

Title:

Five fundamental constraints on theories of the origins of music

Authors:

Bjorn Merker, Iain Morley, and Willem Zuidema

Corresponding author:

Bjorn Merker
Fjälkestadv. 410-82
SE-29194 Kristianstad
Sweden
E-mail: gyr694c@tninet.se
Phone: +26 (0)46 2880707

Five fundamental constraints on theories of the origins of music

Bjorn Merker, Iain Morley, Jelle Zuidema

Abstract:

The origin of music has been a topic of controversy since the mid-18th century. It entered the domain of evolutionary theorizing proper with Darwin's treatment in his "The descent of man and selection in relation to sex" of 1871. The fact that the topic has not come to rest in the intervening century and a half reflects its inherent complexity and difficulty. Human music is exceedingly diverse structurally: a traditional African percussion ensemble performance and contented humming of a tune in solitude share little that might guide theories of music origins. Moreover, no animal species exhibits a pattern diversity in its expressive behavior even remotely approaching the scope of human music. Faced with such obstacles, theorizing needs constraints from the nature of music, the capacities it engages, and the contexts in which it occurs in order to make progress. Here we propose and examine five fundamental such constraints that bear on theories of how music may have originated and where some of its features may have come from. They are, 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.

Key words:

Culture; Entrainment; Evolution; generativity; Music; Vocal learning.

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 (Brown DE 1991; Brown S and Jordania 2011; von Humboldt 1836). 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 (1998) 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 behavior 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 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" (Byrne 2000, 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 homologies, 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 favoring their evolution (see Fitch, this volume). 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.

Constraint # 1: Cultural transmission

Music, like language, is a complex product of cultural history. Its present-day patterns rest on traditions extending back for many tens of thousands of years of intergenerational transmission of learned cultural lore (Higham et al. 2012; Morley, 2013). 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, intergenerational 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 (Bartlett 1932; Kirby 1998; Kirby et al. 2008; Mesoudi and Whiten 2004, 2008). Thus, to go in search of evolutionary explanations for all aspects of cultural phenomena that have such an origin 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 transmissibility, including small differences in learnability and ease of processing, come to transform the cultural corpus in predictable ways, amounting to a process of informational "compression" over many generations. This can lead to a tight fit between properties of the learner and properties of the cultural lore, and to similarities between the traits of behaviours of different, separated, populations, all without invoking 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., Deacon, 1997; Kirby, 2001; Zuidema, 2003; Christiansen & Chater, 2007). This work has led to a growing consensus (1) that pure cultural evolution is a powerful mechanism, (2) that many features of languages are potentially best understood as resulting from cultural adaptation to (preexisting) hominin cognitive and physiological features, and (3) 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 (Christiansen and Chater 2008; Zuidema 2003). 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 (2013) discusses the finding of Smith & Lewicki (2006) 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 preexisting 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 appear explicable as consequences of cultural evolution. Till recently, the failure of most musical tuning systems to conform to the mathematics of small integer ratios were grounds for rejecting Pythagoras' proposal that small integer frequency ratios account for the preception of musical consonance and harmonicity (Burns 1999). However, recent modeling of the cumulative effects of physiological nonlinearities at each way-station of the ascending auditory pathway has disclosed the presence of "resonance neighborhoods" at whole integer ratio spacings on the tonotopic maps of the auditory system (Large 2011; Large and Almonte 2012; Lerud et al. 2014). This finding not only accomodates a wide range of tuning systems and musical scales found world-wide, 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 (Large 2011; see also Gill and Purves 2009; Bidelman and Krishnan 2009; Cousineau et al. 2012; McDermott et al. 2010; Shapira Lots and Stone 2008).

The auditory resonance biases explored by Ed Large are likely to be shared by all mammals, because his modeling relied only on quite elementary properties of the neural circuitry in question. They did not evolve for purposes of music, in other words, but as incidental design features of an auditory system evolved to

process natural sounds efficiently by recruiting the tuning properties of interacting excitation-inhibition (for which see Merker 2013). Not being confined to humans (see, for example, Wright et al. 2000; Kuhl, 1988), they are 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 require an explanation.

The fact that those musical practices are, indeed, cultural patterns learned 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 (Merker 2006). Through the external loop of inter-generational transmission of learned musical lore, the 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 #3), but one emancipated from innate song templates, for which there is precedence in the true mimics among the birds (Baylis 1982; Dowsett-Lemaire 1979; Mayfield 1934). 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 Kuhl, 1988).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 (Kirby 2001, 2002), the constraint is a real one. One proposal for what might meet it invokes the special circumstances attending the emergence of *Homo erectus* as selection pressures for the evolution of vocal learning for song in early *Homo*. (Merker and Okanoya 2007; Merker 2012). An increasingly refined vocal communication system for the accurate communication and extraction of emotional information from vocal prosody is likely to have contributed to this process as well (Morley, 2002; 2013; 2014).

More generally, placing the ineluctably 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.

Constraint # 2: Generativity, or infinite variety by finite means

Music, like language, is generative, i.e. it produces infinite pattern variety by finite means (von Humboldt 1836/1971). The key to that variety in both music and language is of course not recursion (Hauser et al. 2003) but combinatorics (Abler 1989; Merker 2002; von Humboldt 1836/1971). 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 (Fisher 1930), i.e. they must retain their individuality on combining (see Fig. 1, a, b). When that is the case, each such combination "creates something which is not present per se in any of the associated constituents" (von Humboldt 1836/1971, p. 67), making infinite pattern variety possible (see Fig. 1, c). 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; see Table I).

< Insert Figure 1 here >

< Insert Table 1 here >

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 (Merker 2002). 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 # 4).

In other words, musical notes are not simply pitches. Rather, they are individuated pitch locations within a discretised 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 conventionalised 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 figure can return and which can be used repeatedly in the development of a musical pattern.

This discretisation of the pitch continuum into determinate "pitch sets" supplies music with combinable pitch elements featured in musical melodies and chords (Balzano, 1982; Burns 1999; Krumhansl, 1990). 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 # 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 Balzano 1982; Merker 2006). Perhaps accounts of how discrete combinatoriality (culturally) evolved in phonology (Oudeyer, 2006; Zuidema and de Boer, 2009) can be adapted to music.

If so, there is no need to burden theories of music origins with Darwinian accounts of the origin of musical notes, scales, and tuning systems, but however one conceives of the matter, the point here is only this, that 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.

Constraint # 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, sounding models through a process requiring intact hearing and feedback from one's own voice (Clement et al. 1996; Doupé and Kuhl 1999; Konishi 2004; Liu et al. 2009; Oller and Eilers 1998). The process by which they are acquired is technically known as vocal production learning (Janik and Slater

1997, 2000), a dedicated and highly specialized capacity that has no other common uses besides song and speech in our lives.

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, because other primates lack this capacity (Janik and Slater 1997, 2000), which 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 behavior, socially or environmentally contingent selection among innate calls and their variants, and their learned modification, as detailed in Janik and Slater (2000). 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, and 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 (Konishi 2004). It has been studied in detail above all in birds (Thorpe 1961; Ziegler & Marler 2004), 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 and Brenowitz (2005). 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, the true mimics among the birds (mynahs, many species of parrots, lyrebirds, butcherbirds, mockingbirds, etc., see Baylis 1982) might eventually supplement the invaluable knowledge on 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 learning proper is found only in humans, cetaceans, pinnipeds, elephants, and bats, and among birds in oscine songbirds, parrots and hummingbirds (Janik and Slater, 1997; Jarvis 2006; Kroodsma and Baylis 1982). 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" (Nottebohm 1976) on the path to human music.

Since vocal learning is an apparent human autapomorphy (a trait that evolved *de novo* in our lineage), a theory of music origins must either resign itself to simply assuming it as a precondition without accounting for it, or supply an account of how we came into possession of this capacity at or after parting ways with our common ancestor with chimpanzees. Its origin in our lineage could have been driven by either song, speech, or other factors, and virtually forces the theorist to come to grips with the order of precedence among song and speech in our ancestry (see Cross et al. 2013, Fig. 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" (Nottebohm 1976, 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 # 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, the stubborn fact is our possession of vocal production learning, a fact which any theory of the origins of music leaves unaccounted for at its peril.

Constraint # 4: Entrainment

So far the constraints considered 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 behavior to one another with perfect synchrony.

The type of temporal coordination of inter-individual behavior 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" (Ermentrout 1991). It consists of mutual phase-locking with zero (or even slightly negative) phase lag between the periodic signals of two or more signalers 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 endeavor to make its phenomenology 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 (reviewed by Madison 2000; Repp 2005). The zero (or even slightly negative) phase lag of such entrainment means that its timing mechanism is predictive. A punctate behavior 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 (Arom 1991, pp. 179ff; Merker et al. 2009). It makes an upcoming beat in the sequence perfectly predictable, and therefore targetable by the predictive timing mechanism (Fraisse 1982; Merker et al. 2009). Positive asynchronies large enough to come within reach of auditory reaction time, such as those reported for macaques in a synchronization task (Zarco et al. 2009; Merchant & Honing 2014) are thus automatically excluded as evidence for entrainment of a kind relevant to the human capacity. That asynchronies to irregular sequences are even larger in these macaques is hardly surprising, given the greater demands of reacting to individual events in an irregular sequence.

In interindividual 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 (Ermentrout 1991). Both are well documented for human sensorimotor synchrony (Repp 2005). 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, which in humans

centers on 2 Hz, which is also the human locomotor tempo (Fraisse 1982; MacDougall and Moore 2005). Entrainment precision is dependent on predictability, so variance in period length has to be small, typically exhibiting a standard deviation of a few (two to five) percentage points in human tapping performance (Madison 2000; Repp 2005).

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 by small whole integer ratios for harmonically related tempos (Sismondo 1991). It is from this unique correspondence between the individual events of separate time series that asynchronies and their variability is calculated as a measure of synchronization skill (Repp 2005). 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 (Nagasaka et al. 2013) - 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 by Patel and colleagues of dancing movements performed to rhythmic music by the famous cockatoo Snowball bears on the definition of entrainment as well (Patel et al. 2009; see also Hoeschele et al., this volume). 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 behavior expected of a mechanism dedicated to entrainment through provisions for phase and period correction. Such a mechanism locks on to a steady entraining signal and stays locked with only minor asynchronies as long as the signal remains reasonably steady. The reason the bird does not do so are at this point unknown, but one possibility is that producing dance-like movements mimicking those of its human keepers takes precedence over sustained and precise behavioral 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 (Serpell 1981). 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 behavior 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 crickets, cicadas, crabs and perhaps frogs (Greenfield 1994, 2005). To these may be added some marine bioluminescent crustaceans (Morin 1986), and the rattan ant (Attenborough 1995), 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 (Ermentrout 1991). They entrain their behavior 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 (Buck et al. 1981; Buck 1988; Ermentrout 1991).

That is not to say that the mechanism by which these insects achieve their impressive synchrony is the same

as the human. They only share with us those features of it needed to achieve behavioral entrainment with perfect synchrony. We can easily entrain to a synchronous cicada chorus, but cicadas are unlikely to entrain to our favorite 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 beside our own of genuine beat-based group synchrony that plays a role in the natural behavior of the species in question. These species therefore may provide us with hints about our own path to this rare behavior 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 behavioral repertoire of non-human species in the wild are confined to special cases of mate attraction and predation defense (reviewed by Greenfield 1994, 2005). Among the former the so called "beacon effect" (Buck and Buck 1978) takes pride of place, featuring thousands of synchronously signaling 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% (Buck et al. 1981). Through their entrained luminescent signaling, single trees of permanent congregation are converted to flashing "beacons" visible from all directions despite the foliage that obstructs single lines of sight in the tropical rain forest. 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 neighboring 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 (Tuttle and Ryan 1982). 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 the behavior 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 (Attenborough 1995). 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. It results from entrainment of the alarm behavior 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 produces the unexpected rattle. Amplitude summation (auditory "beacon effect") is the key to this defensive use of inter-individual synchrony.

The narrow compass of behavioral 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. All the more noteworthy is this constraint in that it pertains specifically to human music and little else in our behavioral repertoire except the music-related disciplines of dance and drill. As such it provides an invaluable asset in evolutionary scenario building, and has been exploited in a conjectural account of its origin in our lineage by one of us (Merker 1999/2000; Merker et al. 2009).

Constraint # 5: Motivational basis

Wherever humans live, and however they have organized their societies, they exhibit a behavioral peculiarity of gathering from time to time to sing and dance together in a group (Brown DE 1991; Brown S and Jordania 2011; von Humboldt 1836). By featuring both human song (Constraint # 3) and entrainment (in

the dancing movements and perhaps clapping performed in synchrony with the singing/music, Constraint # 4) such behavior 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 behavior of a musical kind on record.

Add to this ubiquity the fact that such gatherings may be staged entirely without instruments (as in the traditional trance dance of the hunter-gatherers of the Kalahari Desert, Katz 1982; see also Trehub, Becker and Morley, this volume), and it would seem that in this human propensity for occasional group singing and dancing we are in the presence of something like a prototypical musical behavior, the core of the human capacity for music around which other variation occurs. One is assisted in becoming aware of the peculiarity and specificity of this behavioral propensity by attempting to imagine other human behaviours fulfilling the same purposes under those circumstances in which we typically gather to sing and dance together in a group – it is very difficult to do so.

The ubiquity and specificity of this putatively prototypical musical behavior 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 (2000) 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 signaling (“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 (2005), though lacking the pulse-based rhythmicity of drumming in the musical sense. Though these displays do not feature any metric structure resembling isochrony, nor any rhythmic entrainment between individuals, they 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 co-ordination, 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 (Ghiglieri, 1984; Reynolds and Reynolds, 1965; Sugiyama, 1969, 1972; Wrangham, 1975, 1979).

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. The Mokondi Massana “spirit plays” featuring ritual singing and dancing are a significant aspect of the culture of BaYaka pygmy hunter-gatherers inhabiting the Congo-Brazzaville rainforest, which they share with chimpanzees. “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 and the other without musical content, 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 regularising 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 is presumed to have characterized the last common ancestor of humans and chimpanzees (Ember 1978; Merker et al. 2009; Pusey 1979). This pattern is strikingly reminiscent of the male clumping combined with female migration that provides the occasion for synchronous chorusing in the insect examples we have cited, and is assumed to have driven its evolution via the so called "beacon effect" (Merker 1999/2000; Merker et al. 2009). As noted in the latter citation, such a scenario is eminently compatible with the central tenet of the coalition signaling scenario proposed by Hagen and Bryant (2003).

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, this social-emotional aspect of music far antedates the specifically musical form it was given by humans, the secret of this same social-emotional aspect should no longer be sought in its latter-day musical content, but rather in the decidedly non-musical adaptations of our hominoid ancestors along paths of homology.

The evolutionary context

In the course of detailing the foregoing constraints, we have noted that some distinguishing aspects of music (for example, 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 (for example, 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 (Gingras et al., this volume), 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 (Frank, 1998).

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 (sexual selection; Darwin, 1878). 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 since 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 (for example, 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 and chimpanzees (around 6-7 million years ago) and the appearance of our own species (around 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 in both 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 (e.g. Klein, 2009; Elton, 2008). Meanwhile the physiological characteristics and ecological contexts of early *Homo* (*Homo habilis*, early African *Homo erectus*, and their descendants such as *Homo 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 (e.g. Klein, 2009; Elton, 2008). The efficacy of any proposed alterations to ancestral 'carnival displays', coalitional displays (e.g. Hagan and Bryant, 2003), or size-exaggeration vocalisations (Fitch, 2000), for example, needs to be situated

within the context of these changes (see also Morley, 2013).

The mating strategies of human ancestors, and the social organisation 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. In contrast, monogamous species compete for access to a single female with which a long-term cooperative bond must also be formed. In each case the types of cooperation and competition, and with whom cooperation and competition occurs, vary, and the behaviours leading to success in negotiating alliances and long-term bonds, in display directed at the same sex, and directed at the opposite sex, vary accordingly (see, for example, Lewin & Foley, 2004, 164-192). In the case of human ancestors, high levels of sexual dimorphism and rapid developmental life-history in australopithecines comparable to that of chimpanzees (Robson and Wood, 2008) have been taken to indicate broadly similar polygynous mating strategies and male-female relations as are exhibited by chimpanzees (e.g. Klein, 2009 and references therein). Meanwhile, trends towards a reduction in sexual dimorphism and increased altriciality in the infants of early *Homo* indicates the development of increased cooperative long-term pair-bonding (e.g. Klein, 2009 and references therein). As noted, these developments have a direct influence upon the forms and efficacy of behaviours related to display, sexual selection, pair-bonding and vocalisation between adults (see also Merker 2012). 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. Falk, 2004), and the learning opportunities of infants and juveniles (e.g. Coqueugniot et al. 2004; Robson and Wood, 2008; O'Connell and DeSilva, 2013).

Similarly the potential value and form of vocal learning capabilities should be tested and understood against the backdrop of physiological constraints for vocalisation capabilities. For example, MacLarnon and Hewitt (1999) 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 vocalisations of extended duration, with greater control over intensity and pitch contour. They conclude that the level of fine control of breathing musculature in early African *H. erectus* (c. 1.8 million years ago) was not increased relative to that of chimpanzees or the earlier australopithecines, but that by the time of its immediate descendent, *H. heidelbergensis* (from around 5-600,000 years ago) the level of control would have been equivalent to that of modern *H. sapiens*.

Whilst we are able to conclude that neural connections allowing deliberate planning, fine control and integration of laryngeal, orofacial and respiratory musculature used in vocalisations (as would be required for vocal learning) emerged in the lineage since our last common ancestor with chimpanzees, the fossil record of hominin brain changes does not provide a record of changes to internal structures such as these. However, if the development of the thoracic neurology for voluntary extended breathing control was to be useful for vocalisation, 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 developed in tandem, during the 1 million-or-so-year period between early African *H. erectus* and *H. heidelbergensis* (see also Morley, 2013).

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; in 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. The distinctions between them have not always been clearly made in the literature discussing evolutionary rationales for musical behaviours, however. 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.

Footnotes:

Footnote 1. This 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 birdsong, 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 (Taylor 2008, 2009, 2011). 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:

- Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12, 1-13.
- Arom S. African (1991). *Polyphony and Polyrhythm: Musical Structure and Methodology*. Cambridge, UK: Cambridge University Press.
- Attenborough D (1995). *The Private Life of Plants: A Natural History of Plant Behaviour*. Princeton, NJ: Princeton University Press.
- Balzano, G. (1982). The pitch set as a level of description for studying musical pitch perception. In M. Clynes (Ed.), *Music, Mind and Brain* (pp. 321-351). New York: Plenum Press.
- Bartlett, F. C. 1932 *Remembering*. Oxford, UK: Macmillan.
- Baylis JR (1982). Avian vocal mimicry: its function and evolution. In DE Kroodsma and EH Miller (Eds.), *Acoustic Communication in Birds*, Vol. 2, pp. 51-83. Academic Press, New York.
- Beecher MD and Brenowitz EA (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, 20, 143–149.
- Bidelman, G.M., Krishnan, A., (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*. 29, 13165-13171.
- Buck, J. (1988). Synchronous rhythmic flashing in fireflies. II., *Q. Rev. Biol.* 63, 265-289.
- Buck J., Buck E., Hanson F., Case, J. F., Mets, L., Atta, G. J. (1981). Control of flashing in fireflies. IV, Free run pacemaking in a synchronic *Pteroptyx*. *Journal of Comparative Physiology* 144, 277-286.
- Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 215-264). San Diego: Academic Press.
- Byrne RW (2000). Evolution of Primate Cognition. *Cognitive Science*, 24, 543–570.

Christiansen, MH & Chater N (2009). Language as shaped by the brain. *Behavioral and Brain Sciences* 31:489–558.

Clement CJ, Koopmans-van Beinum FJL, Pols CW (1996) Acoustical characteristics of sound production of deaf and normally hearing infants. *Spoken Language* 3: 1549–1552.

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.

Cousineau M, McDermott JH, & Peretz I (2012) The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences* 109(48):19858–19863.

Cross I, Fitch WT, Aboitiz F, Iriki A, Jarvis ED, Lewis J, Liebal K, Merker B, Stout D and Trehub SE (2013). Culture and evolution. In MA Arbib (Ed.), *Language, Music and the Brain*, pp. 540–562. Cambridge, MA: The MIT Press.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.

Doupé AJ and Kuhl PK (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* 22, 567–631.

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.

Elton, S. (2008) The environmental context of human evolutionary history in Eurasia and Africa. *Journal of Anatomy* 212, 377–393.

Ember CR (1978). Myths about hunter–gatherers. *Ethnology*, 17: 439–448.

Ermentrout B (1991). An adaptive model for synchrony in the firefly *Pteroptyx malaccae*. *Journal of Mathematical Biology*, 29: 571–585.

Falk, D. (2004b) Prelinguistic evolution in early hominins: whence motherese? *Behavioural and Brain Sciences* 27, 491–541.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: The Clarendon Press.

Fitch WT (2005). The Evolution of Music in Comparative Perspective. *Annals of the New York Academy of Sciences* 1060: 1–20. doi:10.1196/annals.1360.004

Fitch, W. (2000) The evolution of speech: a comparative review. *Trends in Cognitive Science* 4, 258–67.

Fraisse P. (1982). Rhythm and tempo. In Deutsch D (Ed), *The Psychology of Music*, pp. 149–180. London: Academic Press.

Geissmann T (2000). Gibbon song and human music from an evolutionary perspective. In NL Wallin, B Merker and S Brown, eds, *The origins of music*, pp. 103–123. The MIT Press, Cambridge, MA.

Ghiglieri MP (1984). *The Chimpanzees of Kibale Forest*. New York: Columbia University Press.

- Gill KZ & Purves D (2009) A Biological Rationale for Musical Scales. *PLoS ONE* 4(12):e8144.
- Greenfield MD (1994). Cooperation and conflict in the evolution of signal interactions. *Annual Review of Ecological Systems*, 25: 97–126.
- Greenfield MD (2005). Mechanisms and evolution of communal sexual displays in arthropods and anurans. *Advances in the study of behavior* 35: 1-62.
- Hagen EH and Bryant GA (2003). Music and dance as a coalition signaling system. *Human Nature*, 14: 21–51.
- Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how does it evolve? *Science* 298:1569–1579.
- Higham T, Basell L, Jacobi R, Wood R, Bronk Ramsey C and Conard NJ (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of Geißenklösterle. *Journal of Human Evolution* 62, 664–676. DOI: 10.1016/j.jhevol.2012.03.003
- Janik, V. M. and Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26, 59–99.
- Janik VM and Slater PJB (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, 60, 1–11.
- Jarvis ED (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science* 5: 5–14.
- Katz R (1982). *Boiling Energy: Community Healing Among the Kalahari Kung*. Cambridge, MA; Harvard University Press.
- Kirby, S. (1998). Language evolution without natural selection: From vocabulary to syntax in a population of learners. Technical Report, Edinburgh Occasional Papers in Linguistics, 98–1, Department of Linguistics, University of Edinburgh.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102-110.
- Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, chapter 6, pages 121-148. Springer Verlag, London.
- Kirby, S., Cornish, H. & Smith, K. (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences USA* 105: 10681–10686. doi: 10.1073#pnas.0707835105
- Klein, R. (2009) *The Human Career: Human Biological and Cultural Origins* (3rd Edition). Chicago University Press, Chicago.
- Konishi M (2004). The role of auditory feedback in birdsong. In HP Ziegler and P Marler, eds, *The Behavioral Neurobiology of Birdsong*. *Annals of the New York Academy of Sciences*, 1016, 463–475.
- Kroodsma, D. E. & J. R. Baylis (1982). A world survey of evidence for vocal learning in birds. In *Acoustic communication in birds*. D. E. Kroodsma & E. H. Miller, Eds.: pp. 311–337. Academic Press. New York.
- Kuhl, P. (1988) Auditory perception and the evolution of speech. *Human Evolution* 3, 19–43.

- Large EW (2011). A dynamical systems approach to musical tonality. In R Huys and V Jirsa (Eds.), *Nonlinear Dynamics in Human Behavior*, pp. 193-211. New York: Springer.
- Large EW and Almonte FV (2012). Neurodynamics, tonality, and the auditory brainstem response. *Annals of the New York Academy of Sciences* 1252(1):E1-E7.
- Lerud KD, Almonte FV, Kim JC, & Large EW (2014) Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hearing Research* 308: 41-49. doi:10.1016/j.heares.2013.09.010
- Lewin, R. and Foley, R. (2004) *Principles of Human Evolution* (2nd Edition). Oxford: Blackwell.
- Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. (In D. Scarborough and S. Sternberg (Eds.), *An invitation to cognitive science, Volume 4: Methods, models, and conceptual issues*. Cambridge, MA: MIT Press.), p. 130.
- Liu W-c, Wada K and Nottebohm F (2009). Variable food begging Calls are harbingers of vocal learning. *Public Library of Science One* 4(6): e5929. doi:10.1371/journal.pone.0005929
- MacDougall HG and Moore ST (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology*, 99: 1164–1173, 2005.
- MacLarnon, A., and Hewitt, G. (1999) The evolution of human speech: the role of enhanced breathing control. *American Journal of Physical Anthropology* 109, 341–63.
- Madison, G. (2000). On the nature of variability in isochronous serial interval production. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 95-113). Lisse, The Netherlands: Swets and Zeitlinger.
- Mayfield GR (1934). The mockingbird's imitation of other birds. *Migrant*, 5, 17–19.
- McDermott JH, Lehr AJ, & Oxenham AJ (2010) Individual Differences Reveal the Basis of Consonance. *Curr. Biol.* 20(11):1035-1041.
- Merchant H and Honing H (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience* 7, Article 274, 1-8. doi:10.3389/fnins.2013.00274
- Merker B (1999/2000). Synchronous chorusing and the origins of music. *Musicae Scientiae Special issue*: 59–73.
- Merker B (2002). Music: The missing Humboldt system. *Musicae Scientiae*, 6: 3-21.
- Merker B (2005). The conformal motive in birdsong, music and language: An introduction. In G Avanzini, L Lopez, S Koelsch and M Majno, eds, *The Neurosciences and Music II: From Perception to Performance*. *Annals of the New York Academy of Sciences*, 1060, pp. 17–28.
- Merker B (2006). The uneven interface between culture and biology in human music (commentary). *Music Perception*, 24: 95-98 (doi:10.1525/mp.2006.24.1.95)
- Merker B and Okanoya K (2007). The natural history of human language: Bridging the gaps without magic. In C. Lyon, L. Nehaniv & A. Cangelosi (Eds.). *Emergence of Communication and Language*, pp. 403-420. London: Springer-Verlag.
- Merker B (2012). The vocal learning constellation: imitation, ritual culture, encephalization. In N.

Bannan (Ed.). Music, language and human evolution, pp. 215-260. Oxford: Oxford University Press.

Merker B (2013). Cortical gamma oscillations: the functional key is activation, not cognition. *Neuroscience and Biobehavioral Reviews* 37: 401–417.

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

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

Mesoudi A and Whiten A (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Phil. Trans. R. Soc. B* (2008) 363, 3489–3501. doi:10.1098/rstb.2008.0129

Morin, J. G. (1986). Firefleas of the sea: Luminescent signaling in marine ostracode crustaceans. *Fla. Entomol.* 69, 105–121.

Morley, I. (2014) A multi-disciplinary approach to the origins of music: perspectives from anthropology, archaeology, cognition and behaviour. *Journal of Anthropological Sciences* 92, 147-177.

Morley, I. (2013) *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford University Press, Oxford.

Morley, I. (2002). Evolution of the physiological and neurological capacities for music, *Cambridge Archaeological Journal*, Vol.12, No. 2, 195-216.

Nagasaka,Y.,Chao,Z.C.,Hasegawa,N.,Notoya,T.,andFujii,N.(2013). Spontaneous synchronization of arm motion between Japanese macaques. *Scientific Reports* 3, 1151. doi:10.1038/srep01151

Nottebohm F (1976). Discussion paper. Vocal tract and brain: A search for evolutionary bottlenecks. In SR Harnad, HD Steklis and J Lancaster, eds, *Origins and evolution of language and speech*. Annals of the New York Academy of Sciences, 280, 643–649.

O'Connell, C. & DeSilva, J. (2013) Mojokerto revisited: Evidence for an intermediate pattern of brain growth in *Homo erectus*. *Journal of Human Evolution* 65, 156-161.

Oller DK, Eilers RE (1998) The role of audition in infant babbling. *Child Development* 59: 441–449.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Oudeyer, P-Y. (2006). *Self-Organization in the Evolution of Speech*. Studies in the Evolution of Language, Oxford University Press. (Translation by James R. Hurford)

Patel, A.D., Iversen, J.R., Bregman,M.R., and Schulz,I.(2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology* 19, 827–830. doi:10.1016/j.cub.2009.03.038

Payne K (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In NL Wallin, B Merker and S Brown, eds, *The Origins of Music*, pp. 135–150. The MIT Press, Cambridge, MA.

Pusey A (1979). Inter-community transfer of chimpanzees in Gombe National Park. In Hamburg D, and McCown E (Eds), *The Great Apes*, pp. 465–479. Menlo Park, CA: Benjamin/Cummings.

Repp BH (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12: 969–992.

- Reynolds V and Reynolds R (1965). Chimpanzees of the Budongo forest. In Devore I (Ed), *Primate behavior: Field studies of monkeys and apes*. New York: Holt, Rinehart and Winston, pp. 368–424.
- Robson, S. & Wood, B. (2008) Hominin life history: reconstruction and evolution. *Journal of Anatomy* 212, 394-425.
- Serpell J (1981). Duets, greetings and triumph ceremonies: Analogous displays in the parrot genus *Trichoglossus*. *Z. Tierpsychol.* 55: 268-283.
- Shapira Lots I & Stone L (2008) Perception of musical consonance and dissonance: An outcome of neural synchronization. *Journal of The Royal Society Interface* 5(29):1429-1434.
- Sismondo E (1990). Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science*, 249: 55–58.
- Sugiyama Y (1969). Social behavior of chimpanzees in the Budongo Forest, Uganda. *Primates*, 9: 225–258.
- Sugiyama Y (1972). Social characteristics and socialization of wild chimpanzees. In Poirer FE (Ed), *Primate Socialization*. New York: Random House, pp. 145–163.
- Taylor H (2008). Decoding the song of the pied butcherbird: An initial survey. *Transcultural Music Review* 12 (12): 1-30.
- Taylor H (2009). *Towards a Species Songbook: Illuminating the Vocalisations of the Australian Pied Butcherbird (Cracticus nigrogularis)*. Doctoral Dissertation, University of Western Sydney.
- Taylor H and Lestel D (2011). The Australian pied butcherbird and the nature-culture continuum. *Journal of Interdisciplinary Music Studies* 5(1): 57-83.
- Thorpe WH (1961). *Bird song*. Cambridge University Press, Cambridge.
- von Humboldt, W. (1836). *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Royal Academy of Sciences. English translation (1971), G. C. Buck and F. Raven (Trans.). *Linguistic variability and intellectual development*. Baltimore, MD: University of Miami Press.
- Wrangham RW (1975). *The Behavioural Ecology of Chimpanzees in Gombe National Park, Tanzania*. PhD thesis. Cambridge, UK: University of Cambridge.
- Wrangham RW (1979). On the evolution of ape social systems. *Social Science International*, 18: 335–368.
- Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, 129, 291–307.
- Zarco, W., Merchant, H., Prado, L., and Mendez, J. C. (2009). Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology* 102, 3191–3202. doi:10.1152/jn.00066.2009
- Ziegler HP and Marler P (2004). *Behavioral Neurobiology of Birdsong*. Vol. 1016. *Annals of the New York Academy of Sciences*. New York, NY.
- 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)*, ed. S. Becker, S. Thrun & K. Obermayer,

pp. 51–58. MIT Press.

Zuidema, W. (2013). Language in Nature: on the Evolutionary Roots of a Cultural Phenomenon. In: Philippe Binder and Kenny Smith (eds.), The Language Phenomenon, Berlin: Springer

Willem Zuidema and Bart de Boer (2009), The evolution of combinatorial phonology. Journal of Phonetics Volume 37, Issue 2, April 2009, Pages 125-144.

FIGURE LEGEND:

Figure 1:

Figure 1. Cartoon illustration of the "particulate principle of self-diversifying systems", following Abler (Abler 1989). **(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 miniscule sample of the infinite generativity of a combinatorics of as few as two discrete non-averaging elements.

TABLE 1:

Table heading:

Table 1
The principal open-ended or “self-diversifying” systems

Table complete:

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 'paticles'	atoms	nucleotides	notes	phonemes
Particle type	physical		conventional	
domain	nature		culture	