by an oscillation period. For instance, the indicator for seconds in a clock produces isochronous events (i.e. one tick every second) [S4-S7].

Synchrony refers to two or more series of events happening at the same time. Synchrony can be intermittent (elsewhere called 'relative coordination' [S8]) or full (a.k.a. 'absolute coordination'), depending on whether some or all the events of the two series happen at the same time. In its most general sense, synchrony need not (but can) entail: (i) isochrony, e.g. two musical instruments playing in unison (non-isochronous synchrony), or (ii) entrainment, e.g. two clocks, one in London, the other in Edinburgh are in synchrony without any interaction or entrainment (non-entrained synchrony) [S4, S5, S9]. Likewise entrainment does not necessarily entail synchrony, e.g. two swings can entrain in antisynchrony (constant phase difference equal to half period). Finally, entrained oscillators are, by definition and to a first approximation, isochronous. Partially orthogonal to synchrony, *synchronization* can be defined as essentially phase-locking (i.e. a stable phase relationship), which does not necessarily imply simultaneity of events.

Here we define dance broadly as consistent production of movement sequences timed to an external rhythm. Dance may involve entrainment but does not need to. While not sufficient for dance, entrainment, synchrony, antisynchrony, isochrony and other temporal interaction modes are experimentally tractable constructs that are common in human dance and may be analyzed and compared between species.

The vocal learning - beat perception and synchronization hypothesis [S10, S11] is focused on 0-lag synchrony to music, and has relied heavily on findings that show intermittent phase relationships

between behavior and external stimuli that some would classify as 'relative coordination' [S8]. Other researchers have focused more broadly on entrainment, that is, stable phase relationships, 0-lag or not, between behavior and external stimuli, whether to music or something simpler.

We base our preliminary meta-analysis on definitions advanced by the VLH such that a positive finding requires some evidence of ability to move in time to musical stimuli, while negative evidence includes inability to move in time to music or simpler rhythmic stimuli. However, a more inclusive analysis of rhythmic entrainment would include positive evidence of flexible auditory-motor entrainment to simple or complex stimuli. We have computed a second analysis in line with these broader criteria that may better reflect broad rhythmic faculty. Moreover, at least four alternative analyses are possible (which all yield less favorable results to the VL hypothesis than the one we adopt here) depending on four alternative classifications of tested species' rhythmic capacities:

1. If one requires consistent *entrainment* to *music*, Ronan the California sea lion is the only attested non-human animal to date [S12, S13].
2. If one requires relative coordination or entrainment to music, the number of individual animals included in the meta-analysis will still be fairly small [S10, S12, S13].
3. If one accepts a broader definition of relative coordination, synchrony or entrainment to isochronous stimuli, many more species (e.g. positive evidence from apes) must be included in the analysis [S16, S17, S27].
4. Finally, limited entrainment and coordination to conspecific signals is found in a broad range of species, including insects and anurans [S5, S14, S15].

Overview of meta-analyses

The framework for the evolution of dance proposed by Laland and colleagues [S18] draws on an influential hypothesis in comparative psychology. The vocal learning - beat perception and synchronization hypothesis states that neural adaptations supporting vocal mimicry are a precondition for predictive and flexible motor entrainment (based on auditory cues alone) [S11]. This hypothesis received support in 2009 when a cockatoo, species capable of vocal production learning, was the first non-human animal shown able to intermittently synchronize to music [S10]. Although the vocal learning - beat perception and synchronization hypothesis is the most well-known hypothesis in the field, alternative hypotheses exist [S6, S7, S15, S19-S28]. Recent findings have challenged the vocal learning hypothesis: predictive and flexible motor entrainment (based on auditory cues alone) to music has been shown in a California sea lion, a species with no attested vocal learning capacities [S12, S13].

While we recognize the importance of Patel's hypothesis [S11], it is premature to accept it as proven - as Laland and colleagues do [S18]. Here we show how current evidence is quite agnostic about Patel's hypothesis, and we suggest which and how many crucial species must be tested before the inference predictive and tempo-flexible entrainment to an auditory beat->vocal learning is solid. We do this by advancing a novel Bayesian meta-analysis of the comparative entrainment evidence, and a (classical,

frequentist) statistical power analysis. Of note, a number of assumptions are required in formulating this meta-analysis. In all cases, we have selected those most favorable to the VLH. In the Bayesian analysis, we ask: if we find an entrainer, what's the relative likelihood that this is a vocal learner vs. that it is not? We find that this relative probability is quite low, ranging from 1:200 (favoring non vocal learners) to 2:1 (slightly favoring vocal learners), suggesting that one should still be quite agnostic about the entrainment-vocal learning link. This quantitative result is due to two factors, which can be explained without formulas. (See next paragraphs for mathematical details.) First, the prior probability of finding a vocal learner in the animal kingdom is quite low, as vocal learners are only a tiny subset of all animals [S29-S35]. However, as animals in entrainment experiments have been preferentially sampled from among vocal learners, the evidence for entrainment given vocal learning (or non vocal learning) needs to be weighted. Second, the conditional probability of finding flexible entrainment given that an animal is a vocal learner is higher than, though still of the same order of magnitude of, the conditional probability for entrainment given non vocal learning. Finally, our frequentist power meta-analysis treats every species previously tested on entrainment and vocal learning capacities as a data point, and it computes the minimum amount of additional data points needed in order for the vocal learning - beat perception and synchronization link to be statistically significant across species. Together, these analyses suggest that, in order to support the vocal learning - flexible/predictive synchronization *link*, in turn a precondition for the vocal learning - beat perception and synchronization hypothesis to hold, researchers should show: (a) lack of entrainment in 8 more non-vocal learners, (b) entrainment in 36 more vocal learners, or (c) vocal learning capacities in animals previously thought incapable of vocal learning. In quantifying number of vocal learning and vocal non-learning species shown capable of entrainment, and in determining proportion of vocal learning to non vocal learning species, we have striven to be very conservative. That is, we have interpreted ambiguous or contentious evidence in favor of the vocal learning hypothesis, not alternative hypotheses. More in general, the meta-analysis we propose is highly preliminary, and has some obvious limitations (see 'Caveats and future directions'). We however believe in its value as a first attempt at comparing all empirical evidence within a common framework.

Bayesian meta-analysis, with formulas

Conditional probability provides a set of principles that allow us to compute the degree of belief we should ascribe to competing scientific hypotheses. We ask: if we pick a random animal species and this shows evidence for flexible entrainment, how likely is this species to be capable of VL? This is denoted by P(VL|E), the probability of vocal learning given successful entrainment. How likely is it instead that a randomly-picked entraining species is not a VL? This is denoted by P(¬VL|E), the probability of no vocal learning conditional on entrainment.

Relative likelihoods are often more useful and easier to compute than absolute probabilities. A relative likelihood function describes the degree of relative strength of evidence between two inferential processes. For an animal showing evidence of flexible entrainment, how many times more likely is it

that this species is incapable vs. capable of VL? This is simply the ratio of the two probabilities just mentioned:

$$RL=P(\neg VL|E)/P(VL|E) \tag{1}$$

We use the term (relative) 'likelihood' instead of 'probability' to stress that we now compare relative plausibility of hypotheses on the parameters X, X' (which are possible partitions of animal species into groups) in order to maximize the function $P(X|E)/P(X'|E)$. Using Bayes' theorem, one can express the two unknown conditional probabilities using other probabilities, which in turn can be estimated from empirical research. Hence,

$$P(VL|E)=P(E|VL)P(VL)/P(E), \tag{2}$$

that is the probability of an 'entrainer' being a vocal learner equals the probability of a vocal learner being an entrainer, multiplied by the overall probability of that animal being a vocal learner, divided by the probability of entraining. A similar function, with analogous interpretation, can be calculated for the non vocal learning case:

$$P(\neg VL|E)=P(E|\neg VL)P(\neg VL)/P(E). \tag{3}$$

As we are only interested in relative likelihood, the number of probabilities to estimate is greatly reduced, and we do not need to estimate or calculate $P(E)$, the denominator in (2) and (3), explicitly:

$$P(\neg VL|E)/P(VL|E)=(P(E|\neg VL)P(\neg VL)/P(E))/(P(E|VL)P(VL)/P(E)). \tag{4}$$

Simplifying and rearranging the expression we obtain:

$$RL=P(\neg VL|E)/P(VL|E)=(P(E|\neg VL)/P(E|VL))*(P(\neg VL)/ P(VL)) . \tag{5}$$

This equation states that, if we pick a random species and show it capable of flexible entrainment, the relative likelihood of this species being a incapable vs. capable of VL is given by the overall probability of not being a VL divided by the overall probability of being a VL, times the probability that a non vocal learner can entrain divided by the probability that a vocal learner can entrain. All these four probabilities can be empirically estimated from past experiments [S10, S12, S16, S17, S27, S36-S39]. Notice that here we estimate these in the most favorable way to the vocal learninghypothesis, counting all cases of dubious entrainment as non-entrainment—each of the three equivocal studies and the "negative" study could be interpreted to support some degree of entrainment in the species studied. A value for RL close to 1 would mean that if an animal can entrain, no a-posteriori inference can be drawn on its vocal learning capacities. Patel's hypothesis would predict RL close to 0, because all

entrainers are likely to be vocal learners, hence P(¬VL|E)->0 and P(VL|E)->1. For a value of RL above 1, the higher the value, the stronger the evidence against Patel's hypothesis.

Table S1. (Related to Figure 1.) Contingency table showing species' count for animals with (VL) or without (¬VL) vocal learning, and lacking (¬E) or showing (E) entrainment capacities. Cases of intermittent entrainment to non-musical stimuli are denoted by ≈E. These cases are considered as negative results in the main analysis, and positive in a less conservative re-analysis.

	E	≈E	¬E	sum VL
VL	17=1+1+15 Patel, et al. [S10]; Hasegawa, et al. [S38]; Schachner, et al. [S36].	0	0	17
¬VL	1 Cook, et al. [S12].	3 Large & Gray [S27]; Bregman, et al. [S39]; Hattori, et al. [S16, S17].	1 Zarco, et al. [S37].	5
sum E	18	3	1	22

The conditional probabilities can be estimated from Table S1. All vocal learners examined to date have been capable of entrainment: P(E|VL)=1/1. In this first analysis, for the non vocal learners tested for entrainment, we adopt the most conservative estimate (i.e. favorable to the VL hypothesis): We only include Ronan the sea lion, who is capable of human-like entrainment at different tempi and stimuli [S12, S13], hence P(E|¬VL)=1/5. Combining the two estimates above we obtain P(E|¬VL)/P(E|VL)=1/5. How about the relative diffusion of vocal learning across animal species, P(¬VL)/P(VL)? If we only focus on the amniotes, there are about 25,000 species among non-avian reptiles, birds, and mammals. We focus on amniotes as the smallest monophyletic group including all species tested to date for entrainment. Several estimates can be used for the amount of vocal learning. Here we provide two: the most conservative, estimating only how many species have been shown beyond doubt to be open ended vocal learners, and the least conservative, counting the number of species in all amniote orders considered vocal learners [S29, S31-S35, S40-S50]. If we only include species for which vocal production learning has been rigorously shown, these are in the order of 10-10^2, hence P(¬VL)/P(VL)≈10^3, meaning that for every attested vocal learner, there are 1000 species which are not proven vocal learners, resulting in RL≈1000*1/5=200. If we include all members of clades containing a mix of species with limited, some, developed, or unstudied vocal learning capacities, P(VL)≈7100/25000 and P(¬VL)≈(25000-7100)/25000 =17900/25000. These are in fact the Trochiliformes (300 species), Psittaciformes (400), Passeriformes (5000), Chiroptera (1300), Cetaceans (100), plus a few isolated species, namely *Homo sapiens, Phoca vitulina, Elephas maximus*, and *Loxodonta africana*. Using this least conservative estimate, the relative likelihood amounts to RL≈1/5*17900/7100≈1/2. In other words, the likelihood of an entrainer being incapable vs. capable of VL is between 200:1 and 1:2, depending on whether one uses the most or least conservative estimate in

attributing vocal learning capacities to non-human animals. In the first case, 200:1, it would be quite safe to bet against vocal learning, while in the second case, 1:2, one should bet in favor of vocal learning, though it would be quite a risky bet. Finally, if we assume a uniform prior, that is $P(\neg VL)=P(VL)$, we obtain

$$RL=P(\neg VL|E)/P(VL|E)=P(E|\neg VL)/P(E|VL), \tag{6}$$

implying $RL=1/5$. Only by assuming that every amniote is equally likely to be as not to be a vocal learner, do we obtain strong *a posteriori* support for Patel's hypothesis under the current available data. Finally, if we rerun the analysis considering the debated cases in Table S1 as successful entrainment ('relative coordination' [S8]), we obtain $P(E|\neg VL)/P(E|VL)=4/5$. Combining this with the previous analysis, we obtain RL between 800:1 and 2:1, depending on how conservative our estimate is in attributing vocal learning capacities to non-human animals. This less conservative estimate is hence unfavorable to the VL hypothesis. However, notice that Patel's hypothesis predicts that only vocal learners will exhibit predictive and flexible entrainment (based on auditory cues alone), rather than any form of entrainment.

Notice that these estimates are only a first attempt at empirical meta-analysis of the flexible/predictive entrainment-vocal learning link, and rely on a number of assumptions that can be improved upon. In particular, vocal learning capacities are not uniformly and independently distributed across species and higher level taxonomic groups. An estimate of $P(VL)$ taking into account the phylogenetics, mutation rate and nuanced degrees of vocal learning across all 25000 amniotes would provide a better estimate of $P(VL)$, and hence of the strength of Patel's hypothesis. Moreover, the analysis assumes that the species populating Table S1 have been randomly and uniformly sampled from the 25000 amniotes, which is not precise. In particular, sampling of the 15 species in [S36] suffers from a number of biases, most notably that vocal learners accounted for 2/3 of the sampled species. We do not believe this invalidates our conclusions of agnosticism towards the vocal learning hypothesis, as taking this sampling bias into account would have resulted in even less support for the vocal learning hypothesis.

The advantage of this inferential approach to Patel's and other hypotheses on the origins of rhythm is that, much like human inferential reasoning, its strength can be continuously updated and refined as soon as new empirical data are available. In particular, one can ask: How does the equation need to change in order for Patel's hypothesis to gain overwhelming support? For RL to converge to 0, as predicted by Patel, either $P(E|\neg VL)/P(E|VL)$ or $P(\neg VL)/P(VL)$ need to tend to 0. This means that empirical research must show either (a) that several non vocal learners are unable to flexibly entrain, or (b) that more animal species are actually capable of vocal learning.

Statistical power meta-analysis

We have observed that more data points need to be collected across species to provide decisive support in favor or against Patel's hypothesis. Here we suggest how many additional experiments are needed.

When designing an experiment, and deciding on the number of subjects or trials to run, it is advisable to do a power analysis to know the minimum sample size (subjects or experimental trials) necessary to detect a statistically significant effect. In single experiments, the minimum sample size is the number of individuals to be tested, and the hypothesis under scrutiny is whether that species is capable of accomplishing the task. The same idea can be applied to a higher level of reasoning: what is the minimum number of species to test so that the hypothesis of entrainment-vocal learning association is supported (or, more precisely, the null hypothesis of no entrainment-vocal learning association can be rejected)?

Table S2. (Related to Figure 1.) Number of species which are capable vs. incapable of vocal production learning vs. entrainment beyond doubt (integer numbers) or need to be tested (variables x and y). Patel's hypothesis predicts a diagonal matrix, i.e. large x (almost all entertainers will be vocal learners) and y (non vocal learners will be incapable of entrainment).

	E	¬E
VL	2+x Patel, et al. [S10]; Hasegawa, et al. [S38].	0
¬VL	1 Cook, et al. [S12].	1+y Zarco, et al. [S37].

Fisher's exact test tests the hypothesis that there is some significant association between variables, in this case between vocal learning and entrainment. If we compile a 2x2 contingency table, where we classify all tested species in four categories depending on their vocal learning and entrainment abilities, each of the four cells counts the number of species that satisfy the corresponding constraint (Table S2). What is the minimum number of additional species to test in order to detect a significant association between entrainment and vocal learning? How many more vocal learners x need to be shown able to entrain? Or how many more non vocal learners y need to be shown incapable of flexible entrainment? If we fix our alpha level to .05, a power analysis of Fisher's exact test performed using a custom Python script reveals that either $x \geq 8$ or $y \geq 36$. This means that 8 additional non vocal learning species are necessary to establish an association between VL and E, assuming that all these 8 species will fail at entraining. Alternatively, and more laboriously, 36 additional vocal learners need to be tested to achieve statistical significance, assuming that all these species will succeed at entraining. Notice that this number is beyond the 15 vocal learners demonstrating some evidence of entrainment in youtube videos [S36]. The message from the current and previous analysis is quite clear, although a bit counterintuitive: to support or confute Patel's hypothesis, researchers studying entrainment should focus on species showing no vocal learning. Research on vocal learning should instead focus on showing vocal learning abilities in previously unattested species, rather than showing entrainment capacities in vocal learners.

Caveats and future directions

A few caveats apply to the analysis above. First, our approach is quite conservative, because it is usually more difficult to prove a negative rather than a positive beyond doubt. That is, as soon as a particular species or individual is shown able to entrain, testing can stop and the result of the experiment is positive. If, however, an entrainment experiment fails on a first attempt, a variety of reasons might account for this failure, and genuine inability at mastering the task is only one of the possible explanations. A potential entrainer might fail to entrain in a particular experiment, but potentially succeed in other setups, if given an opportunity. For this reason, we suggest that future research trying to show lack of entrainment in additional species should focus on those non-vocal learners which are easier to train using classical conditioning techniques. For instance, negative results in highly trainable domestic dogs may be explained most parsimoniously by a fundamental inability to entrain. However, lack of entrainment in rabbits, or corvids, could be best explained by these species' neophobia and lack of adequate operant training techniques optimized for these animals. Along these lines, the vocal learning without entrainment cell of Tables S1 and S2 could be populated by testing highly-trainable vocal learners, such as dolphins and zebra finches. Also in this case, the animals' trainability will make it comparatively easier to attribute repeated failures to entrain to a genuine inability, rather than methodological confounds.

Second, both the Bayesian analysis and the frequentist power analysis deal with the data needed to show a relationship, rather than a necessary/sufficient implication between vocal learning and entrainment. The closest behavioral research can get to showing causality between vocal learning and entrainment is having the two matrices in Tables S1 and S2 converge towards a triangular matrix, with the count of the bottom-left entry much smaller than the others. Research in neuroscience and genetics will be needed to actually show a true causal relationship.

Third, Laland and colleagues' hypothesis concerns not only vocal but also motor imitation. Hence, a similar meta-analysis for general motor imitation vs. entrainment should be performed as soon as the paucity of comparative data on motor imitation is replaced by cross-species studies. Ideally, in the long term, comparative research on the evolutionary origins of dance might want to classify species along the following behavioral dimensions:

- A. vocal imitation,
- B. non-vocal motor imitation,
- C. action sequencing and patterning,
- D. single-interval timing,
- E. natural behavior producing multiple temporal intervals,
- F. entrainment,
- G. presence of proto-dance behaviors.

Co-occurrence of a subset of these features across species will lend support to specific hypotheses for the evolutionary biology of dance. In particular, co-occurrence of A and F will support Patel's

hypothesis, and with the addition of B and C, Laland and colleagues' hypothesis. An intriguing strand of research linking A and C is already being pursued [S30]. Co-occurrence of D, E and G will instead lend support to our hypothesis.

Fourth, it has been suggested that the last common ancestor of all pinnipeds might have been a vocal learner, and then some daughter species, such as California sea lions, might have lost this ability [S51]. This is a hypothesis worth exploring with additional comparative research in pinnipeds, as it could provide insights on (i) Patel's hypothesis, (ii) pinniped communication in general, and (iii) the evolution of human speech by comparative approach [S52], as phylogenic divergence time in pinnipeds is comparable to that in Old world primates [S52].

Fifth, and finally, vocal learning is increasingly being recognized as a graded trait, rather than an all-or-none capacity [S30]. Hence, it might be possible to reframe Patel's original 'binary' hypothesis into a 'graded' one, namely a "vocal *control* and rhythmic synchronization hypothesis" (Fitch, pg. 308) [S53]. However, graded vocal learning or vocal control must be quantifiable in order to make the hypothesis testable. We suggest that a first approach at meta-testing the "vocal *control* and rhythmic synchronization hypothesis" (pg.308) [S53] would be to: (i) find a salient aspect of vocal control that can be described in one dimension (as opposed to a multi-dimensional space); (ii) quantify the degree of vocal control along this dimension for every animal species, so that a 'vocal control index' is assigned to each species, and these indexes can be ordinally (if not cardinally) compared across species (e.g. cockatoos should have a higher vocal control index than chickens); (iii) find an equivalent quantity to describe and compare rhythmic entrainment capacities in each animal species which have been or will be tested; (iv) perform a non-parametric correlation (e.g. Spearman's rho) to test whether higher vocal control predicts higher entrainment capacities, independently of their magnitude. Correlation values close to 1 would then support Fitch's "vocal *control* and rhythmic synchronization hypothesis" [S53].

Species mentioned in the main paper and future modeling work.

Interesting species to explore to test our framework for the evolutionary origins of human dance include, among others, fiddler crabs, *Uca annulipes* [S54-S56], and foot-flagging frogs, *Staurois latopalmatus* [S57-S59].

Quantitative tools originally developed to model timing in music [S60] have already been adapted to modeling interdependent movements [S61, S62], and could be fruitfully extended to model and compare dance-like behaviors across species [S63, S64].

Author Contributions. A. R. and P. C. conceived and wrote the paper.

Supplemental References

S1. Strogatz, S.H. (2003). Sync: The emerging science of spontaneous order, (Hyperion).

S2. Strogatz, S.H., and Stewart, I. (1993). Coupled oscillators and biological synchronization. Scientific American *269*, 102-109.

S3. Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). Synchronization: a universal concept in nonlinear sciences, Volume 12, (Cambridge university press).

S4. Ravignani, A., Bowling, D., and Kirby, S. (2014). The psychology of biological clocks: a new framework for the evolution of rhythm. In The Evolution of Language: Proceedings of the 10th international conference., S.R. EA Cartmill, H Lyn, H Cornish, ed. (World Scientific), pp. 262-269.

S5. Ravignani, A., Bowling, D.L., and Fitch, W.T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. Frontiers in Psychology *5*, 1118.

S6. Merker, B., Madison, G.S., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. Cortex *45*, 4-17.

S7. Fitch, W.T. (2012). The Biology and Evolution of Rhythm: Unravelling a Paradox. In Language and Music as Cognitive Systems, P. Rebuschat, M. Rohmeier, J.A. Hawkins and I. Cross, eds. (Oxford, UK: Oxford University Press), pp. 73-95.

S8. Collins, D.R., Park, H., and Turvey, M.T. (1998). Relative coordination reconsidered: A stochastic account. Motor Control *2*, 228-240.

S9. Polak, R., Jacoby, N., and London, J. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of Malian jembe drumming. Frontiers in Neuroscience *10*.

S10. Patel, A.D., Iversen, J.R., Bregman, M.R., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. Current Biology *19*, 827-830.

S11. Patel, A.D. (2006). Musical rhythm, linguistic rhythm, and human evolution. Music Perception *24*, 99-104.

S12. Cook, P., Rouse, A., Wilson, M., and Reichmuth, C.J. (2013). A California Sea Lion (*Zalophus californianus*) Can Keep the Beat: Motor Entrainment to Rhythmic Auditory Stimuli in a Non Vocal Mimic. Journal of Comparative Psychology *127*, 1-16.

S13. Rouse, A.A., Cook, P.F., Large, E.W., and Reichmuth, C. (2016). Beat Keeping in a Sea Lion as Coupled Oscillation: Implications for Comparative Understanding of Human Rhythm. Frontiers in Neuroscience *10*.

S14. Hartbauer, M., and Römer, H. (2016). Rhythm generation and rhythm perception in insects: The evolution of synchronous choruses. Frontiers in Neuroscience *10*.

S15. Wilson, M., and Cook, P.F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. Psychonomic Bulletin & Review, 1-13.

S16. Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. Scientific Reports *3*, 1566.

S17. Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2015). Distractor Effect of Auditory Rhythms on Self-Paced Tapping in Chimpanzees and Humans. PloS one *10*, e0130682.

S18.   Laland, K., Wilkins, C., and Clayton, N. (2016). The evolution of dance. Current Biology *26*, R5-R9.

S19.   Iversen, J.R. (2016). In the beginning was the beat: Evolutionary origins of musical rhythm in humans. In The Cambridge Companion to Percussion, R. Hartenberger, ed. (Cambridge University Press).

S20.   Merchant, H., and Honing, H. (2013). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. Frontiers in neuroscience *7*.

S21.   Merker, B. (2000). Synchronous chorusing and the origins of music. Musicae Scientiae *3*, 59-73.

S22.   Hagen, E.H., and Bryant, G.A. (2003). Music and dance as a coalition signaling system. Human Nature *14*, 21-51.

S23.   Hagen, E.H., and Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. Musicae Scientiae *13*, 291-320.

S24.   Honing, H., Merchant, H., Háden, G.P., Prado, L., and Bartolo, R. (2012). Rhesus Monkeys (*Macaca mulatta*) Detect Rhythmic Groups in Music, but Not the Beat. PloS one *7*.

S25.   Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. Evolution and Human Behavior *31*, 354-364.

S26.   Large, E.W. (2008). Resonating to musical rhythm: theory and experiment. Psychology of time, 189-232.

S27.   Large, E.W., and Gray, P.M. (2015). Spontaneous Tempo and Rhythmic Entrainment in a Bonobo (*Pan Paniscus*). Journal of Comparative Psychology.

S28.   Fitch, W.T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. Frontiers in systems neuroscience *7*.

S29.   Janik, V.M., and Slater, P.B. (1997). Vocal learning in mammals. Advances in the Study of Behavior *26*, 59-99.

S30.   Petkov, C.I., and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Frontiers in evolutionary neuroscience *4*.

S31.   Adret, P. (1992). Vocal learning induced with operant techniques: An overview. Netherlands Journal of Zoology *43*, 125-142.

S32.   Egnor, S.E.R., and Hauser, M.D. (2004). A paradox in the evolution of primate vocal learning. Trends in Neurosciences *27*, 649-654.

S33.   Fitch, W.T. (2000). The evolution of speech: a comparative review. Trends in Cognitive Sciences *4*, 258-267.

S34.   Nottebohm, F. (1975). A zoologists's view of some language phenomena with particular emphasis on vocal learning. In Foundations of language development : a multidisciplinary approach, Volume 1, E.H. Lenneberg and E. Lenneberg, eds. (New York: Academic Press), pp. 61-103.

S35.   Schusterman, R.J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. The evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication, 41-70.

S36.   Schachner, A., Brady, T.F., Pepperberg, I.M., and Hauser, M.D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. Current Biology *19*, 831-836.

S37.   Zarco, W., Merchant, H., Prado, L., and Mendez, J.C. (2009). Subsecond Timing in Primates: Comparison of Interval Production Between Human Subjects and Rhesus Monkeys. Journal of Neurophysiology *102*, 3191-3202.

S38.   Hasegawa, A., Okanoya, K., Hasegawa, T., and Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. Scientific reports *1*.

S39.   Bregman, M.R., Iversen, J.R., Lichman, D., Reinhart, M., and Patel, A.D. (2013). A method for testing synchronization to a musical beat in domestic horses (*Equus ferus caballus*). Empirical Musicology Review *7*, 144-156.

S40.   Fitch, W.T., and Jarvis, E.D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In Language, Music, and the Brain: A Mysterious Relationship, Volume 10, M.A. Arbib, ed. (Cambridge, Massachusetts: MIT Press), pp. 499-539.

S41.   Foote, A.D., Griffin, R.M., Howitt, D., Larsson, L., Miller, P.J., and Hoelzel, A.R. (2006). Killer whales are capable of vocal learning. Biology Letters *2*, 509-512.

S42.   Poole, J.H., Tyack, P.L., Stoeger-Horwath, A.S., and Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. Nature *434*, 455-456.

S43.   Ralls, K., Fiorelli, P., and Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. Canadian Journal of Zoology *63*, 1050-1056.

S44.   Reichmuth, C., and Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. Current opinion in neurobiology *28*, 66-71.

S45.   Reiss, D., and McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. Journal of Comparative Psychology *107*, 301.

S46.   Scharff, C., and Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. The Journal of Neuroscience *11*, 2896-2913.

S47.   Schusterman, R.J., and Reichmuth, C.J. (2008). Novel sound production via contingency learning in the Pacific walrus (*Odobenus rosmarus divergens*). Animal Cognition *11*, 319-327.

S48.   Stoeger, A.S., Mietchen, D., Oh, S., de Silva, S., Herbst, C.T., Kwon, S., and Fitch, W.T. (2012). An Asian elephant imitates human speech. Current Biology *22*, 2144-2148.

S49.   Boughman, J.W. (1998). Vocal learning by greater spear-nosed bats. Proceedings of the Royal Society of London B: Biological Sciences *265*, 227-233.

S50.   Knörnschild, M., Nagy, M., Metz, M., Mayer, F., and von Helversen, O. (2009). Complex vocal imitation during ontogeny in a bat. Biology Letters, rsbl20090685.

S51.   Patel, A.D. (2014). The evolutionary biology of musical rhythm: was Darwin wrong? PLoS biology *12*, e1001821.

S52.   Ravignani, A., Fitch, W.T., Hanke, F.D., Heinrich, T., Hurgitsch, B., Kotz, S.A., Scharff, C., Stoeger, A.S., and de Boer, B. (2016). What pinnipeds have to say about human speech, music, and the evolution of rhythm. Frontiers in Neuroscience *10*.

S53.   Fitch, W.T. (2015). The Biology and Evolution of Musical Rhythm: An Update. Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff, 293.

S54.   Backwell, P., Jennions, M., Passmore, N., and Christy, J. (1998). Synchronized courtship in fiddler crabs. Nature *391*, 31-32.

S55.   Kahn, A.T., Holman, L., and Backwell, P.R. (2014). Female preferences for timing in a fiddler crab with synchronous courtship waving displays. Animal Behaviour *98*, 35-39.

S56.   Reaney, L.T., Sims, R.A., Sims, S.W., Jennions, M.D., and Backwell, P.R. (2008). Experiments with robots explain synchronized courtship in fiddler crabs. Current Biology *18*, R62-R63.

S57.   Starnberger, I., Preininger, D., and Hödl, W. (2014). From uni-to multimodality: towards an integrative view on anuran communication. Journal of Comparative Physiology A *200*, 777-787.

S58.   Preininger, D., Boeckle, M., and Hödl, W. (2009). Communication in noisy environments II: Visual signaling behavior of male foot-flagging frogs *Staurois latopalmatus*. Herpetologica *65*, 166-173.

S59.   Boeckle, M., Preininger, D., and Hödl, W. (2009). Communication in noisy environments I: acoustic signals of *Staurois latopalmatus* Boulenger 1887. Herpetologica *65*, 154-165.

S60.   Wing, A.M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. Journal of The Royal Society Interface *11*, 20131125.

S61.   Honisch, J.J., Elliott, M.T., Jacoby, N., and Wing, A.M. (2016). Cue properties change timing strategies in group movement synchronisation. Scientific reports *6*, 19439.

S62.   Elliott, M.T., Chua, W.L., and Wing, A.M. (2016). Modelling single-person and multi-person event-based synchronisation. Current Opinion in Behavioral Sciences *8*, 167-174.

S63.   Ravignani, A. (2015). Evolving perceptual biases for antisynchrony: a form of temporal coordination beyond synchrony. Frontiers in Neuroscience *9*.

S64.   Ravignani, A. (2014). Chronometry for the chorusing herd: Hamilton's legacy on context-dependent acoustic signalling-a comment on Herbers (2013). Biology Letters *10*, 20131018.