PHILOSOPHICAL TRANSACTIONS B

Opinion piece

THE ROYAL SOCIETY
PUBLISHING

Vocal learning as a preadaptation for the evolution of human beat perception and synchronization

Aniruddh D. Patel[1,2]

[1]Department of Psychology, Tufts University, Medford, MA, USA
[2]Program in Brain, Mind, and Consciousness, Canadian Institute for Advanced Research, Toronto, Canada

ADP, 0000-0003-2229-9341

The human capacity to synchronize movements to an auditory beat is central to musical behaviour and to debates over the evolution of human musicality. Have humans evolved any neural specializations for music processing, or does music rely entirely on brain circuits that evolved for other reasons? The vocal learning and rhythmic synchronization hypothesis proposes that our ability to move in time with an auditory beat in a precise, predictive and tempo-flexible manner originated in the neural circuitry for complex vocal learning. In the 15 years since the hypothesis was proposed a variety of studies have supported it. However, one study has provided a significant challenge to the hypothesis. Furthermore, it is increasingly clear that vocal learning is not a binary trait animals have or lack, but varies more continuously across species. In the light of these developments and of recent progress in the neurobiology of beat processing and of vocal learning, the current paper revises the vocal learning hypothesis. It argues that an advanced form of vocal learning acts as a preadaptation for sporadic beat perception and synchronization (BPS), providing intrinsic rewards for predicting the temporal structure of complex acoustic sequences. It further proposes that in humans, mechanisms of gene-culture coevolution transformed this preadaptation into a genuine neural adaptation for sustained BPS. The larger significance of this proposal is that it outlines a hypothesis of cognitive gene-culture coevolution which makes testable predictions for neuroscience, cross-species studies and genetics.

This article is part of the theme issue 'Synchrony and rhythm interaction: from the brain to behavioural ecology'.

1. Introduction

Recent years have seen growing research on the evolutionary foundations of human musicality, defined as the spontaneously developing cognitive and sensorimotor abilities supporting musical behaviour [1]. Among these abilities is the capacity to synchronize rhythmic movements to an auditory beat. This 'beat perception and synchronization' (BPS) is a form of entrainment, whereby periodic motor and auditory rhythms become temporally coordinated with the close temporal alignment of movements to auditory beats. BPS is a prominent aspect of human musical behaviour, is not part of everyday spoken language and is foundational to dance [2]. Because BPS is culturally widespread, emerges without explicit instruction, begins developing early in life and builds on infant predispositions to move rhythmically to beat-based rhythms [3–9], it is relevant to evolutionary questions about musicality. Does BPS reflect any evolved neural specializations for music processing, or does it rest entirely on brain circuits that evolved for other reasons? This question goes to the heart of a long-standing debate about evolution and music. In *The Descent of Man* (1871), Darwin [10] argued humans had evolved to be musical, while in *The Principles of Psychology*

(1890), William James [11], who admired Darwin and believed the human mind was full of instincts, implied that music was a purely cultural invention built on brain mechanisms that evolved to serve other functions. Nearly 150 years later, we are nowhere near consensus on this debate, with detailed arguments on both sides (e.g. [12–14] versus [15–17]). BPS is important to this debate because it is the focus of a growing body of research relevant to evolutionary questions, spanning neuroscience, cross-species studies, genetics, cross-cultural work and developmental psychology (e.g. [18–21]).

Thanks to research in cognitive neuroscience, several key behavioural and neural features of BPS are well known. One such feature is the spontaneous tendency of listeners to predict the timing of beats, as shown by experiments where people tap to the beat of auditory rhythms: taps fall within a few tens of milliseconds of beats and often anticipate the beat, indicating precise temporal prediction of beat times [22]. Another key feature is tempo flexibility. When engaging in BPS humans have a preference for beats separated by approximately 500 ms, near a natural frequency of human movement [23]. However, humans exhibit precise and predictive temporal synchronization (e.g. taps closely aligned to beats) across a wide range of tempi, spanning inter-beat-intervals ranging from approximately 250 ms to approximately 1 s [24]. Precise temporal prediction and tempo flexibility enable tight coordination in group music-making and dance. Thus, these abilities are probably important to the social functions and psychological consequences of synchronized, collective musical behaviour [25–27].

In terms of neuroscience, BPS can be separated into perceptual and motor processes, which do not always co-occur because listeners can perceive a beat while remaining still (although see [28–30]). Yet brain imaging has revealed that even in the absence of movement, beat perception and motor system activity are linked. When humans perceive a beat in an auditory rhythm, several motor areas, including cortical premotor and subcortical motor control areas, are strongly active and interacting (e.g. [31–34]). Overall, the networks involved in beat perception without movement and in BPS have a great deal of overlap [35,36], and there is growing interest in the idea that motor system activity plays a causal role in predicting the timing of beats even when humans do not move to the beat [37–41].

According to one line of theorizing and computational modelling, BPS taps into ancient and widespread brain mechanisms of entrainment. These mechanisms involve endogenous cortical neural oscillations which become coupled to stimulus-driven oscillations of brain activity caused by sensory input [42,43]. This view reflects an old intuition that musical rhythm processing is rooted in fundamental aspects of animal biology. Darwin voiced this intuition in *The Descent of Man* (1871), when he wrote 'The perception, if not the enjoyment, of musical cadences and of rhythm is probably common to all animals, and no doubt depends on the common physiological nature of their nervous systems' [10, vol. 2, p. 333]. One appeal of the coupled oscillator framework is that it has been productively applied to diverse forms of biological entrainment, ranging from circadian rhythms to synchrony in insect choruses [44,45], and is thus attractive for its generality. Indeed, commenting on his model of adaptive coupled oscillators for the synchronization of rhythmic firefly flashing in tropical trees, Ermentrout [45, p. 584] noted 'In a broader

context, the equations we consider in this paper are similar to those describing an interconnected oscillatory neural network… The only significant difference between the firefly tree and an oscillatory neural network is in the time scale (milliseconds for the neural network and seconds for the firefly tree) and the space scale (microns for the neural system and meters for the insects).' Another reason for the appeal of a coupled oscillator framework for BPS is research suggesting that neural oscillations have an important role to play in sensory, motor and cognitive processing across a range of species, including humans [46–48].

The coupled oscillator perspective on BPS entails the idea that many animal species should be capable of this form of rhythmic entrainment, since the purported brain mechanisms are very general. Yet BPS is notably absent in our most familiar animal companions, such as, dogs, cats and farm animals, even though these species have lived with humans and their music for thousands of years [49,50]. Wilson & Cook [51] have suggested that this does not reflect the lack of a capacity for BPS and is owing to other factors that inhibit BPS from developing spontaneously. These could include a lack of motivation for BPS, inattention to auditory rhythms because of their ecological irrelevance for the animal, or the fact that the rhythms of human music are not at tempi suited to the natural frequencies of the animal's movements. Consistent with this view, Cook and colleagues showed that a California sea lion (*Zalophus californianus*), which did not spontaneously engage in BPS when exposed to rhythmic music, could learn to synchronize her head bobs to a musical beat via operant training [52].

A very different hypothesis suggests that the capacity for BPS is limited to a narrow range of species. The 'vocal learning and rhythmic synchronization hypothesis' proposed that BPS relies on specialized auditory–motor forebrain circuitry which originally evolved to serve complex vocal learning [53]. In complex vocal learning, an animal requires auditory input to develop its normal species-specific vocalizations, because this input forms an auditory template which guides the development of the animal's own vocalizations [54]. Complex vocal learning occurs in a few groups of mammals, including cetaceans, pinnipeds and humans uniquely among primates, and in three groups of birds: songbirds, parrots and hummingbirds.

The vocal learning hypothesis (VLH) predicts that only species with complex vocal learning are capable of BPS. This motivated a number of studies testing the hypothesis, including research demonstrating BPS in parrots, which supported it [55,56]. However, the sea lion study mentioned above challenged the hypothesis, because sea lions do not show evidence of complex vocal learning. While there are open questions about whether the sea lion study refutes the VLH, as discussed in §3 below, there are other reasons to revisit the hypothesis. One such reason is the growing view that vocal learning should not be considered a dichotomous trait that animals have or lack, but a more continuous trait along which animals vary, or a modular trait with distinct subcomponents which can dissociate in different species [57–60]. Another reason to revisit the hypothesis is that parrots appear to be the only nonhuman vocal learners to show spontaneous BPS to human music,[1] despite the fact that songbirds have complex vocal learning and are sometimes extensively exposed to human music as pets [56,63]. This is important because recent neural research has revealed

that the parrot vocal learning system is more elaborate than that of songbirds [64]. A final reason to revisit the VLH is the considerable amount of neurobiological research on beat processing and vocal learning which has taken place in the 15 years since the hypothesis was published.

Collectively, these factors suggest it is time to reconsider the hypothesis that complex vocal learning is related to the emergence of BPS in our species. The current paper addresses this goal and is organized as follows: I first clarify the scope, background and evolutionary implications of the original VLH. I next critically examine the data that challenge the hypothesis and then discuss the implications of a continuum/modular view of vocal learning for the hypothesis. Based on these considerations, I propose a revision of the hypothesis whereby an advanced form of vocal learning acted as a preadaptation for BPS. I then suggest that once this form of vocal learning evolved in our lineage, it interacted with ancestral primate rhythmic vocal behaviours to produce sporadic BPS as a fortuitous trait, i.e. intermittent BPS in response to periodic auditory rhythms. Finally, I propose that our capacity for BPS was sharpened by gene-culture coevolution, leading to evolved neural specializations for sustained BPS in humans. I describe the predictions this view makes in terms of neuroscience, cross-species studies and genetics, and close by discussing the larger significance of research on BPS for theoretical issues in evolutionary biology.

2. The original vocal learning and rhythmic synchronization hypothesis

In this section, I discuss the scope, background and evolutionary implications of the original vocal learning and rhythmic synchronization hypothesis [53] (henceforth, VLH). Understanding the scope of the VLH is particularly important, because it explains why the hypothesis has not been challenged to date by any studies other than the sea lion research mentioned above.

(a) Scope

The VLH focuses on BPS, which differs in several ways from well-known cases of rhythmic entrainment in other species, as noted in earlier publications [53,65]. First, BPS includes the capacity to align rhythmic movement to a periodicity perceived in complex auditory rhythms, distinguishing it from synchronization to quasi-metronomic auditory patterns as seen in crickets and katydids [66–68]. Second, BPS involves predictive and temporally precise alignment of movement with beats over a wide tempo range (e.g. from approximately 50% slower to 100% faster than one beat every 500 ms, as noted previously), contrasting with the narrower ranges of tempi over which some insects can synchronize their periodic sonic pulses in a phase-aligned fashion [69–71]. Third, BPS often involves movements which are not themselves aimed at sound production, such as head bobbing or rhythmic movements of the arms or trunk [72], unlike rhythmic entrainment in insect acoustic chorusing, which is aimed at sound production.

To date, research on synchronization of movement to an auditory beat in monkeys has focused on metronomes (e.g. [73,74]) or tempo-varying metronomes [75], and thus does not challenge the VLH, as a defining feature of BPS is the ability to synchronize to periodicities perceived in complex auditory rhythms. Nevertheless, such research has been a valuable source of neural data on how primate brains coordinate rhythmic movements with rhythmic sounds, and these data have informed neural theories of human beat perception [41]. Interestingly, primate research suggests that monkeys and humans have similar capacities in terms of single-interval or 'absolute' timing, but differ in capacities related to beat-based or 'relative' timing, supporting the gradual audiomotor evolution hypothesis [76]. This hypothesis is consistent with neural research on rhythm perception in monkeys which finds that they are sensitive to isochrony but do not appear to perceive a beat in complex rhythms [77,78].

In terms of predictive and tempo-flexible synchronization with a beat, a striking finding from monkey research is that when trained to synchronize movements to a metronome, their spontaneous tendency is to move reactively rather than predictively with respect to the stimulus, unlike humans [73,79]. Recent research has shown monkeys can be trained to synchronize predictively if every predictive movement is rewarded [75,80], raising the idea that species may lie along a continuum of ability or proclivity for reactive versus predictive motor synchronization with a beat. Interestingly, while rats appear to share monkeys' spontaneous tendency for reactive motor synchronization to metronomes [81], Hattori et al. [82] demonstrated that a chimpanzee showed spontaneous predictive synchronization of movement to a metronome. However, this only occurred at one tempo near the animal's spontaneous tapping tempo, and the animal showed no evidence of tempo flexibility. Research with a bonobo has shown that it occasionally synchronized predictively during concurrent rhythmic drumming with a human partner [83], but because the bonobo and human could see each other it is unclear if the animal is capable of BPS without support from visual rhythmic signals. This is a concern as research with other primates reveals that they are better at synchronizing movement with discretely timed visual versus auditory rhythms, unlike humans [22,75]. Finally, despite anecdotal reports of spontaneous BPS in horses trotting to music without a rider onboard who might cue them to the beat, empirical work has not provided evidence of tempo flexibility in this behaviour, although methods for testing this have been developed and research is currently underway [84,85].

(b) Background and evidence

When the VLH was proposed in 2006, there were no known cases of BPS in nonhuman animals, either spontaneous or trained, and it had been suggested that BPS was uniquely human [86]. The VLH was motivated by a synthesis of behavioural, neural and cross-species research. Behavioural experiments had shown that humans were far better at extracting a beat and synchronizing to it in complex auditory versus visual rhythms matched in temporal structure [22]. (Later work extended this auditory advantage to matched complex tactile rhythms [87,88].) This suggested that beat perception involved specialized auditory–motor processing. When considering evolutionary forces which might have strengthened auditory–motor processing in humans, complex vocal learning was a plausible candidate as it involves tight coupling between auditory input and motor output in order to match vocal movements to an auditory model [89,90]. In terms of neural work, brain imaging had revealed

that even in the absence of movement, beat perception engages a number of motor regions of the human brain [31,91], including premotor and basal ganglia (striatal) regions. Interestingly, it was known from research on birds that the evolution of complex vocal learning was associated with neural specializations in premotor and striatal regions [90]. Collectively, these findings led to the VLH, which proposed that the neural circuitry for complex vocal learning was a necessary prerequisite for the capacity for BPS [53].

The key prediction of the VLH hypothesis was that vocal non-learning animals would not be capable of BPS, while only animals with complex vocal learning would have this capacity. (It is worth noting that while the VLH claimed the neural circuitry for complex vocal learning was a *necessary* foundation for BPS, it never claimed it was *sufficient* [65].) Research demonstrating BPS to music in large parrots, such as the sulfur-crested cockatoo 'Snowball', supported the hypothesis [55,56] and experimental work on pecking to a metronome in small parrots (budgerigars) also proved consistent with the hypothesis [92], though synchronization to a beat in more complex auditory stimuli remains to be studied in smaller parrots.

(c) Evolutionary implications

The VLH suggested that brains shaped by evolution for complex vocal learning had 'BPS potential' as a byproduct of their wiring. As parrots are not known to engage in BPS as part of their natural behaviour [93], the discovery of BPS in these animals supported the theoretical position that key components of human musicality rely on brain circuits that evolved for other reasons, and that humans have not evolved neural specializations for music processing [2,11,15–17,94].

3. A critical examination of data challenging the vocal learning hypothesis

To date, the only challenge to the VLH is from a study of a California sea lion ('Ronan') who was trained to synchronize her head bobs to a musical beat and who showed tempo flexibility in generalizing this ability to novel, untrained tempi ranging from 20% slower to 10% faster than the original tempo of 130 beats per minute (BPM) [52]. Because sea lions are not known to be complex vocal learners, this excellent study challenges the VLH. Yet several facts suggest that this study does not refute the VLH.

First, while sea lions are traditionally considered a vocally inflexible species, the upper limits of their vocal learning capacities are not yet known [95] and have not been studied using modern methods applied to other pinnipeds [96]. This is a concern because there is strong evidence of complex vocal learning in pinnipeds such as grey seals and harbour seals [96,97]. Indeed, it is already known that adult sea lions are more vocally flexible than macaque monkeys, as the former, but not the latter, can be trained to vocalize on command [95]. (Interestingly, juvenile macaques (*Macaca mulatta*) can be trained to vocalize on cue, but unlike sea lions this ability is lost as they grow into adulthood, even as they retain the ability to make non-vocal movements on command [98,99].) Given sea lions' voluntary control of vocalization, it would be interesting to see if they can be trained to modulate

spectral or temporal aspects of their calls, which would indicate even greater vocal flexibility.

Second, Ronan was tested as a juvenile, between 3 and 4 years of age, before the age of sexual maturity in this species, which is around 4–5. Juveniles in some mammalian and avian species are more vocally flexible than adults [100,101]. Furthermore, related to a point made by Schachner [102], the intensive operant training Ronan received in synchronizing movements to sounds (first with metronomes, later with music), combined with the heightened plasticity of juvenile mammalian brains [103], raises the question of whether Ronan developed unusually strong auditory–motor forebrain circuitry compared to normal sea lions. This could be addressed via structural brain imaging of auditory–motor forebrain connections in Ronan compared to conspecifics without early auditory–motor training, using recently developed *in vivo* magnetic resonance imaging (MRI) protocols for sea lions [104].

Third, as noted by Merker *et al.* [105], the way Ronan synchronized her movements to music was unlike human BPS in an important respect. Specifically, there was a strong change in the phase relationship of her head bobs to the musical beat as a function of tempo. When tested for her ability to synchronize at novel tempi, Ronan's head bobs lagged considerably behind the beat by an average of approximately 90° at the fastest tempo (143 BPM) and occurred considerably before the beat by an average of approximately 60° at the slowest tempo (117 BPM) (fig. 5 of [52]). This is unlike human BPS, in which rhythmic movements and beats remain much more phase aligned across a comparable range of tempo variation. Ronan's pattern of phase leads and lags is reminiscent of an oscillator with a single intrinsic period driven by nearby frequencies [106]. Subsequent elegant work with Ronan studying her synchronization to metronomes with phase or tempo perturbations [107], and modelling her behaviour with an oscillator capable of period and phase correction, showed that Ronan exhibited a low degree of period coupling to the stimulus, below the range of period coupling reported in prior human studies. This suggests her synchronization abilities may reflect different mechanisms from those used by humans, perhaps relying more heavily on subcortical circuits.

Based on the above points, I believe that rejecting the VLH on the basis of Ronan's rhythmic entrainment abilities is premature. Nevertheless, the pioneering studies of Cook, Rouse and colleagues [52,107] are an important challenge to the VLH and motivate further work on pinnipeds' rhythmic synchronization abilities. As a clade with a broad range of vocal flexibility across species [95], pinnipeds are a particularly promising group for studying phylogenetic relationships between vocal learning and rhythmic synchronization in mammals [108,109].

4. Vocal learning as a continuous or modular trait

The original VLH proposed that complex vocal learning provided the evolutionary and neural foundations for BPS. As noted above, complex vocal learners require auditory input to develop their normal species-specific vocalizations, because this input acts as a template that guides vocal development [54]. Complex vocal learning is associated with specializations of auditory–motor forebrain circuitry which support sophisticated auditory–motor neural interactions

[110]. This type of vocal learning can be distinguished from limited vocal learning, where the latter is 'the ability to fine-tune acoustic features of species-specific vocalizations that can develop in the absence of auditory input because innate motor programs can generate the species-specific pattern' [54, p. 3]. While humans are the only primate with complex vocal learning, limited vocal learning occurs more broadly in primates, including in chimpanzees [111,112]).

The distinction between complex and limited vocal learning highlights the fact that vocal learning is not a binary trait that animals have or lack. Indeed, this has been known for some time. Songbird researchers have long distinguished between closed-ended and open-ended complex vocal learners, where the former, such as zebra finches, do not modify their repertoire after the initial song learning, while the latter, such as European starlings, continue to modify their repertoire as adults. Furthermore, complex vocal learners can differ in other ways, including in whether they copy a tutor model or improvise on tutor songs, and whether they 'copy only tutor songs that fit tightly constrained species-specific parameters or copy essentially anything they hear' [113].

Of direct relevance to the VLH, recent theoretical and empirical work suggests a continuum in vocal learning capacities across species, or regards vocal learning as comprised multiple distinct abilities that may be targeted independently by evolutionary pressures [57–59]. These include the ability to flexibly coordinate the timing of innate vocalizations in social interactions with conspecifics, such as in the antiphonal calling seen in marmosets and singing mice [114,115], and the ability to modify vocalizations as a function of social and auditory experience, or 'vocal plasticity'. Vocal plasticity is the most relevant for the VLH, because substantial modification of the spectral and temporal properties of vocalizations based on auditory experience requires extensive auditory–motor crosstalk in the forebrain, and extensive, precise auditory–motor cortical crosstalk is a foundation of BPS. In this regard, the 'vocal learning continuum hypothesis' [57,58,60] is of particular interest, as it suggests that vocal plasticity varies in a quasi-continuous way across vertebrates. As shown in figure 1, the hypothesis groups parrots and humans together as high vocal learners at a far end of this continuum, with greater vocal plasticity than complex vocal learners.

This placement makes intuitive sense because it has long been known that parrots have extraordinary vocal plasticity, being able to imitate human speech and other sounds with high fidelity. In support of placing them in a distinct category relative to other complex vocal learners, recent neural research using gene expression and neural tract tracing [64,116] has revealed that parrots have a uniquely elaborated vocal learning system, consisting of a 'core' system similar to that found in songbirds, and a 'shell' system unique to parrots (figure 2).

While the relative contributions of the core and shell systems to parrot vocal plasticity remain to be understood, one intriguing possibility is that these distinct yet interconnected systems facilitate independent control of the syrinx (the avian analogue of the mammalian larynx) and tongue during vocalization. It has been found that parrots modulate both of these anatomical structures to shape vocal sounds [117], whereas in songbirds the acoustic structure of songs is largely produced by the syrinx rather than the tongue [110]. Regardless of the precise functional significance of the elaborated vocal control system of parrots, Chakraborty & Jarvis [116] propose that it enhances vocal–auditory–motor
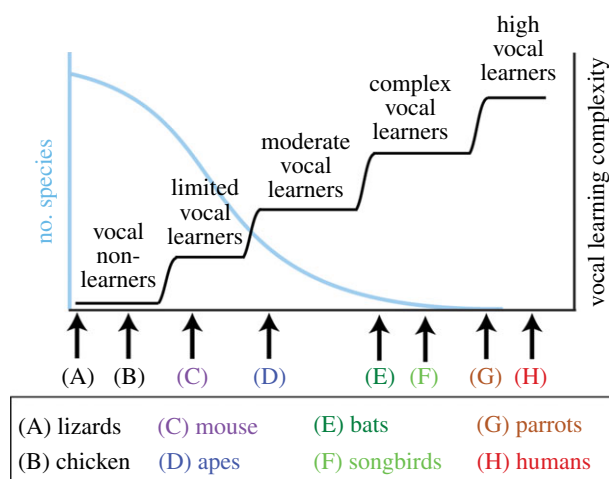

Figure 1. The vocal learning continuum hypothesis, from Petkov & Jarvis [57], updated by Jarvis [60]. Diagram of hypothesized stepwise continuous ability of vocal learning among vertebrates (right y-axis), from simple to more complex forms (x-axis). As vocal learning complexity increases, there are a decreasing number of species with the ability (left y-axis). (A–H) Proposed example species at each step on the continuum. The continuum ranges from lizards that do not vocalize and have no vocal learning, to nonhuman primates with limited vocal learning, to songbirds with complex vocal learning, to parrots and humans with high vocal learning. (Figure and caption modified from [60] with permission from the author.) (Online version in colour.)

integration compared to the songbird system and suggest that this is related to parrots' abilities to synchronize body movements to musical rhythms.[2]

The discovery of a dual pathway system for vocal learning and control in parrots is fascinating in light of an influential model of human speech processing in which the sensorimotor control of speech relies on two pathways within a complex 'dorsal stream' which bidirectionally connects cortical premotor and auditory regions via the temporo-parietal cortex [119]. These two pathways are probably important to our ability to independently control the larynx and tongue in shaping vocal sounds [120,121]. Figure 3a,b shows a schematic of this model, which distinguishes dorsal and ventral streams in spoken language processing, and which has two pathways within the dorsal stream, labelled dorso-dorsal and dorso-ventral. In this model, dorsal stream pathways are involved in speech sound-to-articulation mapping and ventral stream pathways are involved in mapping perceived sounds onto lexical representations. Figure 3b shows a more elaborate diagram of dorsal stream pathways based on a review of long-distance neural connections involved in auditory language processing [122].

Relevant to the VLH, neural studies in humans indicate the involvement of dorsal stream regions in BPS and in beat perception without overt movement, including the dorsal premotor cortex and parietal cortex, which are connected by the blue pathway in figure 3c [32,34,35,123]. Patel & Iversen [38] proposed that dorsal stream pathways are crucial for communicating temporal predictions about beat timing from premotor to secondary auditory regions, via the parietal cortex. Based on prior neuroanatomical research in monkeys [124], Patel and Iversen further proposed that a specific part of the dorsal stream is much more strongly developed in humans than in monkeys owing to the evolution of vocal learning in our lineage,
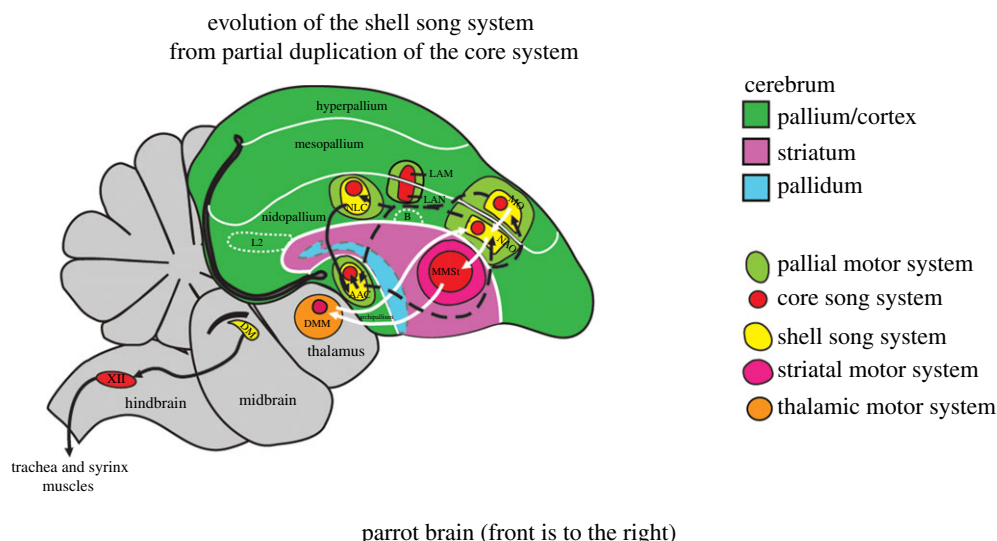
evolution of the shell song system
from partial duplication of the core system

parrot brain (front is to the right)

Figure 2. Schematic of the vocal system in a parrot brain, adapted from Chakraborty & Jarvis [116] with permission from the authors and The Royal Society. Red regions, core song system (similar to songbirds); yellow regions in pallium/cortex, shell song system (unique to parrots). The shell system is proposed to have evolved out of a partial duplication of the core song system. Black solid arrows, posterior vocal motor pathway; white solid arrows, anterior vocal motor pathway; dashed arrows, connections between core and shell systems. Not all connections are shown for simplicity. See caption of the original figure in [116] for definitions of acronyms. (Online version in colour.)
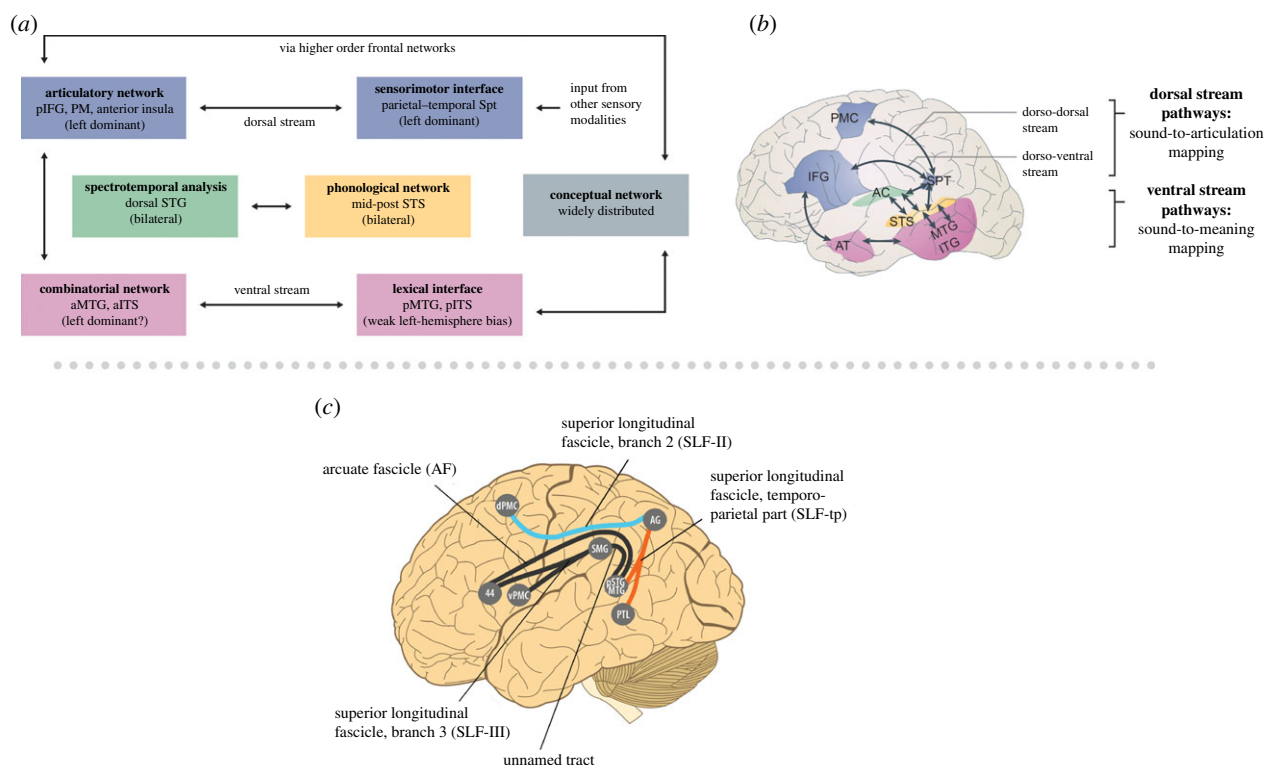

Figure 3. (a,b) Dual stream model of spoken language processing, adapted from Hickok & Poeppel [119]. Colours in the functional modules of (a) are matched to brain regions in (b), which shows neural pathways with dashed lines. Acronyms in (b): PMC, premotor cortex; IFG, inferior frontal gyrus; SPT, sylvian parieto-temporal area; AC, auditory cortex; STS, superior temporal sulcus; MTG, middle temporal gyrus; ITG, inferior temporal gyrus AT, anterior temporal cortex. (c) A more detailed view of dorsal stream pathways involved in spoken language (from [50], adapted from [122]). Of particular interest for BPS are connections shown in orange and blue: orange connections link secondary auditory regions in the posterior superior temporal gyrus/middle temporal gyrus (pSTG/MTG) and parietal regions near the angular gyrus (AG), and blue connections link regions near the angular gyrus to the dorsal premotor cortex (dPMC). These connections correspond to two branches of the superior longitudinal fasciculus (SLF): the temporo-parietal branch (SLF-tp) and the second branch (SLF II). Both tracts appear to play a role in sound-to-articulation mapping, which is part of vocal learning, and have been proposed to support auditory–motor interactions serving beat perception [38]. Other acronyms in (c): PTL, posterior temporal lobe; SMG, supramarginal gyrus; vPMC, ventral premotor cortex; 44, Brodmann area 44 (part of Broca's area). (Online version in colour.)

namely the orange fibre tract in figure 3c linking auditory regions in posterior superior temporal gyrus to regions around the angular gyrus. The relative weakness of this connection in monkeys could help explain why they do not spontaneously move predictively when synchronizing to auditory rhythms.

Stepping back to the larger picture, the additional neural regions and pathways for vocal learning in parrots compared to complex vocal learning songbirds are intriguing when juxtaposed with the complexity of the human dorsal auditory stream, which is involved in vocal learning and BPS. This juxtaposition, combined with the fact that only parrots and humans show spontaneous BPS to music, suggest that spontaneous BPS has an evolutionary relationship to high vocal learning rather than to complex vocal learning more generally.

5. The revised vocal learning and rhythmic synchronization hypothesis

The original VLH proposed that the capacity for BPS relied on neural substrates that first evolved to serve complex vocal learning. Based on the considerations reviewed above, here I propose the revised vocal learning and rhythmic synchronization hypothesis or 'rVLH'. Like the VLH, the rVLH is focused on BPS, which differs in several ways from synchronous rhythmic behaviours seen in insects and many other species, as previously discussed. However, unlike the original VLH, the rVLH shifts the focus from complex vocal learning as a prerequisite for BPS. Rather, it seeks to explain why spontaneous BPS occurs in 'high vocal learners' such as humans and parrots, who have a behaviourally and neurally more elaborate form of vocal learning than complex vocal learners (figure 1 and §4 above). The rVLH proposes that high vocal learning is a preadaptation for spontaneous, sporadic BPS to periodic auditory rhythms. This is because:

(i) high vocal learning provides intrinsic rewards for predicting the temporal structure of complex auditory sequences, because such predictions scaffold vocal learning of such sequences;

(ii) temporal predictions about auditory periodicities in the hundreds-of-milliseconds range are made via action-like neural processes in forebrain motor planning regions; and

(iii) in high vocal learners, these motor planning regions are in tight reciprocal communication with forebrain auditory regions throughout life.

The first reason above is based on cognitive research on statistical learning, a form of implicit learning without external reinforcement which involves detecting patterns and regularities in the environment [125]. Research suggests that statistical learning of auditory sequences is an active process in which predictions are continuously formed and compared to incoming input in order to update mental models of sequence structure [125,126]. Santolin & Saffran [127] point out that statistical learning of auditory patterns by humans begins in infancy as part of language acquisition and also occurs in nonhuman animals. Notably, they argue that 'statistical learning is likely to drive vocal learning in organisms that must learn to produce structured vocalizations' (p. 59). These authors also discuss research indicating that parrots surpass songbirds in the ability to learn underlying patterns in sound sequences (see also [128]). Point (i) above asserts that part of the cognitive system of high vocal learners is a mechanism that provides them with an intrinsic reward for accurately predicting the temporal structure of such sequences, to facilitate learning of this structure. This relates to an idea raised later in this paper, namely that selection

on the motivation for BPS can be conceptually distinguished from the selection on the capacity for BPS.

The second reason in the above list is based on recent theorizing about the role of the motor system in predicting timing in periodic auditory rhythms [41]. This theorizing draws heavily on neurophysiological research on rhythmic timing in nonhuman primates (e.g. [74,129,130]). The involvement of the motor system in such predictions, combined with the above idea of intrinsic rewards for such predictions, provides an explanation for spontaneous rhythmic movement to auditory rhythms in high vocal learners.

The third reason in the above list is based on the importance of strong reciprocal connections between forebrain premotor and auditory regions for high vocal learning and for BPS. Circuit-level research on the role of such connections in vocal learning has only been conducted in songbirds, who have complex vocal learning [131]. Because high vocal learners surpass complex vocal learners in vocal flexibility and in the neural complexity of their vocal learning system (cf. §4), the strength and plasticity of such auditory–motor interactions is probably stronger in high vocal learners than in complex vocal learners. The rVLH argues that these strong connections allow rapid two-way communication between forebrain auditory and motor planning regions, scaffolding the spontaneous predictive movements to auditory rhythms seen in high vocal learners.

The rVLH's claim that high vocal learning is a preadaptation for spontaneous, sporadic BPS to periodic auditory rhythms requires clarification of some terms. The first is 'preadaptation', which is 'an evolutionary change that adapts organisms to one set of environmental conditions but in addition and quite fortuitously positions them for a new surge in adaptive evolution' [132, p. 13]. Examples of preadaptation abound in evolutionary biology, as evidenced by research in palaeontology, developmental biology and genetics [133]. Feathers, for example, were a preadaptation for flight, evolving in therapod dinosaurs long before flight evolved in their avian descendants [134]. Feathers originally served non-flight-related functions such as thermal insulation, conferring flight-related aerodynamic benefits as a fortuitous consequence of their structure. Only later were the flight-related properties of feathers a direct target of natural selection, resulting in changes in feather structure supporting powered flight [135]. Just as the aerodynamic properties of early feathers were a fortuitous byproduct of their structure, the rVLH suggests that a capacity for sporadic BPS was a fortuitous byproduct of the neural circuitry for high vocal learning.

A second term requiring clarification is 'sporadic.' Sporadic BPS is the type of BPS observed in parrots, whereby rhythmic movements are phase aligned to an auditory beat during sporadic 'bouts' of several seconds surrounded by stretches of little movement or of rhythmic movement not synchronized to the beat. This is what was observed in Snowball, who exhibited BPS in bouts with a median of 16 head bobs, and tended to gravitate to a head bob tempo near 126 BPM during unsynchronized movement to music [55,136]. When presented with 10 different novel tempi relative to the original musical tempo of 109 BPM, Snowball synchronized in a phase-aligned manner at nine of these tempi spanning 98–130 BPM [55]. While statistical analyses showed that this amount of synchronization was very unlikely to happen by chance, Snowball's sporadic BPS is distinct from the sustained BPS observed in adult humans, where rhythmic movements

remain phase aligned to a beat for much longer periods, even in musically untrained individuals [72]. Interestingly, sporadic BPS may be more representative of how young children move to music [9,137].

The third term that requires clarification is 'spontaneous', meaning BPS that emerges without explicit instruction or physical rewards, as observed in humans and in parrots. Snowball the cockatoo, for example, was never explicitly trained to move rhythmically to music using food rewards, unlike the sea lion Ronan. Of course, parrots, like human children, often receive positive attention from human adults for BPS, and such attention is doubtless rewarding in species that form strong and lasting social bonds, as parrots and humans both do [138]. Indeed, such social rewards may amplify BPS behaviour. However, by focusing on *intrinsic* rewards, the rVLH entails the idea that attention and social rewards alone cannot account for spontaneous BPS to music. This distinguishes it from a proposal made by Wilson & Cook [51, p. 1655], who suggest that parrots engage in BPS because 'these birds bond with their caretakers and are highly sensitive to social reward, making it particularly likely that they will pick-up behaviours that humans find amusing'. A problem with this proposal is that dogs bond strongly with their carers and are highly sensitive to social reward [139], yet do not show spontaneous BPS to music [56,118].

A focus on the intrinsic rewards of temporal prediction in sequence processing bears a resemblance to a suggestion by Merker [140] that vocal learners have an intrinsic motivation for high-fidelity copying of sounds, because such copying is needed for vocal learning and is typically not reinforced by immediate external rewards. Merker refers to this motivational mechanism as a 'conformal motive' and suggests that in parrots this motivation to copy could extend to non-vocal body movements, leading them to imitate humans moving to the beat of music [105]. Consistent with this view, parrots can imitate non-vocal movements [141]. However, an intrinsic motivation to imitate non-vocal movements also seems to occur in chimpanzees. For example, a juvenile chimpanzee observing an adult crack a nut with an anvil stone and pounding stone will imitate the adult's actions without any reinforcement and often without success [142]. Yet despite this intrinsic motivation to imitate movements, among the cases of enculturated chimpanzees raised by humans and exposed to music during their development, there are no reports of spontaneous BPS in these animals [118].

Before closing this section, it is worth emphasizing that while the rVLH claims that the neural circuitry for high vocal learning is a *necessary* prerequisite for spontaneous BPS, it does not claim that it is *sufficient*. To date, parrots and humans are the only species known to engage in BPS spontaneously. In addition to high vocal learning, parrots and humans also share the ability to imitate non-vocal movements and a tendency to live in complex social groups and form long-term bonds [65,118]. The extent to which these factors also act as preadaptations for spontaneous BPS merits further research [143].

6. The primate heritage in the origins of human beat perception and synchronization

The rVLH suggests that a brain adapted for high vocal learning will fortuitously show a predisposition for spontaneous,

sporadic BPS. Yet for BPS to occur, exposure to rhythmic sound is needed. In the case of parrots living with humans, human music provides such rhythms. Assuming human ancestors had evolved high vocal learning, what would be the source of rhythmic sounds in their environment? Research on chimpanzees and bonobos, the closest living relatives of humans, shows that rhythmic vocalizations are part of their natural social behaviour. These vocalizations include pant hooting in chimpanzees and high-hooting in bonobos, and both species show evidence of temporal coordination with conspecifics when making these vocalizations [144–146]. Furthermore, short episodes of rhythmic drumming on tree buttresses are part of chimpanzee display behaviour in the wild [147], and an untrained chimpanzee in captivity has been filmed drumming steadily on a barrel for more than 30 s [148]. Thus, it seems plausible that rhythmically structured sounds produced in social contexts were present in the last common ancestor of humans and chimpanzees/bonobos [149]. The rVLH suggests that such sounds, in the context of a human ancestor that was a high vocal learner, could have led to spontaneous, sporadic BPS.

In light of this suggestion, a recent study by Hattori & Tomonaga [150,151] is of particular interest. These researchers found that enculturated chimpanzees exposed to complex rhythms made rhythmic rocking and swaying movements. These movements were not entrained to the beat and occurred whether the rhythms were beat-based or not, thus differing from BPS in important ways. Nevertheless, the movements seemed to reflect positive engagement, were made without any reinforcement and did not resemble distress responses or stereotyped behaviours sometimes seen in poorly treated animals. Furthermore, despite the fact that the chimpanzees were free to leave the testing area at any time, one male chimpanzee stayed closer to the sound source when the sounds were on versus off, suggesting attraction to the stimulus. Furthermore, this chimpanzee also made a few different types of rhythmic movements in response to rhythmic sounds, including head bobbing and hand clapping. When combined with field observations of chimpanzee 'rain dancing' (ritualized movements in response to loud sounds such as rain or waterfalls [152]), this study suggests that a predisposition to move rhythmically to loud, complex sound patterns may have been in place in human ancestors prior to the evolution of high vocal learning. In the framework of the rVLH, this would facilitate the occurrence of sporadic BPS once high vocal learning evolved.

7. Human beat perception and synchronization and gene-culture coevolution

There is growing evidence that gene-culture coevolution has shaped some important human biological traits. Convincing cases coming from biological adaptations to diet, including the evolution of lactose tolerance around 10 000 years ago in certain populations that practiced dairying, and much more ancient and species-wide anatomical changes associated with the control of fire and associated dietary changes [153–155]. In this section, I suggest that BPS has been the locus of cognitive gene-culture coevolution, with a gradual transition from sporadic to sustained BPS in human ancestors leading to evolved neural specializations for sustained BPS in humans. (Recall that in sporadic BPS, as observed in parrots,

accurate synchronization to a beat occurs in short bouts of a few seconds separated by stretches of little movement or of unsynchronized rhythmic movement. In sustained BPS, accurate synchronization is maintained over longer stretches of time, as seen in a dance around the world today.) The current proposal aligns with modern theories suggesting a prominent role for gene-culture coevolution in the emergence of human musicality (e.g. [13,156–159]).

The previous section argued that sporadic BPS arose fortuitously when the neural circuitry for high vocal learning evolved in a human ancestor that already had rhythmic social vocalizations. Below I suggest why a transition from sporadic to sustained BPS occurred via gene-culture coevolution, and then outline predictions of this view in terms of neuroscience, cross-species studies and genetics. I close this section by briefly discussing two important issues for future work in this area.

(a) Factors leading to gene-culture coevolution for beat perception and synchronization

For BPS to become a target of gene-culture coevolution, early humans would first have to employ sporadic BPS in a cultural behaviour that became widespread. Here, I would like to draw an analogy to the early use of fire use in human cultures. Fire use has a long history in the genus *Homo*, dating back at least 1.5 million years [160] and is widely believed to have led to changes in human biology via gene-culture coevolution, including anatomical and physiological adaptations to eating cooked food [154,161]. Current research suggests that fire use began with early humans taking advantage of 'fortuitous fire', such as from lightning strikes, sometimes transporting it to safe places such as caves [160]. The larger point is that fire use probably became widespread in human culture long before the ability to make fire at will and long before fire-driven gene-culture coevolution. This illustrates Richersen *et al.*'s [162, p. 8985] point that 'Culture normally evolves more rapidly than genes, creating novel environments that expose genes to new selective pressures'.

In the case of fire, the spread of fire use based on scavenging fire from natural sources is easy to understand owing to controlled fire's use, e.g. in keeping warm and fending off predators. It is less clear why early humans would use and spread behaviours based on sporadic BPS. If sporadic BPS first emerged in the context of rhythmic social vocalizations in an ancestor with high vocal learning, as suggested in §6 above, perhaps early humans used sporadic BPS for social purposes. For example, Mehr *et al.* [14] have suggested that synchronized rhythmic vocalizations and body movements could act as signals of coalition strength aimed at other groups, and Merker *et al.* [163] suggest synchronized rhythmic vocalizations could have a 'beacon-like' effect, attracting potential mates and deterring rivals from a distance. If sporadic BPS supported such behaviours and the behaviours led to advantages in survival or reproduction, then displays using sporadic BPS could have spread through human groups via purely cultural dynamics, e.g. via imitation of more successful groups.

Yet if this occurred, why would gene-culture coevolution favour a transition to sustained BPS? Fisher & Ridley [164, p. 930] have suggested that 'The smallest, most trivial new habit adopted by a hominid species could—if advantageous—have led to the selection of genomic variations that sharpened that habit'. A possible advantage of increasingly

sustained (versus sporadic) BPS in social contexts is that it scaffolds sustained interpersonal synchrony, which can in turn influence subsequent social behaviour [165]. A large body of empirical work finds that group members engaging in sustained interpersonal synchrony are subsequently more cooperative and feel more socially connected to each other [166–168], and may also show more in-group conformity and obedience [169]. This may partly reflect the blurring of self-other boundaries owing to the way sustained interpersonal synchrony interacts with neural mechanisms of action, perception and prediction in the brain [27,170]. If these social effects were more pronounced when episodes of interpersonal synchrony were more sustained, and facilitated behaviours outside of musical contexts that promoted survival, then gene-culture coevolution could favour genetic variants enhancing the capacity and proclivity for sustained BPS. Here, 'capacity' refers to the neural mechanisms supporting sustained BPS, while 'proclivity' refers to neural mechanisms that make sustained BPS rewarding, especially in social contexts. (Selection on the proclivity for BPS, independent of selection on capacity, has been suggested previously [61].) With these ideas in mind, I turn to the predictions that emerge from this proposal.

(b) Predictions of the hypothesis: neuroscience, cross-species studies and genetics

Several predictions emerge from the hypothesis that our current capacity for BPS is a product of gene-culture coevolution which refined a preadaptation for BPS rooted in high vocal learning. Because this hypothesis combines the revised rVLH with the claim of subsequent gene-culture coevolution, I will refer to it as the *rVLH*, using italics to distinguish it from the rVLH as outlined in §5 above, which makes no claim for gene-culture coevolution. Importantly, the predictions listed below are not made by the view that human rhythmic synchronization to a beat reflects widespread entrainment mechanisms with no special relation to the auditory–motor neural specializations of high vocal learners.

(i) Human neuroscience

The *rVLH* predicts significant overlap in the neural circuitry of BPS and high vocal learning in regions and pathways of the auditory dorsal stream, because BPS is seen as an evolutionary offshoot of high vocal learning. Such specialization could occur via evolutionary mechanisms of brain area duplication or brain pathway duplication/elaboration [116,171]. Note that 'overlap' does not mean 'identity'. Owing to evolved neural specializations, the *rVLH* suggests that there should be some neural populations or pathways which show selective involvement in BPS versus vocal learning. One way to conceptualize this combination of sharing and specialization is suggested by recent computational modelling work on the neural relationship between music and speech processing. In a study in which hierarchical artificial neural networks were optimized for speech and music recognition [172], the best-performing networks showed substantial shared processing at early stages, followed by separate speech and music regions and pathways at higher stages (figure 4). This organization is consistent with neural research suggesting that music and speech processing have significant overlap in subcortical and primary auditory
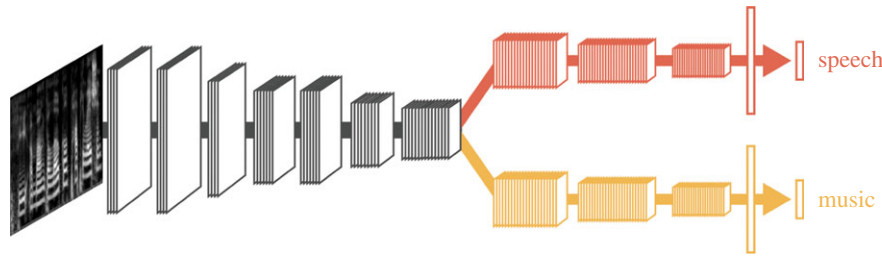
Figure 4. An optimized hierarchical artificial neural network for recognition of sounds as speech or music, after Kell *et al.* [172]. Auditory input is shown at the left (spectrogram-like representation of sound). Lower level processing stages shared by speech and music are shown in black and white, higher level stages and streams unique to each domain are shown in colour. (Online version in colour.)

cortical regions [173–175] yet also have neurally specialized processing in non-primary auditory cortex [176–178].

The *rVLH*'s prediction of neural overlap between BPS and high vocal learning is consistent with studies finding links between childhood speech-related abilities and non-linguistic beat processing [179–181]. At the same time, the fact that the *rVLH* posits evolved neural specializations for BPS is consistent with the existence of congenital 'beat deafness', in which individuals with normal hearing and musical pitch perception have severe problems perceiving and/or synchronizing with a musical beat [182,183].

The *rVLH* further predicts that human developmental neuroscience will reveal experience-expectant plasticity in circuits underlying the capacity for sustained BPS [184,185]. Finally, as noted at the end of the previous subsection, the *rVLH* suggests that natural selection acted not only on the capacity for sustained BPS, but also the proclivity or motivation to engage in this behaviour. The hypothesis thus calls for research on the relationship between BPS and reward-related activity in the brain. In particular, the *rVLH* suggests that the neural rewards humans experience when engaging in BPS-based interpersonal synchrony [186] may reflect evolved neural specializations in the striatum, which is involved both in beat perception and reward [41].

(ii) Cross-species studies

The *rVLH* predicts that the more advanced a species' vocal learning capacities (figure 1), the more its behavioural capacity for, and neural mechanisms of, BPS will resemble those of humans. Thus for example, it predicts that grey seals, which have recently been shown to have remarkable vocal mimicry abilities [96] will outperform less vocally flexible pinnipeds in studies of BPS. Similarly, it predicts that cetaceans with advanced vocal learning abilities, such as belugas and bottlenose dolphins [187,188], will outperform less vocally flexible cetaceans on tests of BPS. Counterintuitively, the *rVLH* predicts the brain mechanisms of rhythmic synchronization to a beat are more similar in humans and parrots than in humans and sea lions, owing to the convergently evolved similarities of parrots and humans in vocal learning capacities, and despite the much closer phylogenetic proximity of humans and sea lions.

In terms of neuroanatomy, because the *rVLH* posits that the advanced degree of auditory integration with motor circuits seen in high vocal learners is critical for spontaneous BPS and also posits that the auditory dorsal stream is a key site of this enhanced integration, it makes predictions for research on primate comparative neuroanatomy. Specifically, it predicts significant differences in auditory dorsal stream pathways important for BPS in humans and homologous pathways in other primates owing to the limited vocal learning abilities of those species. For example, the *rVLH* predicts that connections between the secondary auditory cortex and parietal cortex (red pathway linking superior temporal gyrus and parietal cortex in figure 3*c*, see §5) will be much stronger in humans than in monkeys or chimpanzees. This prediction can be tested using structural MRI methods such as diffusion tensor imaging, which are currently used to compare human, monkey and chimpanzee cortical connectivity [189,190].

(iii) Genetics

For BPS to be subject to gene-culture coevolution, it must have a heritable genetic substrate. Recently, a large-scale genome-wide association study with over 600 000 participants found 67 genetic loci associated with BPS, indicating a heritable, polygenic genetic substrate [20]. That is, BPS is a complex or 'polygenic' trait, in which interactions between genetic variants at different loci probably play an important role, rather than a 'Mendelian' trait influenced by variation at a single gene. The results of this study were virtually unchanged by controlling for more general tests of cognition, consistent with results from the twin literature which show that the genetics of rhythm are not solely attributable to general cognitive effects. Importantly, in this new study, genetic variance explained only about 13–16% of phenotypic variance in the beat synchronization trait, indicating that variance in BPS ability is genetically influenced but far from genetically determined.

The *rVLH* makes three distinct predictions about the genetic substrate of BPS. First, like the original VLH, it predicts genetic relationships between BPS and vocal learning. Supporting this prediction, new research [191] finds that the genetic architecture of BPS is significantly enriched for genes expressed during song production in songbird Area X, a key nucleus for vocal learning in avian brains, homologous to human basal ganglia. This is of interest as the basal ganglia plays an important role in human beat perception [33,41]. Second, since the *rVLH* proposes there was natural selection for sustained BPS after sporadic BPS emerged as a preadaptation, it predicts BPS will show independent genetic variation in humans after variance related to vocal learning abilities is accounted for. Third, the *rVLH* predicts that quantitative signatures of natural selection in the genome [192] will be found in the genetic substrate of BPS. In this regard, it is interesting that two loci associated with BPS in the study of Niarchou *et al.* [20] are in 'human accelerated regions', i.e. regions of the genome that are strikingly different from many other species, including our closest living primate relatives [193].

As a theory of gene-culture coevolution, the *rVLH* also motivates research on existing small-scale cultures which have traditionally had a very little collective, synchronized music-making [194]. In cases where such cultures have had limited gene flow with other groups, the rVLH predicts that individuals in those cultures will show signatures of relaxed selection on genetic variants supporting sustained BPS.

(c) Two issues for future research

A key issue for future research on BPS and gene-culture coevolution is a more complete spelling out of the proposed feedback cycles between culture and genes in the evolution of BPS. For example, relationships between collective musical synchrony, behaviour outside of musical contexts, cultural group selection[3] and individual-level selection remain to specified in detail. One goal of such theorizing is generating testable predictions distinct from those made by 'classical' evolutionary theories of musicality not involving interactions between cultural innovation and biological evolution.

Another issue for future work is broadening the lens when considering possible evolved neural specializations for BPS. The current paper has focused on forebrain circuitry as a site of such specializations, including cortical auditory–motor connections and motivation/reward circuitry in the striatum. Yet when humans engage in BPS, these forebrain circuits interact with lower brain regions, creating loops that include the midbrain, cerebellum, brainstem and spinal cord. Evolutionary changes relevant to BPS may not be restricted the forebrain, and future work will profit from an examination of BPS-relevant regions and connections throughout these loops [29,30,195].

8. Conclusion

This paper has reviewed and revised the hypothesis that the evolutionary foundations of human BPS lie in our capacity for complex vocal learning. On the basis of behavioural and neural research on vocal learning and on beat processing, the paper argues that an advanced form of vocal learning, convergently evolved in humans and parrots, acts as a preadaptation

for sporadic BPS. It further argues that in humans, mechanisms of gene-culture coevolution transformed this preadaptation into a genuine neural adaptation for sustained BPS. This larger significance of this proposal for evolutionary biology is that it outlines a scenario of *cognitive* gene-culture coevolution which makes testable predictions for neuroscience, cross-species research and genetics. To date, convincing examples of gene-culture coevolution in humans come from studies of non-neural physiology or anatomy, e.g. related to digestion and diet. While there is great interest and promise in the idea that gene-culture coevolution has shaped human mental abilities, including language [155,162,164,196–198], we currently lack compelling biological evidence for cognitive gene-culture coevolution. The study of BPS has the potential to provide such evidence and could thus pave the way for other studies examining how biology and culture intertwined in the evolution of the human mind.

Endnotes

[1]Elephants in sanctuaries have been reported to synchronize rhythmic movements to an auditory beat [56], but as noted by the lead author of that study, it is not clear if this behaviour was spontaneous or trained, and tempo flexibility was not tested [61]. Elephants are vocal learners, but their exact degree of vocal flexibility is not yet known [62].

[2]Interestingly, there do not appear to be major differences in the vocal learning circuitry of male and female parrots [64], and both male and female parrots have been found to synchronize rhythmic movements to a musical beat [118].

[3]Note that the revised vocal learning hypothesis is not based on biological group selection theory.

References

1. Honing H, ten Cate C, Peretz L, Trehub SE. 2015 Without it no music: cognition, biology and evolution of musicality. *Phil. Trans. R. Soc. B* **370**, 20140088. (doi:10.1098/rstb.2014.0088)

2. Patel, A D. 2008 *Music, language, and the brain*. New York, NY: Oxford University Press.

3. Savage PE, Brown S, Sakai E, Currie TE. 2015 Statistical universals reveal the structures and functions of human music. *Proc. Natl Acad. Sci. USA* **112**, 8987–8992. (doi:10.1073/pnas.1414495112)

4. Nettl B. 2015 *The study of ethnomusicology: thirty-three discussions*, 3rd edn. Urbana, IL: University of Illinois Press.

5. Winkler I, Háden GP, Ladinig O, Sziller I, Honing H. 2009 Newborn infants detect the beat in music. *Proc. Natl Acad. Sci. USA* **106**, 2468–2471. (doi:10.1073/pnas.0809035106)

6. Kirschner S, Tomasello M. 2009 Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* **102**, 299–314. (doi:10.1016/j.jecp.2008.07.005)

7. Zentner M, Eerola T. 2010 Rhythmic engagement with music in infancy. *Proc. Natl Acad. Sci. USA* **107**, 5768–5773. (doi:10.1073/pnas.1000121107)

8. Fujii S, Watanabe H, Oohashi H, Hirashima M, Nozaki D, Taga G. 2014 Precursors of dancing and singing to music in three-to four-months-old infants. *PLoS ONE* **9**, e97680. (doi:10.1371/journal.pone.0097680)

9. Cirelli LK, Trehub SE. 2019 Dancing to metallica and dora: case study of a 19-month-old. *Front. Psychol.* **10**, 1073. (doi:10.3389/fpsyg.2019.01073)

10. Darwin C. 1871 *The descent of man and selection in relation to sex*. London, UK: John Murray.

11. James W. 1890 *The principles of psychology*. New York, NY: Henry Holt and Co.

12. Dunbar RI. 2012 On the evolutionary function of song and dance. In *Music, language, and human evolution* (ed. N Bannan), pp. 201–214. Oxford, UK: Oxford University Press.

13. Savage PE, Loui P, Tarr B, Schachner A, Glowacki L, Mithen S, Fitch WT. In press. Music as a coevolved system for social bonding. *Behav. Brain Sci.*

14. Mehr S, Krasnow M, Bryant GA, Hagen E. In press. Origins of music in credible signaling. *Behav. Brain Sci.*

15. Sperber D. 1996 *Explaining culture: a naturalistic approach*. Oxford, UK: Blackwell.

16. Pinker S. 1997 *How the mind works*. New York, NY: W.W. Norton and Co.

17. Marcus GF. 2012 Musicality: instinct or acquired skill? *Topics Cogn. Sci.* **4**, 498–512. (doi:10.1111/j.1756-8765.2012.01220.x)

18. Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT. 2015 Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B* **370**, 20140093. (doi:10.1098/rstb.2014.0093)

19. Patel AD. 2019 Evolutionary music cognition: cross-species studies. In *Foundations in music psychology: theory and research* (eds J Rentfrow, D Levitin), pp. 459–501. Cambridge, MA: MIT Press.

20. Niarchou M *et al.* 2021 Unravelling the genetic architecture of musical rhythm. *BioRxiv*, 836197.

21. Kirschner S, Ilari B. 2014 Joint drumming in Brazilian and German preschool children: cultural differences in rhythmic entrainment, but no prosocial effects. *J. Cross-Cult. Psychol.* **45**, 137–166. (doi:10.1177/0022022113493139)

22. Patel AD, Iversen JR, Chen Y, Repp BH. 2005 The influence of metricality and modality on synchronization with a beat. *Exp. Brain Res.* **163**, 226–238. (doi:10.1007/s00221-004-2159-8)

23. MacDougall HG, Moore ST. 2005 Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *J. Appl. Physiol.* **99**, 1164–1173. (doi:10.1152/japplphysiol.00138.2005)

24. Repp BH, Su YH. 2013 Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* **20**, 403–452. (doi:10.3758/s13423-012-0371-2)

25. McNeill WH. 1995 *Keeping together in time: dance and drill in human history*. Cambridge, MA: Harvard University Press.

26. Turino T. 2008 *Music as social life: the politics of participation*. Chicago, IL: University of Chicago Press.

27. Tarr B, Launay J, Dunbar RI. 2014 Music and social bonding: 'self-other' merging and neurohormonal mechanisms. *Front. Psychol.* **5**, 1096. (doi:10.3389/fpsyg.2014.01096)

28. Stupacher J, Hove MJ, Novembre G, Schütz-Bosbach S, Keller PE. 2013 Musical groove modulates motor cortex excitability: a TMS investigation. *Brain Cogn.* **82**, 127–136. (doi:10.1016/j.bandc.2013.03.003)

29. Ross JM, Warlaumont AS, Abney DH, Rigoli LM, Balasubramaniam R. 2016 Influence of musical groove on postural sway. *J. Exp. Psychol.* **42**, 308–319. (doi:10.1037/xhp0000198)

30. Coste A, Salesse RN, Gueugnon M, Marin L, Bardy BG. 2018 Standing or swaying to the beat: discrete auditory rhythms entrain stance and promote postural coordination stability. *Gait Posture* **59**, 28–34. (doi:10.1016/j.gaitpost.2017.09.023)

31. Grahn JA, Brett M. 2007 Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* **19**, 893–906. (doi:10.1162/jocn.2007.19.5.893)

32. Chen JL, Penhune VB, Zatorre RJ. 2008 Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* **18**, 2844–2854. (doi:10.1093/cercor/bhn042)

33. Grahn JA, Rowe JB. 2009 Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* **29**, 7540–7548. (doi:10.1523/JNEUROSCI.2018-08.2009)

34. Matthews TE, Witek MA, Lund T, Vuust P, Penhune VB. 2020 The sensation of groove engages motor and reward networks. *Neuroimage* **214**, 116768. (doi:10.1016/j.neuroimage.2020.116768)

35. Chen JL, Penhune VB, Zatorre RJ. 2008 Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* **20**, 226–239. (doi:10.1162/jocn.2008.20018)

36. Kung SJ, Chen JL, Zatorre RJ, Penhune VB. 2013 Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* **25**, 401–420. (doi:10.1162/jocn_a_00325)

37. Arnal LH. 2012 Predicting 'when' using the motor system's beta-band oscillations. *Front. Hum. Neurosci.* **6**, 225. (doi:10.3389/fnhum.2012.00225)

38. Patel AD, Iversen JR. 2014 The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* **8**, 57. (doi:10.3389/fnsys.2014.00057)

39. Morillon B, Baillet S. 2017 Motor origin of temporal predictions in auditory attention. *Proc. Natl Acad. Sci. USA* **114**, E8913–E8921. (doi:10.1073/pnas.1705373114)

40. Proksch S, Comstock DC, Médé, B., Pabst A, Balasubramaniam R. 2020 Motor and predictive processes in auditory beat and rhythm perception. *Front. Hum. Neurosci.* **14**, 375. (doi:10.3389/fnhum.2020.578546)

41. Cannon JJ, Patel AD. 2021 How beat perception co-opts motor neurophysiology. *Trends Cogn. Sci.* **25**, 137–150. (doi:10.1016/j.tics.2020.11.002)

42. Large EW. 2008 Resonating to musical rhythm: theory and experiment. In *The psychology of time* (ed. S Grondin), pp. 189–232. Bingham, UK: Emerald Group Publishing.

43. Large EW, Herrera JA, Velasco MJ. 2015 Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* **9**, 159. (doi:10.3389/fnsys.2015.00159)

44. Honma S. 2018 The mammalian circadian system: a hierarchical multi-oscillator structure for generating circadian rhythm. *J. Physiol. Sci.* **68**, 207–219. (doi:10.1007/s12576-018-0597-5)

45. Ermentrout B. 1991 An adaptive model for synchrony in the firefly *Pteroptyx malaccae*. *J. Math. Biol.* **29**, 571–585. (doi:10.1007/BF00164052)

46. Giraud AL, Poeppel D. 2012 Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511. (doi:10.1038/nn.3063)

47. Buzsáki G. 2019 *The brain from inside out*. New York, NY: Oxford University Press.

48. Haegens S. 2020 Entrainment revisited: a commentary on Meyer, Sun, and Martin (2020). *Lang. Cogn. Neurosci.* **35**, 1119–1123. (doi:10.1080/23273798.2020.1758335)

49. Fitch WT. 2012 The biology and evolution of rhythm: unraveling a paradox. In *Language and music as cognitive systems* (eds P Rebuschat *et al.*), pp. 73–95. Oxford, UK: Oxford University Press.

50. Patel AD. 2014 The evolutionary biology of musical rhythm: was Darwin wrong?. *PLoS Biol.* **12**, e1001821. (doi:10.1371/journal.pbio.1001821)

51. Wilson M, Cook PF. 2016 Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bull. Rev.* **23**, 1647–1659. (doi:10.3758/s13423-016-1013-x)

52. Cook P, Rouse A, Wilson M, Reichmuth C. 2013 A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* **127**, 412. (doi:10.1037/a0032345)

53. Patel AD. 2006 Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* **24**, 99–104. (doi:10.1525/mp.2006.24.1.99)

54. Tyack PL. 2020 A taxonomy for vocal learning. *Phil. Trans. R. Soc. B* **375**, 20180406. (doi:10.1098/rstb.2018.0406)

55. Patel AD, Iversen JR, Bregman MR, Schulz I. 2009 Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* **19**, 827–830. (doi:10.1016/j.cub.2009.03.038)

56. Schachner A, Brady TF, Pepperberg IM, Hauser MD. 2009 Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* **19**, 831–836. (doi:10.1016/j.cub.2009.03.061)

57. Petkov CI, Jarvis E. 2012 Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12. (doi:10.3389/fnevo.2012.00012)

58. Arriaga G, Zhou EP, Jarvis ED. 2012 Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE* **7**, e46610. (doi:10.1371/journal.pone.0046610)

59. Wirthlin M *et al.* 2019 A modular approach to vocal learning: disentangling the diversity of a complex behavioral trait. *Neuron* **104**, 87–99. (doi:10.1016/j.neuron.2019.09.036)

60. Jarvis ED. 2019 Evolution of vocal learning and spoken language. *Science* **366**, 50–54. (doi:10.1126/science.aax0287)

61. Schachner A. 2013 The origins of human and avian auditory-motor entrainment. *Nova Acta Leopold* **111**, 243–253.

62. Stoeger AS, Manger P. 2014 Vocal learning in elephants: neural bases and adaptive context. *Curr. Opin Neurobiol.* **28**, 101–107. (doi:10.1016/j.conb.2014.07.001)

63. Nicolai J, Gundacker C, Teeselink K, Güttinger HR. 2014 Human melody singing by bullfinches (*Pyrrhula pyrrula*) gives hints about a cognitive note sequence processing. *Anim. Cogn.* **17**, 143–155. (doi:10.1007/s10071-013-0647-6)

64. Chakraborty M *et al.* 2015 Core and shell song systems unique to the parrot brain. *PLoS ONE* **10**, e0118496. (doi:10.1371/journal.pone.0118496)

65. Patel A, Iversen J, Bregman M.R., Schulz I. 2009 Studying synchronization to a musical beat in

nonhuman animals. *Ann. NY Acad. Sci.* **1169**, 459–469. (doi:10.1111/j.1749-6632.2009.04581.x)

66. Walker TJ. 1969 Acoustic synchrony: two mechanisms in the snowy tree cricket. *Science* **166**, 891–894. (doi:10.1126/science.166.3907.891)

67. Hartbauer M, Römer H. 2016 Rhythm generation and rhythm perception in insects: the evolution of synchronous choruses. *Front. Neurosci.* **10**, 223. (doi:10.3389/fnins.2016.00223)

68. Greenfield MD, Marin-Cudraz T, Party V. 2017 Evolution of synchronies in insect choruses. *Biol. J. Linn. Soc.* **122**, 487–504. (doi:10.1093/biolinnean/blx096)

69. Sismondo E. 1990 Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* **249**, 55–58. (doi:10.1126/science.249.4964.55)

70. Hartbauer M, Kratzer S, Steiner K, Römer H. 2005 Mechanisms for synchrony and alternation in song interactions of the bushcricket *Mecopoda elongata* (Tettigoniidae: Orthoptera). *J. Comp. Physiol. A* **191**, 175–188. (doi:10.1007/s00359-004-0586-4)

71. Nityananda V, Balakrishnan R. 2007 Synchrony during acoustic interactions in the bushcricket *Mecopoda* 'Chirper' (Tettigoniidae: Orthoptera) is generated by a combination of chirp-by-chirp resetting and change in intrinsic chirp rate. *J. Comp. Physiol. A* **193**, 51–65. (doi:10.1007/s00359-006-0170-1)

72. Burger B, Thompson MR, Luck G, Saarikallio SH, Toiviainen P. 2014 Hunting for the beat in the body: on period and phase locking in music-induced movement. *Front. Hum. Neurosci.* **8**, 903. (doi:10.3389/fnhum.2014.00903)

73. Zarco W, Merchant H, Prado L, Mendez JC. 2009 Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J. Neurophysiol.* **102**, 3191–3202. (doi:10.1152/jn.00066.2009)

74. Gámez J, Mendoza G, Prado L, Betancourt A, Merchant H. 2019 The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS Biol.* **17**, e3000054. (doi:10.1371/journal.pbio.3000054)

75. Gámez J, Yc K, Ayala YA, Dotov D, Prado L, Merchant H. 2018 Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Ann. NY Acad. Sci.* **1423**, 396–414. (doi:10.1111/nyas.13671)

76. Merchant H, Honing H. 2014 Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* **7**, 274. (doi:10.3389/fnins.2013.00274)

77. Honing H, Merchant H, Haden GP, Prado L, Bartolo R. 2012 Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE* **7**, e51369. (doi:10.1371/journal.pone.0051369)

78. Honing H, Bouwer FL, Prado L, Merchant H. 2018 Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: additional support for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* **12**, 475. (doi:10.3389/fnins.2018.00475)

79. Fuchs AF. 1967 Periodic eye tracking in the monkey. *J. Physiol.* **193**, 161–171. (doi:10.1113/jphysiol.1967.sp008349)

80. Takeya R, Kameda M, Patel AD, Tanaka M. 2017 Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Sci. Rep.* **7**, 1–12. (doi:10.1038/s41598-016-0028-x)

81. Katsu N, Yuki S, Okanoya K. 2021 Production of regular rhythm induced by external stimuli in rats. *Anim. Cogn.* 1–9. (doi:10.1007/s10071-021-01505-4)

82. Hattori Y, Tomonaga M, Matsuzawa T. 2013 Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Sci. Rep.* **3**, 1566. (doi:10.1038/srep01566)

83. Large EW, Gray PM. 2015 Spontaneous tempo and rhythmic entrainment in a bonobo (*Pan paniscus*). *J. Comp. Psychol.* **129**, 317. (doi:10.1037/com0000011)

84. Bregman MR, Iversen JR, Lichman D, Reinhart M, Patel AD. 2013 A method for testing synchronization to a musical beat in domestic horses (*Equus ferus caballus*). *Emp. Musicol. Rev.* **7**, 144–156. (doi:10.18061/emr.v7i3-4.3745)

85. Fitzroy AB, Lobdell L, Norman S, Bolognese L, Patel AD, Breen M. 2018 Horses do not spontaneously engage in tempo-flexible synchronization to a musical beat. Poster presented at the 15th Int. Conf. on Music Perception and Cognition, Montreal, QC.

86. Bispham J. 2006 Rhythm in music: what is it? Who has it? And why? *Music Percept.* **24**, 125–134. (doi:10.1525/mp.2006.24.2.125)

87. Ammirante P, Patel AD, Russo FA. 2016 Synchronizing to auditory and tactile metronomes: a test of the auditory-motor enhancement hypothesis. *Psychon. Bull. Rev.* **23**, 1882–1890. (doi:10.3758/s13423-016-1067-9)

88. Gilmore SA, Russo FA. 2021 Neural and behavioral evidence for vibrotactile beat perception and bimodal enhancement. *J. Cogn. Neurosci.* **33**, 635–650. (doi:10.1162/jocn_a_01673)

89. Janik VM, Slater PJ. 1997 Vocal learning in mammals. *Ad. Study Behav.* **26**, 59–100. (doi:10.1016/S0065-3454(08)60377-0)

90. Jarvis ED. 2004 Learned birdsong and the neurobiology of human language. *Ann. NY Acad. Sci.* **1016**, 749. (doi:10.1196/annals.1298.038)

91. Grahn J. 2004 Behavioural and functional imaging studies of rhythm processing. Unpublished doctoral dissertation, University of Cambridge, Cambridge, UK.

92. Hasegawa A, Okanoya K, Hasegawa T, Seki Y. 2011 Rhythmic synchronization tapping to an audio–visual metronome in budgerigars. *Sci. Rep.* **1**, 120. (doi:10.1038/srep00120)

93. Luescher AU. (ed.) 2006 *Manual of parrot behavior*. Ames, IA: Blackwell.

94. Patel AD. 2010 Music, biological evolution, and the brain. In *Emerging disciplines* (ed. M Bailar), pp. 91–144. Houston, TX: Rice University Press.

95. Reichmuth C, Casey C. 2014 Vocal learning in seals, sea lions, and walruses. *Curr. Opin Neurobiol.* **28**, 66–71. (doi:10.1016/j.conb.2014.06.011)

96. Stansbury AL, Janik VM. 2019 Formant modification through vocal production learning in gray seals.

Curr. Biol. **29**, 2244–2249. (doi:10.1016/j.cub.2019.05.071)

97. Ralls K, Fiorelli P, Gish S. 1985 Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Can. J. Zool.* **63**, 1050–1056. (doi:10.1139/z85-157)

98. Hage SR, Gavrilov N, Nieder A. 2016 Developmental changes of cognitive vocal control in monkeys. *J. Exp. Biol.* **219**, 1744–1749. (doi:10.1242/jeb.137653)

99. Gavrilov N, Hage SR, Nieder A. 2017 Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell Rep.* **21**, 2393–2406. (doi:10.1016/j.celrep.2017.10.107)

100. Zhang YS, Ghazanfar AA. 2020 A hierarchy of autonomous systems for vocal production. *Trends Neurosci.* **43**, 115–126. (doi:10.1016/j.tins.2019.12.006)

101. Sakata JT, Yazaki-Sugiyama Y. 2020 Neural circuits underlying vocal learning in songbirds. In *The neuroethology of birdsong* (eds J Sataka, S Woolley), pp. 29–63. Cham, Switzerland: Springer.

102. Schachner A. 2013 If horses entrain, don't entirely reject vocal learning: an experience-based vocal learning hypothesis. *Emp. Musicol. Rev.* **7**, 157–159. (doi:10.18061/emr.v7i3-4.3748)

103. Reh RK, Dias BG, Nelson CA, Kaufer D, Werker JF, Kolb B, Levibne JD, Hensch TK. 2020 Critical period regulation across multiple timescales. *Proc. Natl Acad. Sci. USA.* **110**, 3101–3106.

104. Cook PF et al. 2021 An MRI protocol for anatomical and functional evaluation of the California sea lion brain. *J. Neurosci. Methods* **353**, 109097. (doi:10.1016/j.jneumeth.2021.109097)

105. Merker B, Morley I, Zuidema W. 2018 Five fundamental constraints on theories of the origins of music. In *The origins of musicality* (ed. H Honing), pp. 49–80. Cambridge, MA: MIT Press.

106. Strogatz SH. 2003 *Sync: how order emerges from chaos in the universe, nature, and daily life*. New York, NY: Hyperion.

107. Rouse AA, Cook PF, Large EW, Reichmuth C. 2016 Beat keeping in a sea lion as coupled oscillation: implications for comparative understanding of human rhythm. *Front. Neurosci.* **10**, 257. (doi:10.3389/fnins.2016.00257)

108. Ravignani A, Fitch W, Hanke FD, Heinrich T, Hurgitsch B, Kotz SA, Scharff C, Stoger AS, de Boer B. 2016 What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* **10**, 274. (doi:10.3389/fnins.2016.00274)

109. Ravignani A, Kello CT, de Reus K, Kotz SA, Dalla Bella S, Méndez-Aróstegui M, Rapado-Tamarit B, Rubio-Garica A, de Boer B. 2019 Ontogeny of vocal rhythms in harbor seal pups: an exploratory study. *Cur. Zool.* **65**, 107–120. (doi:10.1093/cz/zoy055)

110. Nieder A, Mooney R. 2020 The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Phil. Trans. R. Soc. B* **375**, 20190054. (doi:10.1098/rstb.2019.0054)

111. Fischer J, Hammerschmidt K. 2020 Towards a new taxonomy of primate vocal production learning. *Phil. Trans. R. Soc. B* **375**, 20190045. (doi:10.1098/rstb.2019.0045)

112. Mitani J, Gros-Louis J. 1998 Chorusing and call convergence in chimpanzees: tests of three hypotheses. *Behaviour* **135**, 1041–1064. (doi:10.1163/156853998792913483)

113. Beecher MD, Brenowitz EA. 2005 Functional aspects of song learning in songbirds. *Trends Ecol. Evol.* **20**, 143–149. (doi:10.1016/j.tree.2005.01.004)

114. Miller CT, Thomas AW. 2012 Individual recognition during bouts of antiphonal calling in common marmosets. *J. Comp. Physiol. A* **198**, 337–346. (doi:10.1007/s00359-012-0712-7)

115. Okobi DE, Banerjee A, Matheson AM, Phelps SM, Long MA. 2019 Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988. (doi:10.1126/science.aau9480)

116. Chakraborty M, Jarvis ED. 2015 Brain evolution by brain pathway duplication. *Phil. Trans. R. Soc. B* **370**, 20150056. (doi:10.1098/rstb.2015.0056)

117. Ohms VR, Beckers GJ, Ten Cate C, Suthers RA. 2012 Vocal tract articulation revisited: the case of the monk parakeet. *J. Exp. Biol.* **215**, 85–92. (doi:10.1242/jeb.064717)

118. Schachner A. 2010 Auditory-motor entrainment in vocal mimicking species: additional ontogenetic and phylogenetic factors. *Commun. Integr. Biol.* **3**, 290–293. (doi:10.4161/cib.3.3.11708)

119. Hickok G, Poeppel D. 2007 The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402. (doi:10.1038/nrn2113)

120. Dichter BK, Breshears JD, Leonard MK, Chang EF. 2018 The control of vocal pitch in human laryngeal motor cortex. *Cell* **174**, 21–31. (doi:10.1016/j.cell.2018.05.016)

121. Chartier J, Anumanchipalli GK, Johnson K, Chang EF. 2018 Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* **98**, 1042–1054. (doi:10.1016/j.neuron.2018.04.031)

122. Gierhan SM. 2013 Connections for auditory language in the human brain. *Brain Lang.* **127**, 205–221. (doi:10.1016/j.bandl.2012.11.002)

123. Ross JM, Iversen JR, Balasubramaniam R. 2018 The role of posterior parietal cortex in beat-based timing perception: a continuous theta burst stimulation study. *J. Cogn. Neurosci.* **30**, 634–643. (doi:10.1162/jocn_a_01237)

124. Lewis JW, Van Essen DC. 2000 Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137. (doi:10.1002/1096-9861(20001204)428:1<112::AID-CNE8>3.0.CO;2-9)

125. Saffran JR. 2020 Statistical language learning in infancy. *Child Dev. Perspect.* **14**, 49–54. (doi:10.1111/cdep.12355)

126. Sherman BE, Graves KN, Turk-Browne NB. 2020 The prevalence and importance of statistical learning in human cognition and behavior. *Curr. Opin. Behav. Sci.* **32**, 15–20. (doi:10.1016/j.cobeha.2020.01.015)

127. Santolin C, Saffran JR. 2018 Constraints on statistical learning across species. *Trends Cogn. Sci.* **22**, 52–63. (doi:10.1016/j.tics.2017.10.003)

128. Tu HW, Dooling RJ. 2012 Perception of warble song in budgerigars (*Melopsittacus undulatus*): evidence for special processing. *Anim. Cogn.* **15**, 1151–1159. (doi:10.1007/s10071-012-0539-1)

129. Merchant H, Averbeck BB. 2017 The computational and neural basis of rhythmic timing in medial premotor cortex. *J. Neurosci.* **37**, 4552–4564. (doi:10.1523/JNEUROSCI.0367-17.2017)

130. Balasubramaniam R, Haegens S, Jazayeri M, Merchant H, Sternad D, Song JH. 2020 Neural encoding and representation of time for sensorimotor control and learning. *J. Neurosci.* **41**, 866–872. (doi:10.1523/JNEUROSCI.1652-20.2020)

131. Roberts TF, Hisey E, Tanaka M, Kearney MG, Chattree G, Yang CF, Shah NM, Mooney R. 2017 Identification of a motor-to-auditory pathway important for vocal learning. *Nat. Neurosci.* **20**, 978. (doi:10.1038/nn.4563)

132. Lumsden CJ, Wilson EO. 1983 *Promethean fire: reflections on the origins of mind*. Cambridge, MA: Harvard University Press.

133. Shubin N. 2020 *Some assembly required: decoding four billion years of life, from ancient fossils to DNA*. New York, NY: Pantheon.

134. Prum RO, Brush AH. 2002 The evolutionary origin and diversification of feathers. *Q. Rev. Biol.* **77**, 261–295. (doi:10.1086/341993)

135. Pan Y *et al.* 2019 The molecular evolution of feathers with direct evidence from fossils. *Proc. Natl Acad. Sci. USA* **116**, 3018–3023. (doi:10.1073/pnas.1815703116)

136. Patel AD, Iversen JR, Bregman MR, Schulz I. 2009 Avian and human movement to music: two further parallels. *Commun. Integr. Biol.* **2**, 485–488. (doi:10.4161/cib.2.6.9373)

137. Eerola T, Luck G, Toiviainen P. 2006 An investigation of preschoolers' corporeal synchronization with music. In *Proceedings of the 9th international conference on music perception & cognition (ICMPC9)* (eds M Baroni *et al.*), pp. 472–476. Bologna, Italy: ESCOM. (doi:10.1101/062703)

138. Spoon TR. 2006 Parrot reproductive behavior, or who associates, who mates, and who cares. In *Manual of parrot behavior* (ed. A Leuscher), pp. 63–77. Ames, IA: Blackwell Publishing.

139. Cook PF, Prichard A, Spivak M, Berns GS. 2016 Awake canine fMRI predicts dogs' preference for praise vs food. *Soc. Cogn. Affect. Neurosci.* **11**, 1853–1862.

140. Merker B. 2005 The conformal motive in birdsong, music, and language: an introduction. *Ann. NY Acad. Sci.* **1060**, 17–28. (doi:10.1196/annals.1360.003)

141. Heyes C, Saggerson A. 2002 Testing for imitative and nonimitative social learning in the budgerigar using a two-object/two-action test. *Anim. Behav.* **64**, 851–859. (doi:10.1006/anbe.2003.2002)

142. Inoue-Nakamura N, Matsuzawa T. 1997 Development of stone tool use by wild chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **111**, 159–173. (doi:10.1037/0735-7036.111.2.159)

143. Laland K, Wilkins C, Clayton N. 2016 The evolution of dance. *Curr. Biol.* **26**, R5–R9. (doi:10.1016/j.cub.2015.11.031)

144. Fedurek P, Machanda ZP, Schel AM, Slocombe KE. 2013 Pant hoot chorusing and social bonds in male chimpanzees. *Anim. Behav.* **86**, 189–196. (doi:10.1016/j.anbehav.2013.05.010)

145. Clay Z, Genty E. 2017 Natural communication in bonobos: insights into social awareness and the evolution of language. In *Bonobos: unique in mind, brain, and behavior* (eds B Hare, S Yamamoto), pp. 105–122. Oxford, UK: Oxford University Press.

146. Fedurek P, Schel AM, Slocombe KE. 2013 The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behav. Ecol. Sociobiol.*, **67**, 1781–1789. (doi:10.1007/s00265-013-1585-7)

147. Babiszewska M, Schel AM, Wilke C, Slocombe KE. 2015 Social, contextual, and individual factors affecting the occurrence and acoustic structure of drumming bouts in wild chimpanzees (*Pan troglodytes*). *Am. J. Phys. Anthropol.* **156**, 125–134. (doi:10.1002/ajpa.22634)

148. Dufour V, Poulin N, Curé, C., Sterck EH. 2015 Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Sci. Rep.* **5**, 11320. (doi:10.1038/srep11320)

149. Fitch WT. 2006 The biology and evolution of music: a comparative perspective. *Cognition* **100**, 173–215. (doi:10.1016/j.cognition.2005.11.009)

150. Hattori Y, Tomonaga M. 2020 Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proc. Natl Acad. Sci. USA* **117**, 936–942. (doi:10.1073/pnas.1910318116)

151. Hattori Y, Tomonaga M. 2021 Reply to Bertolo *et al.*: rhythmic swaying in chimpanzees has implications for understanding the biological roots of music and dance. *Proc. Natl Acad. Sci. USA* **118**, e2017986118. (doi:10.1073/pnas.2017986118)

152. Whiten A, Goodall J, McGrew W.C., Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C. 2001 Charting cultural variation in chimpanzees. *Behaviour* **138**, 1481–1516. (doi:10.1163/156853901317367717)

153. Segurel L, Guarino-Vignon P, Marchi N, Lafosse S, Laurent R, Bon C, Fabre A, Hegay T, Heyer E. 2020 Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS Biol.* **18**, e3000742. (doi:10.1371/journal.pbio.3000742)

154. Wrangham RW, Carmody RN. 2010 Human adaptation to the control of fire. *Evol. Anthropol.* **19**, 187–199. (doi:10.1002/evan.20275)

155. Laland KN, Odling-Smee J, Myles S. 2010 How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* **11**, 137–148. (doi:10.1038/nrg2734)

156. Cross I. 2003 Music and biocultural evolution. In *The Cultural study of music: a critical introduction* (eds M Clayton *et al.*), pp. 19–30. New York, NY: Routledge.

157. Tomlinson G. 2015 *A million years of music: the emergence of human modernity*. Cambridge, MA: MIT Press.

158. Podlipniak P. 2017 The role of the Baldwin effect in the evolution of human musicality. *Front. Neurosci.* **11**, 542. (doi:10.3389/fnins.2017.00542)

159. Patel AD. 2018 Music as a transformative technology of the mind: an update. In *The origins of*

musicality (ed. H Honing), pp. 113–126. Cambridge, MA: MIT Press.

160. Gowlett JA. 2016 The discovery of fire by humans: a long and convoluted process. *Phil. Trans. R. Soc. B* **371**, 20150164. (doi:10.1098/rstb.2015.0164)

161. Wrangham R. 2017 Control of fire in the Paleolithic: evaluating the cooking hypothesis. *Curr. Anthropol.* **58**, S303–S313. (doi:10.1086/692113)

162. Richerson PJ, Boyd R, Henrich J. 2010 Gene-culture coevolution in the age of genomics. *Proc. Natl Acad. Sci. USA* **107**(Suppl. 2), 8985–8992. (doi:10.1073/pnas.0914631107)

163. Merker BH, Madison GS, Eckerdal P. 2009 On the role and origin of isochrony in human rhythmic entrainment. *Cortex* **45**, 4–17. (doi:10.1016/j.cortex.2008.06.011)

164. Fisher SE, Ridley M. 2013 Culture, genes, and the human revolution. *Science* **340**, 929–930. (doi:10.1126/science.1236171)

165. Cirelli LK. 2018 How interpersonal synchrony facilitates early prosocial behavior. *Curr. Opin. Psychol.* **20**, 35–39. (doi:10.1016/j.copsyc.2017.08.009)

166. Rennung M, Göritz AS. 2016 Prosocial consequences of interpersonal synchrony: a meta-analysis. *Zeitschrift für Psychologie* **224**, 168–189. (doi:10.1027/2151-2604/a000252)

167. Mogan R, Fischer R, Bulbulia JA. 2017 To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J. Exp. Soc. Psychol.* **72**, 13–20. (doi:10.1016/j.jesp.2017.03.009)

168. Tarr B, Slater M, Cohen E. 2018 Synchrony and social connection in immersive virtual reality. *Sci. Rep.* **8**, 1–8. (doi:10.1038/s41598-018-21765-4)

169. Gelfand MJ, Caluori N, Jackson JC, Taylor MK. 2020 The cultural evolutionary trade-off of ritualistic synchrony. *Phil. Trans. R. Soc. B* **375**, 20190432. (doi:10.1098/rstb.2019.0432)

170. Seth AK. 2013 Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **17**, 565–573. (doi:10.1016/j.tics.2013.09.007)

171. Mars RB, Eichert N, Jbabdi S, Verhagen L, Rushworth MF. 2018 Connectivity and the search for specializations in the language-capable brain. *Curr. Opin. Behav. Sci.* **21**, 19–26. (doi:10.1016/j.cobeha.2017.11.001)

172. Kell AJ, Yamins DL, Shook EN, Norman-Haignere SV, McDermott JH. 2018 A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* **98**, 630–644. (doi:10.1016/j.neuron.2018.03.044)

173. Tierney AT, Krizman J, Kraus N. 2015 Music training alters the course of adolescent auditory development. *Proc. Natl Acad. Sci. USA* **112**, 10 062–10 067. (doi:10.1073/pnas.1505114112)

174. Norman-Haignere SV, McDermott JH. 2018 Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol.* **16**, e2005127. (doi:10.1371/journal.pbio.2005127)

175. Patel AD. 2014 Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* **308**, 98–108. (doi:10.1016/j.heares.2013.08.011)

176. Leaver AM, Rauschecker JP. 2010 Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* **30**, 7604–7612. (doi:10.1523/JNEUROSCI.0296-10.2010)

177. Norman-Haignere S, Kanwisher NG, McDermott JH. 2015 Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* **88**, 1281–1296. (doi:10.1016/j.neuron.2015.11.035)

178. Albouy P, Benjamin L, Morillon B, Zatorre RJ. 2020 Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* **367**, 1043–1047. (doi:10.1126/science.aaz3468)

179. Woodruff-Carr KW, White-Schwoch T, Tierney AT, Strait DL, Kraus N. 2014 Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl Acad. Sci. USA* **111**, 14 559–14 564. (doi:10.1073/pnas.1406219111)

180. Ozernov-Palchik O, Wolf M, Patel AD. 2018 Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *J. Exp. Child Psychol.* **167**, 354–368. (doi:10.1016/j.jecp.2017.11.009)

181. Ladányi E, Persici V, Fiveash A, Tillmann B, Gordon RL. 2020 Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisc. Rev.* **11**, e1528. (doi:10.1002/wcs.1528)

182. Mathias B, Lidji P, Honing H, Palmer C, Peretz I. 2016 Electrical brain responses to beat irregularities in two cases of beat deafness. *Front. Neurosci.* **10**, 40. (doi:10.3389/fnins.2016.00040)

183. Tranchant P, Lagrois MÉ, Bellemare A, Schultz BG, Peretz I. 2021 Co-occurrence of deficits in beat perception and synchronization supports implication of motor system in beat perception. *Music Sci.* **4**, 1–13. (doi:10.1177/2059204321991713)

184. Greenough WT, Black JE, Wallace CS. 1987 Experience and brain development. *Child Dev.* **58**, 539–559. (doi:10.2307/1130197)

185. Budisavljevic S et al. 2015 Age-related differences and heritability of the perisylvian language networks. *J. Neurosci.* **35**, 12 625–12 634. (doi:10.1523/JNEUROSCI.1255-14.2015)

186. Kokal I, Engel A, Kirschner S, Keysers C. 2011 Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment: if the rhythm comes easily. *PLoS ONE* **6**, e27272. (doi:10.1371/journal.pone.0027272)

187. Ridgway S, Carder D, Jeffries M, Todd M. 2012 Spontaneous human speech mimicry by a cetacean. *Curr. Biol.* **22**, R860–R861. (doi:10.1016/j.cub.2012.08.044)

188. Reiss D, McCowan B. 1993 Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. *J. Comp. Psychol.* **107**, 301. (doi:10.1037/0735-7036.107.3.301)

189. Roumazeilles L et al. 2020 Longitudinal connections and the organization of the temporal cortex in macaques, great apes, and humans. *PLoS Biol.* **18**, e3000810. (doi:10.1371/journal.pbio.3000810)

190. Balezeau F, Wilson B, Gallardo G, Dick F, Hopkins W, Anwander A, Friederici AD, Griffiths TD, Petkov CI. 2020 Primate auditory prototype in the evolution of the arcuate fasciculus. *Nat. Neurosci.* **23**, 611–614. (doi:10.1038/s41593-020-0623-9)

191. Gordon RL et al. 2021 Linking the genomic signatures of human beat synchronization and learned song in birds. *Phil. Trans. R. Soc. B* **376**, 20200329. (doi:10.rstb.2020.0329)

192. Pritchard JK, Di Rienzo A. 2010 Adaptation: not by sweeps alone. *Nat. Rev. Genet.* **11**, 665–667. (doi:10.1038/nrg2880)

193. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR. 2018 Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. *Genome Biol. Evol.* **10**, 166–188. (doi:10.1093/gbe/evx240)

194. Patel AD, Von Rueden C. In press. Where they sing solo: accounting for cross-cultural variation in collective music-making in theories of music evolution. *Behav. Brain Sci.*

195. Ashmore RC, Wild JM, Schmidt MF. 2005 Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J. Neurosci.* **25**, 8543–8554. (doi:10.1523/JNEUROSCI.1668-05.2005)

196. Lumsden CJ, Wilson EO. 1981 *Genes, minds and culture: the coevolutionary process*. Cambridge, MA: Harvard University Press.

197. Dennett DC. 2017 *From bacteria to Bach and back: the evolution of minds*. New York, NY: WW Norton & Company.

198. Laland KN. 2017 *Darwin's unfinished symphony: how culture made the human mind*. Princeton, NJ: Princeton University Press.