



Review

Cite this article: Merker B, Morley I, Zuidema

W. 2015 Five fundamental constraints on theories of the origins of music. *Phil. Trans. R. Soc. B* **370**: 20140095.

Trans. R. Soc. B **370**: 20140095.

<http://dx.doi.org/10.1098/rstb.2014.0095>

One contribution of 12 to a theme issue 'Biology, cognition and origins of musicality'.

Subject Areas:

evolution, palaeontology, theoretical biology, behaviour, cognition

Keywords:

culture, entrainment, evolution, generativity, music, vocal learning

Author for correspondence:

Bjorn Merker

e-mail: gyr694c@tninet.se

Five fundamental constraints on theories of the origins of music

Bjorn Merker¹, Iain Morley² and Willem Zuidema³

¹Fjälkestadv. 410-82, SE-29194, Kristianstad, Sweden

²Institute of Cognitive and Evolutionary Anthropology, and Institute of Human Sciences, University of Oxford, Oxford, UK

³Institute for Logic, Language and Computation, Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands

The diverse forms and functions of human music place obstacles in the way of an evolutionary reconstruction of its origins. In the absence of any obvious homologues of human music among our closest primate relatives, theorizing about its origins, in order to make progress, needs constraints from the nature of music, the capacities it engages, and the contexts in which it occurs. Here we propose and examine five fundamental constraints that bear on theories of how music and some of its features may have originated. First, cultural transmission, bringing the formal powers of cultural as contrasted with Darwinian evolution to bear on its contents. Second, generativity, i.e. the fact that music generates infinite pattern diversity by finite means. Third, vocal production learning, without which there can be no human singing. Fourth, entrainment with perfect synchrony, without which there is neither rhythmic ensemble music nor rhythmic dancing to music. And fifth, the universal propensity of humans to gather occasionally to sing and dance together in a group, which suggests a motivational basis endemic to our biology. We end by considering the evolutionary context within which these constraints had to be met in the genesis of human musicality.

1. Introduction

Music is a cherished art form and a daily source of inspiration and pleasure, as well as occasional irritation, for billions. It is also an extraordinarily complex phenomenon that appears to be not only uniquely human, but a human universal [1–3]. This uniqueness and universality raises the question of how and why the human ability to appreciate and produce music evolved. However, as is the case for language and other aspects of human cognition, it is not obvious how to properly constrain our theorizing so as to avoid producing no more than 'just-so stories'. Evolutionary biologist Richard Lewontin [4] has warned against 'the childish notion that everything that is interesting about nature can be understood. History, and evolution is a form of history, simply does not leave sufficient traces, especially when it is the forces that are at issue. Form and even behaviour may leave fossil remains, but forces like natural selection do not. It might be interesting to know how cognition (whatever that is) arose and spread and changed, but we cannot know. Tough luck.'

Against this blunt pessimism stand those who hold, with Richard Byrne, that 'comparative analysis of the behaviour of modern primates, in conjunction with an accurate phylogenetic tree of relatedness, has the power to chart the early history of human cognitive evolution' [5, p. 543]. With regard to human music, we suspect that neither side of this conceptual divide has rendered good advice to those who would explore its evolutionary origins. Perhaps the pessimism of Lewontin might be overcome by casting the comparative and inferential net wide enough. However, to do so we can no longer, as Byrne does, restrict ourselves to the study of primate homologues, but must explore analogies wherever they are found in the animal kingdom. Some traits do after all evolve *de novo* in a lineage. To understand such novelties, analogous developments in unrelated animals provide invaluable information regarding potential selection pressures and

ecological conditions favouring their evolution. The fruitfulness of such exercises, whether pursuing homologies or analogies, depends on the extent to which they can be constrained by stubborn facts regarding the phenomenon in search of an evolutionary explanation.

Here we focus on a small set of characteristics of human music that should help constrain accounts of its origins. They can be conceived of as basic hurdles that must be cleared along the way to a comprehensive theory of the origins of human music. They were chosen above all for their generality, with the additional desideratum of involving mechanisms that generate consequences for the structural content of music. At present a number of these constraints are difficult to meet, which means that besides their potential bearing on already proposed theories, they pose challenges for and may perhaps even inspire future ones.

2. Constraint no. 1: cultural transmission

Music, like language, is a complex product of cultural history. Its present-day patterns rest on traditions extending back over many thousands of years of inter-generational transmission of learned cultural lore [6,7]. This simple fact, so obvious that it typically is taken for granted in theories of music origins, nevertheless has profound consequences for any attempt to reconstruct the biological background to human music.

If patterns of cultural goods were only matters of human tastes and preferences—a common misconception regarding the nature of culture—the cultural transmission of musical lore would have no systematic or principled bearing on the reconstruction of music origins. However, when sustained over many generations, inter-generational transmission itself exerts profound and predictable effects on the contents of the transmitted lore, even in the complete absence of natural selection or any differential reinforcement of outcomes [8–12]. Thus, to go in search of evolutionary explanations for aspects of music that result from such a cultural process would be a serious mistake. As we shall see, major structural features of music are likely to be shaped by the cultural transmission process itself.

The key insight here is that with each generational transfer, the cultural lore (be it language, music, or any similar system transmitted culturally through learning) has to pass the so-called ‘learner bottleneck’. Any given learner is only exposed to a portion of the cultural lore extant in the population into which they are born and has, moreover, a limited capacity to absorb even the portion to which they are exposed. This means that the many items that make up that lore compete with each other for passage to the next generation. Through this competitive filtering process any and all aspects of the lore that bear on transmittability, including small differences in learnability and ease of processing, come to transform the cultural corpus in predictable ways, amounting to a cumulative process of informational ‘compression’ over many generations. This tends to issue in a tight fit between properties of the cultural lore and properties of the learner, introducing commonalities across the lore of different, separated, populations, all without the agency of biological selection.

In the field of language evolution, this mechanism which we refer to as ‘cultural evolution’ has been extensively studied over the past two decades (e.g. [13–16]). This work has led to a growing consensus (i) that cultural evolution is a powerful

mechanism, (ii) that many features of languages are potentially best understood as resulting from cultural adaptation to (pre-existing) hominin cognitive and physiological features, and (iii) that theorizing about the evolution of the biological basis of language can only sensibly proceed if we explicitly take into account the possibility that cultural evolution has shaped the linguistic phenotype. There is no reason to believe that any of this is any less applicable to the cultural transmission of music than it is to that of language.

Cultural evolution is a gradual, unconscious and obligatory process that extends over many generations, and restructures the cultural corpus in ways that increase its salience, expressive economy, communicative generality and grammatical power, all of which turn on enhanced communicability and learnability in various ways [15,16]. This allows for learners to manage ever larger amounts of cultural content without change in the neural resources devoted to it (through data compression) and lets the cultural products ‘exploit’ existing peculiarities of neural organization. For instance, Zuidema [17] discusses the finding of Smith & Lewicki [18] that the neural code in the auditory nerve of cats appears to be optimized for human speech sounds and argues that this finding only makes sense if the direction of causality is inverted: speech sounds have evolved in a process of cultural evolution to exploit features of a pre-existing general mammalian neural code, i.e. to achieve maximum discriminability under noise and time constraints.

Turning then to music, some major structural features of music widely distributed across cultures might likewise be a consequence of cultural evolution. Until recently, the failure of most musical tuning systems to conform to the mathematics of small integer ratios was grounds for rejecting Pythagoras’s proposal that small integer frequency ratios account for the perception of musical consonance and harmonicity [19]. However, recent modelling of the cumulative effects of physiological nonlinearities at each way-station of the ascending auditory pathway has disclosed the presence of ‘resonance neighbourhoods’ at whole integer ratio spacings on the tonotopic maps of the auditory system [20–22]. This finding not only accommodates a wide range of tuning systems and musical scales found worldwide, but appears capable of accounting for human judgements of consonance, dissonance and tonal stability/attraction in terms of inherent organizational features of our auditory system ([20]; see also [23–27]), as follows.

The pattern of ‘resonance neighbourhoods’ in auditory system tonotopy is likely to be shared by all mammals, being a product of quite elementary properties of the neural circuitry in question. It did not evolve for purposes of music, in other words, but as an incidental by-product of the interaction of excitation and inhibition in a neural system evolved to process natural sounds efficiently [20]. Not being confined to humans [28,29], it is not likely to represent an adaptation to music. The presence in humans of cultural patterns of musical practice conforming to these subtle resonances in auditory physiology accordingly requires an explanation.

The fact that those musical practices are, indeed, cultural patterns formed by inter-generational transmission may supply the explanation. In principle, the formal powers of cultural evolution should suffice to allow musical practice to eventually find its way to the ubiquitous and inherent resonant biases of the auditory system, given a long running cultural tradition of song [30]. Through the external loop of inter-generational transmission of learned musical lore, the

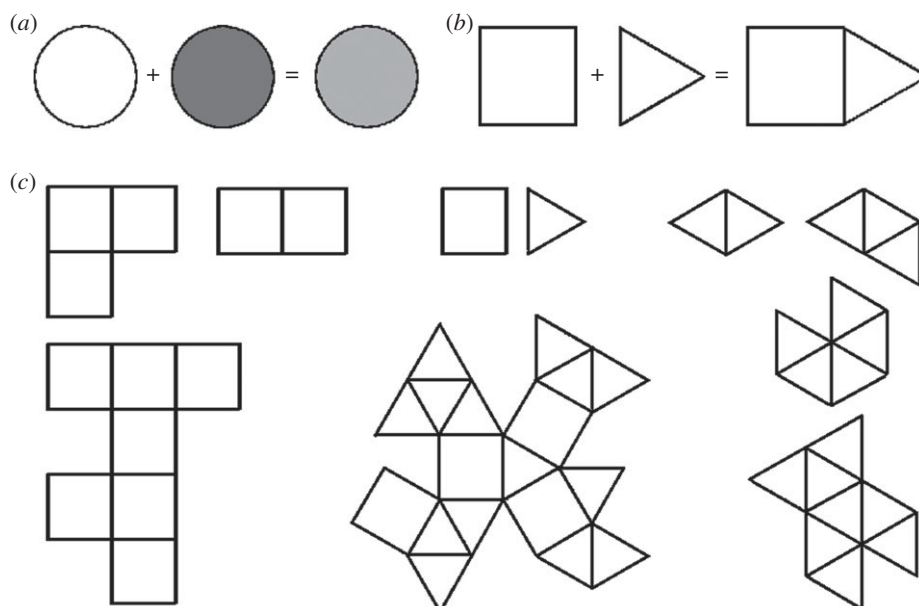


Figure 1. Cartoon illustration of the ‘particulate principle of self-diversifying systems’, following Abler [44]. (a) A ‘blending system’ in which combining ingredients average. Here exemplified by a drop of ink in water: the combining elements do not generate a qualitatively new entity. Other examples are most mixtures of liquids as well as gases, as in weather systems, and patterns of heat conduction. (b) A ‘particulate system’, in which the combining elements generate a qualitatively new entity by retaining their individuality on combining. (c) A minuscule sample of the infinite generativity of a combinatorics of as few as one or two discrete non-averaging elements.

production of musical patterns would pass through the ‘learner bottleneck’ to be shaped by pre-existing biases on the part of learners, specifically the purely perceptual resonant biases just invoked. This assumes not only a capacity for vocal learning (see ‘Constraint no. 3’), but one emancipated from innate song templates, for which there is precedence in the true mimics among the birds [31–33]. From such a starting point, devoid of scales, tonality and small integer ratio consonances, thousands of generations of cultural transmission could eventually externalize even subtle biases in auditory perception in the musical practices of human cultures (see also [28]).¹

The principal constraint this process imposes on theories of music origins is that they must provide a non-arbitrary reason for our distant forebears to have engaged in persistent inter-generational transmission of vocal lore *lacking* the tonal organization of music as we know it for long enough to allow the transmission mechanism to find its way to the resonances already embedded in basic auditory physiology. Since thousands of generations may be required for this to happen [14,37], the constraint is a real one. Perhaps our forebears, like many bird species, maintained cultural traditions of learned song, and became vocal learners as part of the cluster of changes that define the emergence of *Homo* some 2 Myr [38,39]. An increasingly refined vocal communication system for the accurate communication and extraction of emotional information from vocal prosody is likely to also have contributed to this process ([7,40,41] and [42, note 3]).

More generally, placing the historical nature of the pattern-content of human music at the head of the effort to understand its biology greatly facilitates the reconstruction of its evolutionary background. Because its patterns are learned from a corpus of cultural models subject to the transformative dynamics of the inter-generational ‘learner bottleneck’, there is no need to ask evolutionary selectional mechanisms to equip us with those pattern specifics, even when they happen to be cross-culturally widespread, as in the ‘auditory system resonance’ example.

3. Constraint no. 2: generativity or infinite variety by finite means

Music, like language, is generative, i.e. it produces infinite pattern variety by finite means [1]. The key to that variety in both music and language is of course not recursion [43] but combinatorics [1,44,45]. By combining a finite set of elements—discrete pitches and durations—music creates composite patterns without limit. For this to be possible, the combining elements must be non-blending in the sense of not producing an average when combined [46], i.e. they must retain their individuality on combining (figure 1a,b). When that is the case, each such combination ‘creates something which is not present *per se* in any of the associated constituents’ [1, p. 67], making infinite pattern variety possible (figure 1c). There is a total of four major such open-ended generative systems in existence, two of which are natural ones (chemistry and genetics), while two are cultural (music and language; table 1).

In the cases of music and language, the combining elements are conventional, the musical ones arising through a radical reduction in the degrees of freedom available to vocal or instrumental sound production [45]. This is accomplished by discretizing two continua, those of pitch and duration, to yield musical notes with determinate pitch and—in all rhythmic music—proportional durations based on discretizing time through an isochronous pulse (see ‘Constraint no. 4’).

In other words, musical notes are not simply pitches. Rather, they are individuated pitch locations within a discretized pitch continuum. They are fixed reference points on that continuum, between which even glissandi must travel with the same necessity as does any ordinary note if they are to be musical. The designation of a specific location on the pitch continuum as a ‘note’ by a culturally determined ‘pitch standard’—applied with a conventionalized margin of tolerance—lifts that location out of its relation of equivalence to its infinitude of pitch neighbours. It breaks its ‘anonymity’, as it were, and turns it into an individuated and specific musical note to which a musical

Table 1. The principal open-ended or ‘self-diversifying’ systems.

system	chemistry	genetics	music	language
product	all molecules	all life forms	all ‘melodies’	all sentences
combining ‘partides’	atoms	nucleotides	notes	phonemes
particle type		physical		conventional
domain		nature		culture

figure can return and which can be used repeatedly in the development of a musical pattern.

This discretization of the pitch continuum into determinate ‘pitch sets’ supplies music with combinable pitch elements featured in musical melodies and chords [19,47,48]. Pitch sets thus supply the ‘particulates’, the individuated elements, needed for its combinatorial mill. They are found in all musical traditions cross-culturally. Indeed, one of the distinguishing marks of musicianship anywhere is adherence to the pitch locations designated by a pitch standard during musical performance. Not to do so is to sing or play ‘out of tune’, the quintessential demarcation line between musical and other employments of human capacities.

The constraint imposed on theories of music origins by the generativity of music is that no such theory can account for the genesis of music as we know it without giving a credible account of how we came to conquer for ourselves the discretized (‘particulate’) elements without which there can be no open-ended generativity of music. In light of what has been covered under Constraint no. 1, these elements may of course be prime products of a protracted process of cultural transmission exploiting the tonal scaffolding of auditory system resonances along with factors such as the convenience of dividing the octave into steps that maximize the individuation of its intervals (for which see [45,47]).

The ways in which cultural evolution of musical lore might produce particulate elements need further research, including the exploration of computational models. Perhaps accounts of how discrete combinatoriality (culturally) evolved in phonology [49,50] can be adapted to music. In these models the evolution of a repertoire of continuous trajectories through an acoustic space is studied. Discrete structure emerges in these models as a side effect of the neural encoding [49] or of optimization for discriminability [50]. Both of these proposals would seem applicable to music, providing a potential route to superficially combinatorial structure in the musical lore.

If we can make plausible that such a cultural route can lead to productive combinatoriality (generativity) too, there is no need to burden theories of music origins with Darwinian accounts of the origin of musical notes, scales, and tuning systems. However one conceives of the matter, the point here is only this: a credible theory of music origins must furnish such an account, short of which the phenomenal pattern richness of human musical culture remains a cipher.

4. Constraint no. 3: vocal learning

Every song we know how to sing, and every word we know how to pronounce is ours through a highly specialized learning capacity that is conspicuous by its absence in other primates, our closest living relatives included. The vocal patterns of

song and speech are acquired through *motor* learning on the basis of heard, culturally transmitted models through a process requiring intact hearing and feedback from one’s own voice [51–56]. The process by which they are acquired is technically known as vocal production learning [57,58], a dedicated and highly specialized capacity that has no other common uses in our lives besides song and speech.

Since there can be no human singing without it, the origin of our capacity for vocal production learning bears directly on scenarios for the origins of music. The issue is an acute one, since the fact that other primates lack this capacity [57,58] means that we became vocal learners at or after our divergence from the common ancestor we share with chimpanzees. One limb of the comparative method—the tracing of continuities (homologies) with our close evolutionary relatives—is therefore unavailable for reconstructing its origin in this particular case, Byrne’s assertion to the contrary quoted in our introduction notwithstanding.

There are a variety of context- or learning-based modifications of vocal output that do not involve the mechanism of vocal production learning in the technical sense applicable to human song and speech. They include contextual modulation of vocal behaviour, socially or environmentally contingent selection among innate calls and their variants, and their learned modification, as detailed by Janik & Slater [58]. There is no dearth of evidence for such vocal phenomena among primates. They are an integral part of the vocal expressiveness primates share with many other mammals, but occur without reliance on the specialized mechanism of vocal learning.

Vocal learning proper, by contrast, is the ability to convert heard sound patterns that are not in the species-specific innate vocal repertoire into vocal output, using feedback from one’s own voice to achieve the match [55]. It has been studied in detail above all in birds [51,59], among whom there is a rich assortment of vocal learning phenomena, by no means all the same. They differ along at least six major dimensions of classification, as reviewed by Beecher & Brenowitz [60]. Human vocal learning occupies the more advanced end of several of these dimensions in that it is open ended, allowing new patterns to be added throughout life (though with diminished accuracy after puberty), as well as being emancipated from dependence on a species-specific vocal template. Given that the human capacity is an advanced one, comprehensive studies of the true mimics among birds (mynahs, many species of parrots, lyrebirds, butcherbirds, mockingbirds, etc. [33]) are needed to supplement the invaluable knowledge about mechanisms of vocal learning supplied by the bird species typically employed in the study of vocal learning in the laboratory.

As far as is currently known, vocal production learning proper is found only in humans, cetaceans, pinnipeds,

elephants, bats, oscine songbirds, parrots and hummingbirds [57,61,62]. Given its absence in non-human primates, the process by which our ancestors were equipped with this capacity is a major evolutionary event intervening between the last common ancestor and the first singing or speaking humans. It supplies a major biological constraint or 'evolutionary bottleneck' [63] on the path to human music. Its origin in our lineage could have been driven by either song, speech or other factors. This virtually forces the theorist to come to grips with the order of precedence of song and speech in our ancestry ([64, figure 21.1 and accompanying text]).

There is currently no good account of how humans evolved the capacity for vocal production learning. As Nottebohm noted many years ago: 'you might find it much harder to explain this first step, vocal learning, than the latter acquisition of language' [63, p. 645]. Nottebohm's warning, we submit, applies to music no less than to language. One possibility in this regard is that we acquired vocal learning, like some of the songbirds, as a means to sustain cultural traditions of learned song (see 'Constraint no. 1'). Another might be that vocal learning built upon comparable abilities for manual imitative learning and variation that were already developed, or developing. However conceived, our possession of vocal production learning is a fact, and one that any theory of the origins of music leaves unaccounted for at its peril.

5. Constraint no. 4: entrainment

The constraints considered so far apply both to human music (song) and to language (speech), and therefore cannot help us home in on evolutionary factors unique to music as such. This is no longer so for the final two constraints we shall consider, beginning with our capacity to entrain our behaviour to one another with perfect synchrony.

The type of temporal coordination of inter-individual behaviour that is most distinctively musical, hardly occurring outside the domains of music, dance and drill in the human case, and certainly not in speech, is the type of entrainment that features what Ermentrout has dubbed 'perfect synchrony' [65]. It consists of mutual phase-locking with zero (or even slightly negative) phase lag between the periodic signals of two or more signallers sustained consistently at a given tempo, with a capacity to do so at different tempos.

As members of a species in possession of such an entrainment capacity, we tend to take it for granted. Doing so may obscure from us not only its key characteristics but also the reason for the exceedingly sparse distribution of this capacity among animal species (see below). We will therefore endeavour to make its phenomology explicit so as to avoid confusing it with unrelated forms of temporal coordination commonly occurring in the animal kingdom.

The human capacity for perfect synchrony has been well established and explored through more than a century of sensorimotor synchronization studies [66,67]. The zero (or even slightly negative) phase lag of such entrainment means that its timing mechanism is predictive. A punctate behaviour that coincides with (or even slightly leads) a given beat in an isochronous sequence cannot be caused by a reaction to that beat because of reaction time limitations. Predictive timing is made possible by the regular periodicity of the entraining signal, its isochrony, also known as *tactus* or 'pulse' in music [68,69]. It makes an upcoming beat in the sequence perfectly

predictable, and therefore targetable by the predictive timing mechanism [69,70]. Positive asynchronies large enough to come within reach of auditory reaction time, such as those reported for macaques in a synchronization task [71,72], are thus automatically excluded as evidence for entrainment of a kind relevant to the human capacity.

In inter-individual entrainment, the isochrony needed for synchrony must be motorically produced by the entraining individuals on an endogenous (generative) basis. To sustain phase-locking at an average of zero phase lag between such periodic outputs under conditions of a variety of biologically inevitable local perturbations and drift requires mechanisms of phase correction as well as period adjustment [65]. Both are well documented for human sensorimotor synchrony [67]. A mechanism equipped with these features latches on to the regular beat and stays on it as long as that beat stays reasonably steady and lies within the operational tempo range of the entrainment mechanism. In humans that range centres on 2 Hz, which is also the human locomotor tempo [70,73]. Entrainment precision is dependent on predictability, so variance in period length has to be small, typically exhibiting a standard deviation of a few (2–5) percentage points in human tapping performance [66,67].

Entrainment between two or more motoric time series by such a mechanism establishes an unequivocal, unique and rather precise correspondence between the individual events making up the simultaneously unfolding sequences. That correspondence is either one to one or related by small whole integer ratios for harmonically related tempos [74]. It is from this unique correspondence between the individual events of separate time series that asynchronies and their variability are calculated as a measure of synchronization skill [67]. That is, the achievement of beat matching is *presupposed* by these measures, which assess only how precisely in time that matching occurs. One cannot therefore—as was done in a study of purported entrainment in macaques [75]—take the time series of the animal with the slower movement pace as a reference, and for each of its events select the closest match in time from the other animal's record as a basis for calculating asynchrony. There is always such a 'closest event' irrespective of entrainment, and when as in this case (see tempo means and variances in fig. 2 of that study) sequences drift in phase relative to one another some of these events will occur before and some after the reference event, and these will average out to small asynchronies, again irrespective of entrainment.

The study of cockatoo dancing to human music by Patel *et al.* ([76], see also [77]) helps define the entrainment phenomenon further by way of contrast with human performance. The bird's episodic stretches of on-the-beat synchrony emerged from intervening phase drifts over the full phase range in an erratic pattern, while the musical beat remained steady all along. This is not the behaviour expected of a mechanism dedicated to entrainment through provisions for phase and period correction. As already mentioned, such a mechanism locks on to a steady entraining signal and stays locked with only minor asynchronies as long as that signal remains reasonably steady. We do not at this point know why the bird does not do so, but one possibility is that producing dance-like movements mimicking those of its human keepers takes precedence over sustained and precise behavioural matching of the musical beat in the bird's performance. There are no indications that these birds engage in pulse-based synchrony in their natural

habitat, but many monogamous parrots do engage in joint cooperative pair displays, some of which are learned by imitation [78]. Perhaps, then, the imitative capacities of these highly intelligent birds may help account for their imperfectly entrained dancing to rhythmic music in a human environment.

There are, however, animals who, like us, produce steady isochronous signal sequences on an individual basis and mutually entrain to such signals on a group basis with sustained and consistent on-the-beat matching across individuals as part of their natural behaviour in the wild. These species of 'natural synchronizers' are few in number and far removed from us in evolutionary terms, being found among species of fireflies, crickets, cicadas and fiddler crabs [79,80]. To these may be added some marine bioluminescent crustaceans [81], and the rattan ant [82], the latter virtually unstudied (see below).

The champions among these non-human synchronizers are three species of synchronously flashing fireflies and possibly a few species of synchronously chorusing crickets [65]. They entrain their behaviour to one another with a sustained pulse-based rhythmic precision featuring both phase and period correction that equals or exceeds that of human mutual synchrony in music, dance and drill [65,83,84].

That is not to say that the mechanism by which these insects achieve their impressive synchrony is the same as the human mechanism. They only share with us those features of it needed to achieve behavioural entrainment with perfect synchrony. We can easily entrain to a synchronous cicada chorus, but cicadas are unlikely to entrain to our favourite dance music, given the limited auditory scene analysis performed by their fraction-of-a-milligram brains. It is to say, however, that in these animals we have the only documented instances besides our own of genuine beat-based group synchrony that plays a role in the natural behaviour of the species in question. These species therefore may provide us with invaluable hints regarding our own path to this rare behaviour by comparative scrutiny of its functions and evolution in these insects.

The reason for the sparse distribution of the capacity for mutual beat-based entrainment in nature is not far to seek. It resides in the apparent lack of its general biological utility, being virtually useless, with a very few narrowly constrained exceptions. As far as is presently known, the functions it serves where it features in the behavioural repertoire of non-human species in the wild are confined to special cases of mate attraction and predation defence (reviewed in [79,80]). Among the former the so-called 'beacon effect' [85] takes pride of place, featuring thousands of synchronously signalling male fireflies whose entrainment precision falls at the more skilled end of the human range, with a standard deviation in period length of less than 3% [83]. Through their entrained luminescent signalling, single trees of permanent male congregations are converted to flashing 'beacons' visible from all directions despite the foliage that obstructs single lines of sight in the tropical rainforest. Only male fireflies synchronize their flashing, and females are attracted to these displays. The synchronizing males, of course, are in competition for the females who arrive, and at close quarters females prefer more luminous males, who are also bigger in size.

Defensive uses of synchrony are of two principal kinds: evasive and deterrent. Synchronous calling among neighbouring callers may confuse a predator's auditory ability to localize any given caller in the chorus, as appears to be the case for a species of treefrog preyed upon by bats [86]. These amphibians do not, however, call rhythmically, but achieve collective

superposition of calls by calling at very short latency following the first individual to call spontaneously. This renders their behaviour a special case of reaction-time-limited calling and is therefore irrelevant to our topic.

A deterrent use of synchrony is that of the rattan ant, which lives in symbiosis with the rattan vine [82]. When a vine is disturbed by a sudden external impact, it emits an unexpected and potentially unsettling audible rattle at sound levels far beyond what any single ant can produce. The rattle is a result of entrainment of the alarm behaviour of the ants, which consists of rhythmic beating of their gasters against the vine surface. Local clusters of ants do so in synchrony, with lack of entrainment to more distant ants. Many such locally synchronized clusters produce the unexpected rattle. Amplitude summation (auditory 'beacon effect') is the key to this defensive use of inter-individual synchrony.

The narrow compass of behavioural and ecological conditions under which the otherwise useless capacity for entrainment with perfect synchrony has evolved among animals imposes an exceedingly stringent constraint on theorizing about its origin in our case. This is all the more noteworthy in that in the human case the constraint pertains specifically to music and little else in our behavioural repertoire except the music-related disciplines of dance and drill. As such it provides an invaluable asset in evolutionary scenario building. This fortuitous circumstance has been exploited in an account of the origin of the human entrainment capacity proposed by Merker [42,69,87], briefly summarized in the next section.

6. Constraint no. 5: motivational basis

Wherever humans live, and however they have organized their societies, they exhibit a behavioural peculiarity of gathering from time to time to sing and dance together in a group [1–3]. By featuring both human song (Constraint no. 3) and entrainment (in the dancing movements and perhaps clapping performed in synchrony with the singing/music, Constraint no. 4), such behaviour qualifies as human music. Indeed, the fact that it occurs in every human culture, and indeed subculture, without exception, unless deliberately suppressed by severe sanctions against it, marks this phenomenon as the most universal human behaviour of a musical kind on record.

In its ubiquity, this human propensity for occasional group singing and dancing would seem to constitute a prototypical musical behaviour, all the more so as it can be staged entirely without musical instruments (as in the traditional trance dance of the hunter-gatherers of the Kalahari Desert [88], see also [89]). It may in fact represent the motivational core of the human capacity for music from which its many other manifestations may have developed by differentiation, elaboration and specialization. One is assisted in becoming aware of the peculiarity and specificity of this behavioural propensity by imagining that in exactly those circumstances in which we typically gather to sing and dance together in a group, another human culture would gather in groups to draw pictures together instead.

The ubiquity and specificity of this putatively prototypical musical behaviour would seem to require an explanation. In searching for one we enter for the first time onto the grounds of a possible homology, because certain social displays of our closest living primate relatives may provide a biological background to the human tendency.

As pointed out by Geissmann [90] there is an association between ‘loud calls’ (‘distance calls’) and physical display among our closest living primate relatives, the apes. The loud calls used by apes in distance signalling (‘long call’, ‘pant-hoot’, gibbon pair duet, etc.) tend to be accompanied by vigorous physical displays such as locomotor excitement, branch shaking, chest beating and other forms of noise-making called ‘drumming’ by Fitch [91], although lacking the pulse-based rhythmicity of drumming in the musical sense. These displays do not feature any metrical structure resembling isochrony, nor any pulse-based rhythmic entrainment between individuals, but they do provide a precedent for the linkage between vocalization and bodily movement that occurs in human group singing and dancing.

Chimpanzees exhibit a social elaboration of this coupling between voice and physical display into an occasional group frenzy called the ‘carnival display’. On irregular occasions, typically when a foraging subgroup discovers a ripe fruit tree or when two subgroups of the same territory meet after a period of separation, the animals launch an excited bout of loud calling, stomping, bursts of running, slapping of tree buttresses and other means of chaotic noise-making. There are no indications that any kind of inter-individual coordination, let alone rhythmic synchrony, forms part of these chimpanzee group displays. They may last for hours, even a whole night, and induce distant subgroups and individuals on the territory, both male and female, to approach and join the fray [92–97].

Our social–emotional propensity to occasionally gather for excited group displays appears to be shared, in other words, with our closest living relative among the apes, the chimpanzee. We are not alone in sensing a possible connection in this regard. BaYaka pygmy hunter–gatherers inhabit the Congo-Brazzaville rainforest, which they share with chimpanzees. Mokondi Massana ‘spirit plays’ featuring ritual singing and dancing are a significant aspect of BaYaka culture.

When BaYaka . . . hear a chimpanzee ‘carnival display’ from their camp it provokes great hilarity among camp members as one or two of them begin imitating the frenetic actions of the chimpanzees as they pound buttress roots or shriek at the canopy. The camp is launched into laughter as they explicitly ridicule the chimpanzees attempt to stage a ritual (massana), but are incapable of bringing it off properly. Fables such as ‘Chimpanzee you will die’ (sumbu a we) elaborate on this theme describing how chimpanzee tries to get initiated but has to be dissuaded to avoid him being killed during the trials.

(Jerome Lewis, personal communication to BM, 2014, with permission.)

Our propensity for occasional gatherings of excited group displays may in fact be a primitive trait conserved in both lineages from our common ancestor, far predating its elaboration with specifically musical content in our case. If the propensity for an excited social noise-and-movement display is indeed homologous in the two cases, one with musical content and the other without it, this bears directly on theories attributing group or social functions to music. In case of homology the causal arrow may be reversed, the social efficacy deriving not from the musical content of the group activity but from the motivational mechanisms of the group display itself, long antedating its musical elaboration. This ‘group excitement’ factor has to be controlled for in studies designed to explore the emotional or social significance and consequences of human music.

Assuming homology, for the sake of argument, in our case the communal display was eventually elaborated by

the introduction of metric and melodic structure into the chaotic noise-and-movement display. The refinement takes the form of regularizing the pacing of both voice and bodily display, making the even pace of its tempo (isochrony) the means for entraining the behaviour of individuals to one another in an accurately timed group display of rhythmic chanting and dancing. A plausible setting for such a development is the male group territoriality combined with female exogamy—a rare pattern among higher animals—that can be assumed to have characterized the last common ancestor of humans and chimpanzees [69,98,99]. Merker [42,69,87] noted the striking parallelism between this pattern and the male clumping combined with female migration that is the functional and evolutionary key to synchronous chorusing in the insect examples cited in the previous section [79], and proposed it as a selection pressure for the evolution of the human entrainment capacity [42,69,87]. As noted by Merker *et al.* [69], such a scenario is eminently compatible with the central tenet of the coalition signalling scenario proposed by Hagen & Bryant [100].

The constraint we are proposing in this section pertains to the motivational underpinnings of music, rather than to its structural content. Something needs to explain the cross-culturally universal human tendency to gather from time to time for group singing and dancing. No theory of music origins can be considered complete without somehow accounting for this tendency. If, as suggested here, the social function and emotional impact of the gatherings which in our case feature music far antedate their specifically musical content, then it is not to the musical content but to the decidedly non-musical social adaptations of our hominoid ancestors that we should look for the secret of the social function and emotional impact of those gatherings.

7. The evolutionary context

In the course of detailing the foregoing constraints, we have noted that some distinguishing aspects of music (e.g. scale systems) require no Darwinian explanation for their widespread yet unique occurrence in humans; equifinality can occur as a consequence of characteristics of learning mechanisms and existing constraints providing the necessary frameworks for such development. Other underlying abilities required for musical behaviours, however (e.g. vocal learning, entrainment), are likely to be the product of forms of Darwinian selection. In these cases, the question then arises regarding what modes of selection might be operative, and on what, specifically, they might be operating.

There are some general lessons from evolutionary theory that are relevant, but often ignored, in constructing evolutionary scenarios for music (and language).

The first point is that biological evolution is always about genetic change. Even though very few genes involved in music have been identified [101], it is important to recognize that evolutionary scenarios, implicitly or explicitly, assume a sequence of changes in gene frequencies in a population, including the appearance of new genetic variants. Making this assumption explicit helps in avoiding the common fallacies of assuming (implicitly) unrealistic amounts of genetic changes (although that is difficult to quantify), assuming instantaneous adoption of new variants, or ignoring the fact that new variants, arising from mutation, are initially always rare.

A second point is that although evolution involves non-adaptive mechanisms such as random mutation and drift, a series of non-adaptive genetic changes leading to a complex new phenotype is exceedingly improbable. To establish the plausibility of a scenario for a trait shared by all humans, we thus have to show that each new variant conveys a fitness advantage both when it is rare in a population and when it has already become quite common. Moreover, we must show that this advantage applies to the individual that carries it, rather than to the group as a whole (simply assuming selection for the benefit of the group is widely considered a fallacy). Traits that benefit the group rather than the individual can only evolve under quite specific circumstances described by kin selection and social evolution theory *sensu* Frank [102].

A third point is that we need to be aware of the fact that the fitness advantage of a trait might not, or not only, come from its contribution to increased success in reproduction through increased survival (natural selection in the narrow sense, though including benefits associated with individuals' ability to establish effective social alliances), but may also come from the trait's effects on increased success in reproduction via attractiveness to potential partners (Darwinian sexual selection [103]). This could be particularly relevant for the evolution of music, as sexual selection is invariably invoked in understanding the evolution of elaborate animal aesthetic displays (where the connection between display and fitness can be very indirect). Music is nothing if not an aesthetic display (although possibly much else besides). Darwin treated it as such, and proposed sexual selection as the mechanism behind it.

Finally, a fourth point is that in order to confer a selective advantage, a trait or behaviour need not be *essential* for survival, but need only confer a slightly improved likelihood of survival to procreation, and/or a greater rate of procreation—thus perpetuating and increasing the frequency of that trait or behaviour—than would otherwise be the case. There is thus no justification for the common observation that music could hardly be a product of evolution by selection as it is hardly essential for survival. The former observation does not in fact follow from the latter. Furthermore, it does not rule out the possibility that various of the abilities that are used in musical activities may have been initially selectively favoured as a consequence of their fulfilment of other purposes (e.g. interpersonal communication and the establishment of interpersonal relationships), and that musical practices may have developed within the context of those uses; musical behaviours have the potential to fulfil some of those same purposes, or other purposes, potentially in even more effective ways. The co-use of these underlying abilities could lead to increasing interdependence between them, uniting them functionally in this new behavioural system, and potentially leading to further selective processes acting upon those underlying abilities and the behaviours that use them.

Some of the traits that are essential for musical activities may have been a product of biological (natural or sexual) selection, and this could be by conferring a fitness advantage either in the context of their use in musical activities themselves, or in a different context of use. Meanwhile, as observed in the preceding sections, certain properties of music and the traits that support them need not have been the product of biological selection at all.

The constraints outlined in the preceding sections indicate that some of the abilities prerequisite for music (e.g. entrainment

and vocal learning) would appear to have arisen in our lineage, or at the least adapted from existing mechanisms to take on essentially novel form, in the period between our last common ancestor with chimpanzees (approx. 6–7 Myr) and the appearance of our own species (approx. 200 000 years ago). Proposals regarding the emergence of these abilities should be complementary to, and tested against, what we know of the physiology, behaviour and ecology of the hominins in that intervening period. This is no small task as clearly the knowledge both in palaeoanthropology and in the study of primate and human cognition is in a state of constant flux. Nevertheless, some aspects of our understanding in both areas are well-enough established that we should undertake to ensure that proposals regarding the evolution of these capacities do not contradict core understanding in hominin evolution.

For example, hypotheses regarding the development of collective bodily and vocal display behaviours in early hominins from those exhibited by chimpanzees (and presumably our last common ancestor with them) should be framed in the context of changes in the habitat and group size of successive hominin species. It is now well established that gracile australopithecines exploited more open environments and a more omnivorous diet than higher primates of today, but nevertheless continued to exploit wooded environments for shelter and some aspects of subsistence [104,105]. Meanwhile the physiological characteristics and ecological contexts of early *Homo* (*H. habilis*, early African *H. erectus* and their descendants such as *H. heidelbergensis*) indicate that they were exploiting a far greater range of more open environments, lacustrine and riverine habitats, and that carnivory had increasingly taken the place of arboreal frugivory in their subsistence resource exploitation [104,105]. The efficacy of any proposed alterations to ancestral 'carnival displays' [69], coalitional displays [100] or size-exaggeration vocalizations [106], for example, needs to be situated within the context of these changes (see also [7]).

The mating strategies of human ancestors, and the social organization that arises from them, are also relevant to assessing the ecological validity of models regarding foundations of musical behaviours in interpersonal communication and display behaviours. This is because strategies for interpersonal communication, alliance and pair-bond formation, and display behaviours, will vary according to whether, for example, populations are monogamous, polygynous or polyandrous. In polygynous species, for example, males compete with other males for access to multiple females, with little or no long-term alliance commitment to any one female. By contrast, monogamous species form pair-wise long-term cooperative bonds between males and females. In each case the types of cooperation and competition, and with whom cooperation and competition occur, vary, and the behaviours leading to success in negotiating alliances and long-term bonds, and in display directed at the same sex, and directed at the opposite sex, vary accordingly (e.g. [107]).

In the case of human ancestors, high levels of sexual dimorphism and rapid developmental life history in australopithecines comparable to that of chimpanzees [108] have been taken to indicate broadly similar mating strategies and male–female relations to those exhibited by chimpanzees (e.g. [105], and references therein). Meanwhile, trends towards a reduction in sexual dimorphism and increased altriciality in the infants of early *Homo* indicate the development of increased cooperative long-term pair-bonding (e.g. [105] and references

therein). As noted in the previous paragraph, these developments have a direct influence upon the forms and efficacy of behaviours related to display, sexual selection, pair-bonding and vocalization between adults (see also [39]). The development of greater altriciality in infants, a longer developmental process and greater dependence upon adult care also have direct impacts upon the vocal behaviours between adults and infants (e.g. [109]), and the learning opportunities of infants and juveniles (e.g. [108,110,111]).

Similarly the potential value and form of vocal learning capabilities should be tested and understood against the backdrop of physiological constraints for vocalization capabilities. For example, MacLarnon & Hewitt [112] studied the size of the nerve canal in the thoracic vertebrae of australopithecines, early African *H. erectus* (*H. ergaster*) and its later descendant *H. heidelbergensis*. These dimensions provide an indication of the level of fine control of breathing musculature present in the species, allowing controlled vocalizations of extended duration, with greater control over intensity and pitch contour. They concluded that the level of fine control of breathing musculature in early African *H. erectus* (ca 1.8 Myr) was not increased relative to that of chimpanzees or the earlier australopithecines, but that by the time of its immediate descendent, *H. heidelbergensis* (from approx. 5 to 600 000 years ago), the level of control would have been equivalent to that of modern *H. sapiens*. These changes could have served either song or speech, and it is worth noting that on a number of measures such as tidal volume, range of subglottal pressure and muscular control, the biomechanics of human song are more demanding than those of conversational speech [113].

While we are able to conclude that neural connections allowing deliberate planning, fine control and integration of laryngeal, orofacial and respiratory musculature used in vocalizations (as would be required for vocal learning) emerged in the lineage since our last common ancestor with chimpanzees, the hominin fossil record does not preserve neural changes internal to the brain. However, if the development of the thoracic neurology for voluntary extended breathing control was to be useful for vocalization, it would have had to have been accompanied by, or preceded by, the development of the neurological connections in the brain allowing the planning and control of these aspects of vocalization. Conversely, the usefulness of the neurological connections in the brain

allowing integrated voluntary control over the larynx, articulation and breathing would have been increased by the development of greater thoracic breathing control. It would seem likely that these two neurological systems, in the brain and the body, would have evolved in tandem, during the 1 million-or-so-year period between early African *H. erectus* and *H. heidelbergensis* (see also [7]).

The many ways in which evolutionary changes in traits and behaviours relevant to musical behaviours can occur, by biological selection or otherwise, are not mutually exclusive; by contrast, they can interact in important and complex ways, and any or all of them could have operated at various times in the course of human evolution. However, the distinctions between them have not always been clearly made in the literature discussing evolutionary rationales for musical behaviours. It is important that any future proposals do so, and clearly situate such mechanisms within what we know of the social and ecological contexts of human ancestors, and their physiological and neurological capabilities.

Acknowledgements. We thank Henkjan Honing for organizing, and the Lorentz Center, The Netherlands, for hosting the workshop that resulted in this paper. Our thanks also go to our editor Carel ten Cate, Guy Madison and an additional anonymous reviewer, for numerous comments and suggestions that have helped us improve its contents.

Endnote

¹This sketch leaves out the many critical developments a vocal learning tradition must traverse in order to do what we propose it to have done in our case. The vocal learning capacity must first be emancipated from dependence on an innate song template, as in the bird mimics cited in the text. It must also abandon exclusive reliance on the tiny 'vocal gestures' that supply the song elements for most bird-song, learned and unlearned, to include elements more akin to musical notes, i.e. notes sustained at a given pitch with spectral energy concentrated to the fundamental. There is precedent for this in birds such as the pied butcherbird of Australia, a mimic and virtuoso singer [34–36]. Only on the basis of producing such music-like notes is vocal production learning likely to engage auditory system resonances with enough strength and reliability to become a factor in cultural transmission. The requirement that all this be in place if the process we have postulated is to get a start may help explain the rarity of actual tonal phenomena in animal song.

References

1. von Humboldt W. 1836 *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin, Germany: Royal Academy of Sciences. GC Buck, F Raven trans. 1971 *Linguistic variability and intellectual development*. Baltimore, MD: University of Miami Press.
2. Brown DE. 1991 *Human universals*. New York, NY: McGraw-Hill.
3. Brown S, Jordania J. 2013 Universals in the world's musics. *Psychol. Music* **41**, 229–248. (doi:10.1177/0305735611425896)
4. Lewontin RC. 1998 The evolution of cognition: questions we will never answer. In *An invitation to cognitive science, volume 4: methods, models, and conceptual issues* (eds D Scarborough, S Sternberg), p. 130. Cambridge, MA: MIT Press.
5. Byrne RW. 2000 Evolution of primate cognition. *Cogn. Sci.* **24**, 543–570. (doi:10.1016/S0364-0213(00)00028-8)
6. Higham T, Basell L, Jacobi R, Wood R, Bronk Ramsey C, Conard NJ. 2012 Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geißenklösterle. *J. Hum. Evol.* **62**, 664–676. (doi:10.1016/j.jhevol.2012.03.003)
7. Morley I. 2013 *The prehistory of music: human evolution, archaeology, and the origins of musicality*. Oxford, UK: Oxford University Press.
8. Bartlett FC. 1932 *Remembering*. Oxford, UK: Macmillan.
9. Kirby S. 1998 Language evolution without natural selection: from vocabulary to syntax in a population of learners. *Tech. Rep. Edinb. Occ. Pap. Linguistics*, no. 98–1.
10. Kirby S, Cornish H, Smith K. 2008 Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl Acad. Sci. USA* **105**, 10 681–10 686. (doi:10.1073/pnas.0707835105)
11. Mesoudi A, Whiten A. 2004 The hierarchical transformation of event knowledge in human cultural transmission. *J. Cogn. Cult.* **4**, 1–24. (doi:10.1163/156853704323074732)

12. Mesoudi A, Whiten A. 2008 The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Phil. Trans. R. Soc. B* **363**, 3489–3501. (doi:10.1098/rstb.2008.0129)
13. Deacon TW. 1997 *The symbolic species: the co-evolution of language and the brain*. New York: WW Norton & Company.
14. Kirby S. 2001 Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* **5**, 102–110. (doi:10.1109/4235.918430)
15. Zuidema W. 2003 How the poverty of the stimulus solves the poverty of the stimulus. In *Advances in neural information processing systems 15 (Proceedings of NIPS'02)* (eds S Becker, S Thrun, K Obermayer), pp. 51–58. Cambridge, MA: MIT Press.
16. Christiansen MH, Chater N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–509. (doi:10.1017/S0140525X08004998)
17. Zuidema W. 2013 Language in nature: on the evolutionary roots of a cultural phenomenon. In *The language phenomenon* (eds P Binder, K Smith). Berlin, Germany: Springer.
18. Smith EC, Lewicki MS. 2006 Efficient auditory coding. *Nature* **439**, 978–982. (doi:10.1038/nature04485)
19. Burns EM. 1999 Intervals, scales, and tuning. In *The psychology of music* (ed. D Deutsch), pp. 215–264. San Diego, CA: Academic Press.
20. Large EW. 2011 A dynamical systems approach to musical tonality. In *Nonlinear dynamics in human behavior* (eds R Huys, V Jirsa), pp. 193–211. New York, NY: Springer.
21. Large EW, Almonte FV. 2012 Neurodynamics, tonality, and the auditory brainstem response. *Ann. NY Acad. Sci.* **1252**, E1–E7. (doi:10.1111/j.1749-6632.2012.06594.x)
22. Lerud KD, Almonte FV, Kim JC, Large EW. 2014 Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hear. Res.* **308**, 41–49. (doi:10.1016/j.heares.2013.09.010)
23. Lots IS, Stone L. 2008 Perception of musical consonance and dissonance: an outcome of neural synchronization. *J. R. Soc. Interface* **5**, 1429–1434. (doi:10.1098/rsif.2008.0143)
24. Gill KZ, Purves D. 2009 A biological rationale for musical scales. *PLoS ONE* **4**, e8144. (doi:10.1371/journal.pone.0008144)
25. Bidelman GM, Krishnan A. 2009 Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* **29**, 13 165–13 171. (doi:10.1523/JNEUROSCI.3900-09.2009)
26. McDermott JH, Lehr AJ, Oxenham AJ. 2010 Individual differences reveal the basis of consonance. *Curr. Biol.* **20**, 1035–1041. (doi:10.1016/j.cub.2010.04.019)
27. Cousineau M, McDermott JH, Peretz I. 2012 The basis of musical consonance as revealed by congenital amusia. *Proc. Natl Acad. Sci. USA* **109**, 19 858–19 863. (doi:10.1073/pnas.1207989109)
28. Kuhl P. 1988 Auditory perception and the evolution of speech. *Hum. Evol.* **3**, 19–43. (doi:10.1007/BF02436589)
29. Wright AA, Rivera JJ, Hulse SH, Shyan M, Neiworth JJ. 2000 Music perception and octave generalization in rhesus monkeys. *J. Exp. Psychol.* **129**, 291–307. (doi:10.1037/0096-3445.129.3.291)
30. Merker B. 2006 The uneven interface between culture and biology in human music (commentary). *Music Percept.* **24**, 95–98. (doi:10.1525/mp.2006.24.1.95)
31. Mayfield GR. 1934 The mockingbird's imitation of other birds. *Migrant* **5**, 17–19.
32. Dowsett-Lemaire F. 1979 The imitation range of the song of the Marsh Warbler, *Acrocephalus palustris*, with special reference to imitations of African birds. *Ibis* **121**, 453–468. (doi:10.1111/j.1474-919X.1979.tb06685.x)
33. Baylis JR. 1982 Avian vocal mimicry: its function and evolution. In *Acoustic communication in birds*, vol. 2 (eds DE Kroodsma, EH Miller), pp. 51–83. New York, NY: Academic Press.
34. Taylor H. 2008 Decoding the song of the pied butcherbird: an initial survey. *Transcult. Music Rev.* **12**, 1–30.
35. Taylor H. 2009 Towards a species songbook: illuminating the vocalisations of the Australian pied butcherbird (*Cracticus nigrogularis*). PhD thesis, University of Western Sydney, Australia.
36. Taylor H, Lestel D. 2011 The Australian pied butcherbird and the natureculture continuum. *J. Interdisc. Music Stud.* **5**, 57–83.
37. Kirby S, Hurford J. 2002 The emergence of linguistic structure: an overview of the iterated learning model. In *Simulating the evolution of language* (eds A Cangelosi, D Parisi), pp. 121–148. London, UK: Springer Verlag.
38. Merker B, Okanoya K. 2007 The natural history of human language: bridging the gaps without magic. In *Emergence of communication and language* (eds C Lyon, L Nehaniv, A Cangelosi), pp. 403–420. London, UK: Springer.
39. Merker B. 2012 The vocal learning constellation: imitation, ritual culture, encephalization. In *Music, language and human evolution* (ed. N Bannan), pp. 215–260. Oxford, UK: Oxford University Press.
40. Morley I. 2002 Evolution of the physiological and neurological capacities for music. *Camb. Archaeol. J.* **12**, 195–216. (doi:10.1017/S0959774302000100)
41. Morley I. 2014 A multi-disciplinary approach to the origins of music: perspectives from anthropology, archaeology, cognition and behaviour. *J. Anthropol. Sci.* **92**, 147–177. (doi:10.4436/JASS.92008)
42. Merker B. 2000 Synchronous chorusing and human origins. In *The origins of music* (eds NL Wallin, B Merker, S Brown), pp. 315–327. Cambridge, MA: MIT Press.
43. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how does it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
44. Abler WL. 1989 On the particulate principle of self-diversifying systems. *J. Soc. Biol. Struct.* **12**, 1–13. (doi:10.1016/0140-1750(89)90015-8)
45. Merker B. 2002 Music: the missing Humboldt system. *Musicae Scientiae* **6**, 3–21. (doi:10.1177/102986490200600101)
46. Fisher RA. 1930 *The genetical theory of natural selection*. Oxford, UK: The Clarendon Press.
47. Balzano G. 1982 The pitch set as a level of description for studying musical pitch perception. In *Music, mind and brain* (ed. M Clynes), pp. 321–351. New York, NY: Plenum Press.
48. Krumhansl CL. 1990 *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.
49. Oudeyer P-Y. 2006 *Self-organization in the evolution of speech*. Oxford, UK: Oxford University Press.
50. Zuidema W, de Boer B. 2009 The evolution of combinatorial phonology. *J. Phon.* **37**, 125–144. (doi:10.1016/j.wocn.2008.10.003)
51. Thorpe WH. 1961 *Bird song*. Cambridge, UK: Cambridge University Press.
52. Clement CJ, Koopmans-van Beinum FJL, Pols CW. 1996 Acoustical characteristics of sound production of deaf and normally hearing infants. *Spoken Lang.* **3**, 1549–1552. (doi:10.1109/ICSLP.1996.607914)
53. Oller DK, Eilers RE. 1998 The role of audition in infant babbling. *Child Dev.* **59**, 441–449. (doi:10.2307/1130323)
54. Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
55. Konishi M. 2004 The role of auditory feedback in birdsong. *Ann. NY Acad. Sci.* **1016**, 463–475.
56. Liu W-C, Wada K, Nottebohm F. 2009 Variable food begging calls are harbingers of vocal learning. *Public Libr. Sci. One* **4**, e5929. (doi:10.1371/journal.pone.0005929)
57. Janik VM, Slater PJB. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99. (doi:10.1016/S0065-3454(08)60377-0)
58. Janik VM, Slater PJB. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/ambe.2000.1410)
59. Ziegler HP, Marler P (eds). 2004 Behavioral neurobiology of birdsong. *Ann. NY Acad. Sci.* **1016**, 1–788.
60. 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)
61. Kroodsma DE, Baylis JR. 1982 A world survey of evidence for vocal learning in birds. In *Acoustic communication in birds* (eds DE Kroodsma, EH Miller), pp. 311–337. New York, NY: Academic Press.
62. Jarvis ED. 2006 Selection for and against vocal learning in birds and mammals. *Ornithol. Sci.* **5**, 5–14. (doi:10.2326/osj.5.5)
63. Nottebohm F. 1976 Vocal tract and brain: a search for evolutionary bottlenecks. *Ann. NY Acad. Sci.* **280**, 643–649. (doi:10.1111/j.1749-6632.1976.tb25526.x)

64. Cross I *et al.* 2013 Culture and evolution. In *Language, music and the brain* (ed. MA Arbib), pp. 540–562. Cambridge, MA: MIT Press.
65. Ermentrout B. 1991 An adaptive model for synchrony in the firefly *Pteroptyx malaccaea*. *J. Math. Biol.* **29**, 571–585. (doi:10.1007/BF00164052)
66. Madison G. 2000 On the nature of variability in isochronous serial interval production. In *Rhythm perception and production* (eds P Desain, L Windsor), pp. 95–113. Lisse, The Netherlands: Swets and Zeitlinger.
67. Repp BH. 2005 Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* **12**, 969–992. (doi:10.3758/BF03206433)
68. Arom S. 1991 *African polyphony and polyrhythm: musical structure and methodology*, pp. 179ff. Cambridge, UK: Cambridge University Press.
69. Merker B, Madison G, 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)
70. Fraisse P. 1982 Rhythm and tempo. In *The psychology of music* (ed. D Deutsch), pp. 149–180. London, UK: Academic Press.
71. 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)
72. Merchant H, Honing H. 2013 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)
73. 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/jappphysiol.00138.2005)
74. Sisondo E. 1991 Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* **249**, 55–58. (doi:10.1126/science.249.4964.55)
75. Nagasaka Y, Chao ZC, Hasegawa N, Notoya T, Fujii N. 2013 Spontaneous synchronization of arm motion between Japanese macaques. *Sci. Rep.* **3**, 1151. (doi:10.1038/srep01151)
76. 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)
77. Hoeschele M, Merchant H, Kikuchi Y, Hattori Y, ten Cate C. 2015 Searching for the origins of musicality across species. *Phil. Trans. R. Soc. B* **370**, 20140094. (doi:10.1098/rstb.2014.0094)
78. Serpell J. 1981 Duets, greetings and triumph ceremonies: analogous displays in the parrot genus *Trichoglossus*. *Z. Tierpsychol.* **55**, 268–283. (doi:10.1111/j.1439-0310.1981.tb01272.x)
79. Greenfield MD. 1994 Cooperation and conflict in the evolution of signal interactions. *Ann. Rev. Ecol. Syst.* **25**, 97–126. (doi:10.1146/annurev.es.25.110194.000525)
80. Greenfield MD. 2005 Mechanisms and evolution of communal sexual displays in arthropods and anurans. *Adv. Study Behav.* **35**, 1–62. (doi:10.1016/S0065-3454(05)35001-7)
81. Morin JG. 1986 Fireflies of the sea: luminescent signaling in marine ostracode crustaceans. *Florida Entomol.* **69**, 105–121. (doi:10.2307/3494749)
82. Attenborough D. 1995 *The private life of plants: a natural history of plant behaviour*. Princeton, NJ: Princeton University Press.
83. Buck J, Buck E, Hanson F, Case JF, Mets L, Atta GJ. 1981 Control of flashing in fireflies. *J. Comp. Physiol* **144**, 277–286. (doi:10.1007/BF00612559)
84. Buck J. 1988 Synchronous rhythmic flashing in fireflies. II. *Q. Rev. Biol.* **63**, 265–289. (doi:10.1086/415929)
85. Buck J, Buck E. 1978 Towards a functional interpretation of synchronous flashing by fireflies. *Am. Nat.* **112**, 471–492. (doi:10.1086/283291)
86. Tuttle MD, Ryan MJ. 1982 The role of synchronized calling, ambient light, and ambient noise in anti-bat-predator behavior of a treefrog. *Behav. Ecol. Sociobiol.* **11**, 125–131. (doi:10.1007/BF00300101)
87. Merker B. 1999/2000 Synchronous chorusing and the origins of music. *Musicae Scientiae* (Special Issue, 1999-2000), 59–73.
88. Katz R. 1982 *Boiling energy: community healing among the Kalahari Kung*. Cambridge, MA: Harvard University Press.
89. Trehub SE, Becker J, Morley I. 2015 Cross-cultural perspectives on music and musicality. *Phil. Trans. R. Soc. B* **370**, 20140096. (doi:10.1098/rstb.2014.0096)
90. Geissmann T. 2000 Gibbon song and human music from an evolutionary perspective. In *The origins of music* (eds NL Wallin, B Merker, S Brown), pp. 103–123. Cambridge, MA: MIT Press.
91. Fitch WT. 2005 The evolution of music in comparative perspective. *Ann. NY Acad. Sci.* **1060**, 1–20. (doi:10.1196/annals.1360.004)
92. Reynolds V, Reynolds R. 1965 Chimpanzees of the Budongo Forest. In *Primate behavior: field studies of monkeys and apes* (ed. I Devore), pp. 368–424. New York, NY: Holt, Rinehart and Winston.
93. Sugiyama Y. 1969 Social behavior of chimpanzees in the Budongo Forest, Uganda. *Primates* **9**, 225–258. (doi:10.1007/BF01730972)
94. Sugiyama Y. 1972 Social characteristics and socialization of wild chimpanzees. In *Primate socialization* (ed. FE Poirer), pp. 145–163. New York, NY: Random House.
95. Wrangham RW. 1975 The behavioural ecology of chimpanzees in Gombe National Park, Tanzania. PhD thesis, University of Cambridge, Cambridge, UK.
96. Wrangham RW. 1979 On the evolution of ape social systems. *Soc. Sci. Int.* **18**, 335–368. (doi:10.1177/053901847901800301)
97. Ghiglieri MP. 1984 *The chimpanzees of Kibale Forest*. New York, NY: Columbia University Press.
98. Ember CR. 1978 Myths about hunter-gatherers. *Ethnology* **17**, 439–448. (doi:10.2307/3773193)
99. Pusey A. 1979 Inter-community transfer of chimpanzees in Gombe National Park. In *The great apes* (eds D Hamburg, E McCown), pp. 465–479. Menlo Park, CA: Benjamin/Cummings.
100. Hagen EH, Bryant GA. 2003 Music and dance as a coalition signaling system. *Hum. Nat.* **14**, 21–51. (doi:10.1007/s12110-003-1015-z)
101. Gingras B, Honing H, Peretz I, Trainor LJ, Fisher SE. 2015 Defining the biological bases of individual differences in musicality. *Phil. Trans. R. Soc. B* **370**, 20140092. (doi:10.1098/rstb.2014.0092)
102. Frank SA. 1998 *Foundations of social evolution*. Princeton, NJ: Princeton University Press.
103. Darwin C. 1871 *The descent of man and selection in relation to sex*. New York, NY: D. Appleton & Co.
104. Elton S. 2008 The environmental context of human evolutionary history in Eurasia and Africa. *J. Anat.* **212**, 377–393. (doi:10.1111/j.1469-7580.2008.00872.x)
105. Klein R. 2009 *The human career: human biological and cultural origins*, 3rd edn. Chicago, IL: Chicago University Press.
106. Fitch WT. 2000 The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267. (doi:10.1016/S1364-6613(00)01494-7)
107. Lewin R, Foley R. 2004 *Principles of human evolution*, pp. 164–192, 2nd edn. Oxford, UK: Blackwell.
108. Robson S, Wood B. 2008 Hominin life history: reconstruction and evolution. *J. Anat.* **212**, 394–425. (doi:10.1111/j.1469-7580.2008.00867.x)
109. Falk D. 2004 Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* **27**, 491–503. (doi:10.1017/S0140525X04000111)
110. Coqueugniot H, Hublin J-J, Veillon F, Houët F, Jacob T. 2004 Early brain growth in *Homo erectus* and implications for cognitive ability. *Nature* **431**, 299–302. (doi:10.1038/nature02852)
111. O'Connell C, DeSilva J. 2013 Mojokerto revisited: evidence for an intermediate pattern of brain growth in *Homo erectus*. *J. Hum. Evol.* **65**, 156–161. (doi:10.1016/j.jhevol.2013.04.007)
112. MacLarnon AM, Hewitt GP. 1999 The evolution of human speech: the role of enhanced breathing control. *Am. J. Phys. Anthropol.* **109**, 341–363. (doi:10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2)
113. Sundberg J. 1987 *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press.