The natural history of musical rhythm: functional and mechanistic theories on the evolution of human rhythm cognition and the relevance of rhythmic animal behaviors

Corresponding author:
Gabriela-Alina Sauciuc, Lund University, Department of Philosophy / Cognitive Science
Corresponding address: Lund University, Department of Philosophy, Box 192, 221 00 Lund.
Phone number: (+46) 70 64 866 86
e-mail address: gabriela-alina.sauciuc@lucs.lu.se

Co-authors
Tomas Persson, Lund University, Department of Philosophy / Cognitive Science
Corresponding address: Lund University, Department of Philosophy, Box 192, 221 00 Lund
e-mail address: tomas.persson@lucs.lu.se

Elainie Madsen, Lund University, Department of Philosophy / Cognitive Science
Corresponding address: Lund University, Department of Philosophy, Box 192, 221 00 Lund
e-mail address: elainie.madsen@lucs.lu.se

Keywords: rhythmic behaviors, rhythm cognition, music evolution, functional theories, mechanistic hypotheses

Abstract

There has recently been a growing interest in investigating rhythm cognition and behavior in nonhuman animals as a way of tracking the evolutionary origins of human musicality – i.e., the ability to perceive, enjoy and produce music. During the last two decades, there has been an explosion of theoretical proposals aimed at explaining why and how humans have evolved into musical beings, and the empirical comparative research has also gained momentum. In this paper, we focus on the rhythmic component of musicality, and review functional and mechanistic theoretical proposals concerning putative prerequisites for perceiving and producing rhythmic structures similar to those encountered in music. For each theoretical proposal we also review supporting and contradictory empirical findings. To acknowledge that the evolutionary study of musicality requires an interdisciplinary approach, our review strives to cover perspectives and findings from as many disciplines as possible. We conclude with a research agenda that highlights relevant, yet thus far neglected topics in the comparative and evolutionary study of rhythm cognition. Specifically, we call for a widened research focus that will include additional rhythmic abilities besides entrainment, additional channels of perception and production besides the auditory and vocal ones, and a systematic focus on the functional contexts in which rhythmic signals spontaneously occur. With this expanded focus, and drawing from systematic observation and experimentation anchored in multiple disciplines, animal research is bound to generate many important insights into the adaptive pressures that forged the component abilities of human rhythm cognition and their (socio)cognitive and (neuro)biological underpinnings.

1. Introduction

There has recently been a growing interest in investigating rhythm cognition and behavior in nonhuman animals (henceforth 'animals'), as a way of tracking the evolutionary origins of human musicality – i.e., the ability to perceive, enjoy and produce music (e.g., Honing, 2019, and references therein). Currently, there is wide agreement that musicality is a biocultural ability supported (and constrained) by a suite of cognitive and emotional traits. Consistent with this view, a **componential approach** to the evolution of musicality is presently advocated by several scholars (Fitch, 2006; Justus & Hutsler, 2005; McDermott & Hauser, 2005; Patel, 2006), whereby the presence and variability of component traits of human musicality ought to be systematically examined across human cultures and in nonhuman species. This approach offers real potential for mapping musicality-relevant traits that are shared with other species, thereby providing unique insights into **why**, **how** and **when** musicality has evolved in our own species.

Since **rhythm** (together with the use of discrete pitches) is commonly accepted as a putative universal in human music (Savage et al., 2015), abilities related to **rhythm cognition** constitute an obvious candidate topic for the comparative study of musicality. Musical rhythm can be broadly defined as the systematic patterning of sounds in time (McAuley, 2010). One of the most salient features of musical rhythm is *isochrony*, i.e., the periodic recurrence of sounds (or sound groupings) at equal time intervals. Rhythmic isochrony makes a musical piece highly recognizable and facilitates the individuals' synchronization with the music (e.g., Ravignani & Madison, 2017; Savage et al., 2015).

While definitions such as the above emphasize the acoustic modality, musical rhythm must also encompass the regular patterning of body movements in time, since, across cultures of the world, dance is universally coupled to music (Cross, 2001; Savage et al., 2015; Richter &

Ostovar, 2015). Ethnographic records suggest, moreover, that music and dance as separate phenomena and separate concepts have emerged only recently in the cultural history of humans. Indeed, it is not unusual that so-called traditional cultures lack specific concepts (and words) for music and / or dance (Cross, 2001; Fitch, 2006; Merker 2000, 2002; Nettl, 2015). Instead, musical behaviors (e.g., dance, song, tool-based sound making) occur as an indissoluble part of rituals or religious ceremonies, and are often named after the specific ritual or purpose they serve (Cross, 2001; Nettl, 2015). This may contrast with how we customarily consume, experience and conceive of music and dance in post-industrial, contemporary societies. Nevertheless, it is likely to be highly informative about **how** and **why** musicality evolved, as hunter-gatherer societies are often considered a more faithful reflection of the socio-ecological niche in which our ancestors evolved (Mithen, 2005). Notably, evolutionary theories focused on the putative adaptive functions of *musicality* are often theories of *dance* evolution as well, insofar as their focus lies on the benefits of rhythmically moving together (as detailed in sections 2 and 3).

The comparative study on rhythm cognition has begun to take off only during the last decade, especially after the seminal study of Patel et al., (2009), who provided experimental evidence of flexible **audiomotor entrainment** in an animal – the sulfur-crested cockatoo (*Cacatua galerita*) Snowball. Ever since, audiomotor entrainment (i.e., the ability to accurately synchronize bodily movements to an exogenous auditory rhythm) has constituted the main focus of comparative research. Emerging evidence suggests that individuals of several species can (be trained to) align motor responses to metronomic and even musical stimuli (Honing, 2019, and references therein). Although limited, this evidence has rekindled the century-long interest in the origins of human musicality (Darwin, 1871), spurring an effervescence of speculations and hypotheses related to the evolution of human rhythmic abilities.

Traditionally, the theoretical debate around the evolution of musicality has revolved much around whether musicality is an adaptation, i.e., whether it has evolved in response to selective pressures from the socio-ecological environment of our ancestors, and thus fulfilled important adaptive functions for the species (e.g., Fitch, 2006; Huron, 2012). Alternative proposals are that musicality is an exaptation (i.e., a trait originally lacking adaptive benefits, but later co-opted in the service of an adaptive function), a spandrel (i.e., a by-product of other adaptations), or a technology (e.g., Dissanayake 2000, 2009a; Fitch, 2006; Huron, 2001, 2012; Kotz et al. 2018; Mithen, 2005; Patel, 2006). Several scholars have recently argued that such debates are counter-productive, since musicality is not a monolithic trait, but encompasses a conglomerate of abilities with distinct adaptive benefits and evolutionary trajectories (e.g., Fitch, 2011; Kotz et al., 2018; Mithen, 2005; Trainor, 2018). In the evolution of complex traits such as musicality, it is generally the case that adaptation, exaptation and spandrel are intertwined and iterated (Andrews et al., 2002). The alternative approach that has gained momentum during the last decade is to separately investigate constituent components of musicality in order to formulate specific functional and/or mechanistic hypotheses that are amenable to empirical testing in both human and animal populations (e.g., Fitch, 2006; Trainor, 2018).

In this paper, we pursue a twofold purpose: (i) we review the main **functional** (section 2) and **mechanistic** (section 3) theoretical proposals on the evolution of musical rhythm, and the relevant empirical evidence that supports each of them; (ii) we outline a **research agenda** for the field of comparative rhythm cognition, and highlight relevant topics that have received little or no empirical treatment (section 4). The reviews in section 2 and 3 rely on both human and nonhuman data, as the aim is to characterize the current status of each theoretical proposal with respect to empirical support. To help the reader navigate these dense review sections, Tables 1-6 provide summaries and key references for the reviewed proposals.

In the research agenda outlined in section 4, however, we focus narrowly on comparative research, although, whenever relevant, human studies will be also reviewed. Specifically, we call for a widened research focus that will include additional rhythmic abilities besides entrainment (section 4.1), additional channels of perception and production besides the auditory and vocal ones (section 4.2), and a systematic focus on the functional contexts in which rhythmic signals spontaneously occur (4.3). The topics proposed within each of these sections, along with aims, significance and suitable methodology are synthesized in Figures 1-3.

Several reviews have been recently published on topics pertaining to the evolution of musical rhythm, either dedicated broadly to the evolution of musicality (or music, e.g., Savage et al., 2021), or to specific aspects of research on rhythm cognition, such as, for example e.g., isochrony (Ravignani & Madison, 2014), audiomotor entrainment in animals (e.g., Wilson & Cook, 2016), experimental methods in comparative rhythm perception (Bouwer et al., 2021), analytical tools (e.g., Ravignani & Norton, 2017). The focus of our review is narrower compared to the former type of reviews, and broader compared to the latter. We concentrate specifically on rhythm cognition and rhythmic behaviors, but without limiting the scope of our focus to particular aspects of rhythmicity. Our ambition is to cover as many aspects and explanatory angles as possible in order to establish links, and invite potential synergies between traditional disciplines, and more recent areas of research and perspectives.

We dedicate separate sections to functional and mechanistic proposals, in order to stress that, although complementary, these are distinct levels of analysis. **Functional** theories address questions about **why** a trait (e.g., a behavior) has evolved, thus attempting to account for the selective pressures that led to the emergence of a trait and the adaptive functions that it putatively fulfilled, given those specific selective pressures. **Mechanistic** theories address questions about **how** the adaptive function(s) of a trait are achieved, thus accounting for the mechanisms (neuroanatomical, neurochemical, genetic, cognitive, socio-cognitive or motivational) that implement it. By reviewing functional and mechanistic proposals in separate sections we also attempt to counter the illusion, which may emerge from the literature, that certain functions are obligatorily and/ or exclusively coupled with certain mechanisms. Instead, we aim to expose the flexibility of function - mechanisms relationships. This reveals limits in the explanatory power of (past and current) attempts to outline grand evolutionary theories by pinpointing *the one* function or *the one* functional context that had precedence in ancestral times, and may have been the primary driving force in the evolution of musical rhythm – and musicality, in general. In contrast, function - mechanism flexibility suggests that current evolutionary theories of musical rhythm (and musicality) are best treated as complementary.

2. Functional theories and derived hypotheses

In today's world, we spend a good portion of our time immersed in music, whether we relax, exercise, drive, play computer games, participate in demonstrations, etc. Research suggests that the rhythmic component of music exerts a plethora of emotional, aesthetic and social effects on humans, which makes it a powerful regulatory tool for the individual. This also applies at group level, with benefits that seemingly range from facilitating group cohesion, thereby cementing trust and cooperation, to communicating cultural identity (as reviewed in section 2.3). It is far from clear, however, to what extent these individual and social benefits reflect the adaptive significance that rhythm-based behaviors had for ancestral hominins.

In this section, we review the most prominent functional theories on the evolution of musicality and musical rhythm, as well as the empirical evidence they have generated (for a quick overview of these, including key references, see Table 1). Before proceeding, it is important to point out that functional proposals are often generically formulated, in the sense that the envisaged trait is musicality (as if it were a monolithic trait), rather than rhythmicity in particular. These proposals, however, tend to highlight adaptive benefits related to rhythmic displays and/ or synchronization effects, which entails that rhythmicity is likely the relevant aspect.

Insert Table 1 about here

2.1. Musicality as a sexually selected trait

Darwin (1871) suggested that, by analogy with other species, music and dance originated in the mating displays of our ancestors, and speculated that the strong emotional impact that music exerts on humans owes to these courtship origins. He argued that both musical notes and rhythms are vehicles by which primeval displays advertised physical or mental fitness to the opposite sex. This theory has been revived and elaborated by Miller (2000). As an argument for the sexual selection hypothesis, Miller draws a parallel between the high complexity and energy expenditure of mating displays in other species and the musical behaviors of tribal human societies, both of which seemingly involve the largest and most energy-craving muscles in the human body, and commonly feature repeated high stepping, stamping, and jumping. Darwin's initial proposal did not imply a sex bias in the production of protomusical behaviors. Miller, however, sides with the idea that musical behaviors have emerged as a male-specific display, with arguments that, e.g., music production is more widespread in men than women, and peaks in men around 30 years of age, which is claimed to be the prime age for courtship.

According to Miller (2000), musical behaviors function both as honest fitness signals and aesthetic displays. As an honest fitness signal, for example, a well-sustained rhythm – whether enacted in song, drumming or movement – would advertise aerobic fitness, health, strength, stamina, and good motor coordination, as well as cognitive abilities, such as memory, learning, and the processing of complex sequences. Such skills, it is argued, are valuable for persistence hunting, which involves party coordination and prolonged tracking of prey, and is currently practiced by hunter-gatherer communities, and, thus, was likely also practiced by our ancestors.

As aesthetic displays, musical behaviors would be signal traits shaped by evolutionary processes that exploited species-specific sensory biases, thereby setting up sexual-selection pressures in favor of behaviors that appealed to such biases. In the field of animal communication, this process is captured by the **sensory exploitation** hypothesis (e.g., Ryan et al., 1990)*,* according to which the preference for certain signal traits may be the outcome of a sensory bias that already existed in receivers before the signalers evolved the traits to exploit it.

An incontestable merit of Miller's comprehensive proposal is that it sketches several empirically testable hypotheses, and suggests exploratory analyses aimed at assessing his theory. Several of these hypotheses concern the predictive power of musical skills with respect to reproductive success and the physical or mental aptitudes purportedly advertised by musical skill. The exploratory research suggested by Miller revolves around the hypothesis that musical behaviors function as aesthetic display, which may entail that rhythmicity prompts an optimal level of neural excitation in the receiver. Consequently, Miller suggests studies aimed at inventorying stimuli that are optimally exciting – and, thereby, attractive for our species – and assessing whether such optimality ranges extend across closely related species.

Currently, the empirical evidence to evaluate the theory that musicality has evolved as a sexually selected trait comes from three areas of investigation: *ethnographic data (i), experimental studies with humans (ii),* and *animal communication studies (iii)*. The *ethnographic evidence (i)* is ambiguous, with both consistent and divergent records. For example, Malinowksi (1929) and Sachs (1962) provide several examples of musical behaviors used competitively, as male displays. Moreover, a recent study has revealed that, across cultures of the world, singing and instrumental music are predominantly performed by males (Savage et al., 2015). Counter-examples, however, are also abundant in ethnographic records. It is relatively common in traditional societies that older men and women, who are well-beyond the optimal reproductive age, are regarded as the most skilled performers. There are also communities, such as the Moso (from the provinces of Yunnan and Sichuan in China), where the use of song in courtship rituals is an exclusive female prerogative (Namu & Mathieu, 2004). Finally, musical behaviors are found in a broad range of contexts, and serve a variety of purposes, which is inconsistent with the idea of functional specialization.

Experimental studies with humans (ii) aimed at testing the sexual selection hypothesis are currently scarce, and their results are inconsistent. In support of this hypothesis, some studies suggest that musical skill may function as a putative indicator of mental or sexual fitness. For example, both women and men have been reported to prefer sexual partners who demonstrate some music abilities (Tifferet et al. 2012), and women (but not men) give higher ratings of facial attractiveness and dating desirability after listening to music compared to a silent control condition (Marin et al., 2017). Moreover, women rate prospective long-term partners higher with respect to intelligence, health and social status, when these partners are associated with high-quality musical performance (Madison et al. 2018). Finally, musical discrimination skills correlate with general intelligence, a correlation that is accounted for by genetic heritability in males only (Mosing et al., 2015). In the same study, however, Mosing et al. (2015) found that musical skills do not predict measures of mating success, and that musical aptitude and mating success are not genetically correlated (Mosing et al., 2015).

Findings from *animal communication (iii)* studies are often invoked in theoretical proposals related to the hypothesis that musicality – in particular, musical rhythm – has evolved as a sexually selected trait. In animals, the main functional context in which rhythmic behaviors – primarily rhythmic sound production – have been investigated is that of mating displays. For many species, it seems that rhythmic displays have an exclusively reproductive function (Greenfield, 2006), as such displays seem to be strictly regulated by sex hormones (for a review, see Schlinger and Brenowitz, 2002), and are often a male-specific trait (Greenfield, 2006; Greenfield et al., 2021; ten Cate & Spierings, 2019; Party et al., 2014). It is important to note that scholars who invoke animal data as potential evidence to back up the theory of musicality as a sexually selected trait (or any other evolutionary theory of musicality, for that matter) do not claim that animal signals are musical. Instead, features of animal signals that are reminiscent of human musical behaviors (e.g., isochrony, communicative function) are studied as potential *precursors* and behavioral *analogues* meant to inform on the selective pressures that could be conducive to rhythm cognition and rhythmic communication, as it is ultimately found in human musicality. As such, the rhythmic behaviors of animals that will be discussed in this paper are minimally defined as signals exhibiting a rapid succession (typically in the sub-second timescale) of notes or movements that appear to exhibit isochrony. We discuss both solo rhythms and choruses, as evolutionary theories ought to consider building blocks and assess the potential continuity of relevant phenomena. Given the current status of empirical research, it is mostly unclear whether conspecific receivers perceive such signals as rhythmic, i.e. if they grasp the relational structure of isochrony or whether rhythmicity affects them at all. In this paper, we will not discuss irregular rhythms, nor interactive rhythms with offset-driven coordination (e.g. duetting), as this type of coordination is likely to depend on local cues rather than global patterns (e.g. Brown, 2007; Terleph et al., 2017), thereby being reactive rather than predictive.

Rhythmic signals have been studied most extensively in arthropod, anuran, and avian species, whose mating signals – also designated *songs* – have been traditionally likened to human music (Fitch, 2006; Ravignani et al., 2019a). In mammalian species, studies on the presence of rhythm in courtship displays have only recently begun to take off, with evidence coming primarily from marine mammals and bats. For example, throughout the breeding season, males of certain pinniped species and baleen whales (e.g., the bearded seal (*Erignathus barbatus*), harbor seal (*Phoca vitulina*), walrus (*Odobenus rosmarus*), leopard seal (*Hydrurga leptonyx*), blue whale (*Balaenoptera musculus*), humpback whale (*Megaptera novaeangliae*), minke whale (*Balaena rostrata*), fin whale (*Balaenoptera physalus*)), emit intense bouts of rhythmic calls underwater (e.g., Handel et al., 2009; Handel et al. 2012; Handel & Mercado, 2016; Rogers, 2017; Schneider & Mercado, 2019). Similar rhythmically structured calls that coincide with the breeding season have been described in the greater sac-winged bat (*Saccopteryx bilineata,* Burchardt et al., 2019; Knörnschild et al., 2017).

Several hypotheses have been advanced to account for the broad occurrence of rhythmic signaling in a reproductive context. It has been hypothesized that for species with a low-density distribution (e.g., certain marine and parrot species), call rhythmicity functions as a failsafe redundancy that preserves call distinctiveness and enhances its recognition by distant receivers (e.g., Heinsohn et al. 2017; Rogers, 2017). Conversely, in species with high population density, it has been hypothesized that rhythmic calling counters signal masking in noisy environments (e.g., Brumm & Slater, 2006; Serrano & Terhune, 2001). The predictability of rhythmic signals may also provide a mnemonic benefit (as, e.g., proposed for the rhythmic structure of whale song, Schneider & Mercado, 2019) by enhancing auditory processing. When performed as acoustic communal displays, synchronized rhythmicity may have a signal conservation function, by improving the quality of the broadcasted signal or the conspicuity of species-specific rhythms, thereby facilitating species recognition (Greenfield & Schul, 2008). Finally, such communal rhythmic displays may have a beacon effect, whereby the increased peak signal amplitude generated by multiple synchronized callers enhances the broadcasting range of the signal (Hartbauer et al., 2014; Merker, 2000; Merker et al, 2009; Morris et al., 1978).

A common assumption in theories of musicality as a sexually selected trait is that the specific *ways* in which rhythmic signals are executed function as honest fitness indicators, thus informing female receivers of male caller capacities. For example, in marine mammals that call underwater where breathing is not possible, the quality of rhythmic calling might advertise the caller's breath-holding capacity, which in turn is related to endurance and, thus, hunting ability. This hypothesis has been recently corroborated by an empirical study which revealed that, in leopard seals, rhythmic calling is more consistent in large-sized males compared to smaller ones (Rogers, 2017). As such, large-sized males are able to maintain long calling bouts and stable calling rates throughout the breeding season, while the calls of smaller males become shorter and irregular as the breeding season progresses.

Similar hypotheses have been advanced for the rhythmic vocalizing of humpback whales on breeding grounds. Since in this species male songs are updated annually with novel elements that spread across all males in a region (Eriksen et al., 2005), it has been hypothesized that the temporal consistency and structural complexity of a male's rendition of an updated song serve as fitness indicators for attracting females (Chu & Harcourt, 1986; Medrano et al., 1996; Tyack, 1981). An alternative hypothesis is that whale song serves as an inter-male competitive display for establishing a dominance hierarchy, based on age and status information provided by the features of individual song execution (Darling & Béruré, 2001). Finally, a third hypothesis that relates whale song to reproductive function is that whale song functions as a long-range sonar, allowing males to locate females (Frazer & Mercado, 2000), or to locate other individuals, both for mating and other purposes, such as traveling or foraging (Mercado, 2018).

In a number of orthopterans and anurans, rhythmic courtship takes the form of communal displays, often involving a large number of individuals that signal either synchronously or in alternation. The seemingly cooperative appearance of such displays, while occurring in an obviously competitive context, has intrigued scholars for a long time, compelling them to draw parallels to rhythmic synchronization in humans, and thus to human musicality (for discussions, see Ravignani et al., 2014; Wilson & Cook, 2016, and references therein). It has to be stressed, however, that there is a key difference between such animal signals and musical rhythm, in that the former are simple and innate, while the latter can take complex forms that are learned.

Data from playback experiments and rigorous acoustic analyses reveal that chorusing synchrony in insects is not a unitary phenomenon. Even in closely related species, the emergence of synchrony can be achieved through distinct mechanisms that have been shaped by distinct selective pressures (Greenfield & Schul, 2008). In species where the females show a strong orienting response towards the first call emitted in a chorusing bout, synchronous calling emerges incidentally, as the by-product of males' selective attention and a competitive drive to produce the leading signal (for recent reviews, see Greenfield et al., 2017; Greenfield et al., 2021; Hartbauer & Römer, 2016). Accordingly, males time their calls to avoid overlapping with other males, but only monitor the calls of nearest neighbors, which results in call alternation between neighboring males and accidental call synchronization between distant males. Interestingly, at least in some species, communal calling entails greater rhythmic regularity compared to solo calls, although the rhythmic structure attained by the chorus is irrelevant to females (e.g., Greenfield & Schul, 2008).

In species where female response is dependent upon the perception of distinct sound envelopes, a more precise form of synchrony tends to emerge as a ‘cooperative’ process, whereby males adjust their intrinsic calling to align to the rate and phase of neighboring males (e.g., Greenfield et al., 2017; Greenfield et al., 2021). This type of synchrony presents multiple adaptive advantages, including the preservation of species-species acoustic signatures (consistent with the signal conservation hypothesis), and increased broadcasting reach (consistent with the beacon effect hypothesis).

Similar adaptive benefits have been linked to the evolution of musical rhythm by Merker and colleagues (Merker, 2000; Merker et al., 2009). Accordingly, selection pressures related to male communal territoriality and female exogamy – which likely characterized the last common ancestor of humans and chimpanzee – led to the emergence of synchronized rhythmic displays in a human ancestor during the late Miocene, as a way of attracting distant females. Arguably, synchronized stomping and vocalizing would have provided a beacon benefit in this context, allowing the signal to reach farther than single or unsynchronized signals. The potential similarity of adaptive benefits between insect and hominin synchronized displays should not be interpreted as entailing that the two are identical with respect to forms, synchronization mechanisms or, indeed, the putative selection pressures that shaped them. As it will be reviewed in the remainder of section 2 and in section 3, additional adaptive benefits, selection pressures and supporting mechanisms may have been involved in the evolution of human audiomotor entrainment and inter-individual synchronization.

2.2. Music as the emotional sibling of speech

Suggestions that music – in particular, song – and speech have common evolutionary origins date far back in the history of philosophical thinking, although these accounts do not focus specifically on rhythm. Originally, the debate revolved around whether music or speech had precedence over the other (e.g., Darwin, 1871; Rousseau, 1761; Spencer, 1857). An alternative account, that can be traced back to Darwin (1871) and has been significantly developed by Brown (2000b; 2017) postulates that a rudimentary communication system eventually diverged into a referential communicative system (speech) and an emotional communicative system, i.e., music (for additional perspectives on this account, see also Mithen, 2005).

Most arguments for the common origins of musicality and speech are theoretical attempts at bridging data from speech and music research in order to find similarities between the two. Such comparisons suggest that speech and singing engage overlapping neural networks (e.g., Musso et al., 2015; Özdemir et al., 2006), rely on similar mechanisms of emotional modulation (Brown, 2017) and emotional induction (Ma & Thompson, 2015), and attractiveness ratings for speaking and singing correlate within the same individual (Valentova, 2019). These similarities also encompass aspects of rhythm processing and production. For example, the temporal organization of both language and music exploits perceptual grouping, generativity and recursion whereby elements are mentally organized into hierarchically structured sets, whose boundaries are marked by changes in duration or frequency (e.g., Fitch, 2006). As critics have pointed out, however, the similarities between speech and music may be the result of confounding variables, such as common neuromotor systems connected to auditory input and vocal output (Zatorre & Baum, 2012). This is highly plausible considering that ‘common origins’ accounts focus primarily on song (i.e., *vocal* musical behavior), overlooking other musical behaviors such as dance or tool-assisted sound production.

Beyond similarities related to common input and output systems, the production of song and speech diverges with respect to breathing patterns (Leanderson et al., 1987), the use of the vocal apparatus (Sundberg 2018) and underlying processing mechanisms (Zatorre & Baum, 2012). Generally, singing is more demanding than speech, as it requires finer vocal control (Zarate, 2013), higher capacities to regulate breathing and subglottal pressure (Sundberg, 2018), as well as higher energy expenditure, due to engaging all respiratory muscles (intercostal, abdominal and diaphragm), as opposed to speech, which typically implicates only the intercostal muscles (Leanderson et al., 1987; Åkerlund & Gramming,1994). Clinical evidence further suggests divergences with respect to the neuro-cognitive processing of speech and music, as speech processing is intact in patients with amusia (i.e., tone-deafness), while music processing is not impaired in aphasia patients (Mithen, 2005; Peretz & Coltheart, 2003). Finally, and of most relevance to the present paper, rhythm and speech differ with respect to rhythmic structure. Unlike speech, musical rhythm is dominated by isochrony, i.e., evenly spaced time intervals (Kotz et al., 2018; Ravignani & Madison, 2017; Savage et al., 2015). Finally, interactive musical behaviors build on *predictive coordination* of event *onsets* either between individuals or between an individual and another exogenous rhythm source. In contrast, speech is primarily a system of *alternation*, which relies on the between-individual coordination of event *offsets* (Brown, 2007).

These differences notwithstanding, a recent iterated learning study (Ma et al., 2019) found that a single communication system could diverge, across several generations, under the exclusive pressure of communicative function. As such, when nonsense vocalizations were presented as having the function to communicate emotional states, they evolved into vocalizations that were rated as more music-like by naive participants from two different cultures. These vocalizations exhibited wider pitch and intensity variation, slower rates and longer durations compared to average vocalizations. Conversely, nonsense vocalizations whose function was described as referential, were rated as being more speech-like. These findings are consistent with the ‘common origins’ theories reviewed above, as they suggest that demands for functional specialization may trigger a cascade of phonatory and acoustic changes ultimately driving a communication system to split into distinct systems. Since animal vocalizations function as behavioral expressions of emotional states (Bachorowski & Owren, 2003), this putative functional specialization could have been prompted by increased demands for referential communication in ancestral human species. It is also possible, however, that such demands have more protracted origins, which extend to the last common ancestor of hominids, given evidence of referential vocal communication in the other great ape species (e.g., Crockford et al., 2014; O’Bryan et al., 2018).

2.3. Musicality as a biocultural adaptation in the service of group-level functions

It has long been suggested that musicality has emerged under selection pressures related to human sociality, to support group functions with regard to within- and/or between-group cooperation or inter-group competitiveness (e.g., Kogan, 1997; Roederer, 1984). Consistent with these theories, cross-cultural evidence suggests that the rhythmic component of musicality may be inherently social, considering that, across cultures of the world, there is a universal association between isochronous rhythm and musical *group* performance (Savage et al., 2015). While social theories on the evolution of musicality agree that rhythmic isochrony is the crucial feature that facilitates the social functions of musicality, they diverge with respect to the putative primordial function(s) ascribed to proto-musical behaviors. Some theories are relatively unspecific, in that musicality is assigned a generalized group-level function by reinforcing social life (Brown, 2000a; Koelsch, 2014; Loersch & Arbuckle, 2013; Merker et al., 2009; Roederer, 1984). Other theories emphasize a particular adaptive function as primordial for ancestral musical-like behaviors, with the most prominent among these being **social bonding** (Dunbar 2004; Dunbar 2012; Huron 2001; Savage et al., 2021), **coalition signaling** (Bryant, 2013; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al., 2020), and **cooperation** (Brown 2000a; Freeman, 2001; Kogan, 1997; Merker et al., 2009). Since social theories rely on arguments pertaining to human-specific features of sociality, most evidence in their support comes from human studies. Animal behaviors are, nevertheless, referenced by such theories insofar as they are related to adaptive pressures presumed to have spurred the emergence of protomusical behaviors.

According to the **social bonding** theory (e.g., Dunbar, 2004; Dunbar 2012; Huron, 2001), musicality has evolved as a prelinguistic mechanism that supplanted social grooming – the primary bonding mechanism of primates – under pressures exerted by increased group-size in early *Homo* species. Arguably, dyadic bonding mechanisms entailed prohibitive time resources in large groups (Lehman et al., 2007), being thus replaced by novel ones such as laughter, proto-musical chorusing and speech (Dunbar, 2004). This putative shift from tacile to vocal social bonding would have exerted a pressure on the development and refinement of the neuroanatomical mechanisms involved in vocal bonding behaviors, such as the motor control of vocal output and bodily time keeping (Dunbar 2004; Dunbar, 2012). Early arguments in favor of this theory highlighted, for example, the potential of music to simultaneously reach, involve, and thus connect a large number of individuals (Huron, 2001). Moreover, the ancestral neurochemical mechanisms underlying primate bonding through grooming appear to also be activated by rhythmic behaviors (as detailed in 3.4).

Other theories highlight musicality – and in particular rhythmic synchronization – as abilities that enable individuals to coordinate in time, thereby supporting human-specific forms of **social cooperation**, besides the affiliative effects mentioned above (Brown, 2000a; Freeman, 2000; Kirschner & Tomasello, 2009; Merker et al., 2009;). Kogan (1997), as well as Merker and colleagues (2009) argued, for example, that the feelings of bonding engendered by musical behaviors contributed to group solidarity, and promoted altruism, thereby increasing the effectiveness of collective actions (e.g. predator defense, warfare).

The experimental evidence in support of the **social bonding** theory and the **social cooperation** theory is abundant, as musical behaviors, such as communal singing, chanting, dancing and drumming, have a wide range of prosocial effects (Anshel & Kipper, 1988; Pearce et al., 2017; Reddish et al., 2013; Reddish et al 2014; Tarr et al., 2016; Wiltermuth & Heath, 2009) that fail to materialize when only passively listening to music (e.g., Dunbar et al., 2012). Moreover, these effects appear to be specifically triggered by the rhythmic components of music, as the prosocial effects of rhythmic synchronization extend beyond the context of music-making. Indeed, interpersonal rhythmic synchronization (e.g., rocking in chairs, finger tapping) leads to prosocial feelings, such as liking, rapport, trust, affiliation, entitativity (i.e., ‘feeling as one’), cooperation and generosity (Hove & Risen, 2009; Valdesolo et al., 2010; Wiltermuth & Heath, 2009).

The prosocial effects of rhythmic synchronization emerge early in human development, which is also suggestive of ancestral evolutionary roots (e.g., Justus & Hutsler, 2005; McDermott & Hauser, 2005; Trainor 2018). For example, 12-month old infants prefer a person who rocks in synchrony with them as opposed to an asynchronous one, and this preference does not extend to synchrony exhibited by non-social entities (Tunçgenç et al., 2015). Fourteen-month old infants are more helpful after being engaged in interpersonal synchrony (Cirelli et al., 2014), an effect which is present also in the absence of music (Cirelli et al., 2017). In preschoolers, the prosocial effects of interpersonal rhythmic synchronization are documented across a variety of behaviors, such as swinging, music-making, clapping and tapping (Kirschner & Tomasello, 2010; Rabinowich & Meltzoff, 2017a, b; Tunçgenç & Cohen, 2016). In conclusion, from an early stage of human development, rhythmic synchronization triggers a wide array of prosocial effects in both musical and non-musical contexts, thereby suggesting a social bonding function for the rhythmic component of musicality.

In the **coalition signaling** theory on the evolution of musicality, the causal relationship between interpersonal rhythmic synchronization and social bonding is reversed, as compared to the **social bonding** theory. Rather than *facilitating* social bonding, proto-musical behaviors purportedly *advertised* the level of bonding within a group, with bonding being acquired by other means (Bryant, 2013; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al. 2021; Merker, 2000). Accordingly, a putative benefit of synchronous rhythmic displays would have been to confuse and frighten the enemy through the mimicry of a large animal (Merker, 2000). According to an alternative proposal, synchronous displays co-evolved with the complexity of social organization as a way of advertising group fitness in the formation of inter-group alliances

(Hagen & Bryant, 2003). Consequently, it has been hypothesized that complex and well-synchronized group displays signal group stability and higher coordination abilities, since a high degree of synchronization entails long-term commitment and dedication.

Compared to the **social bonding** theory, the empirical evidence supporting the **coalition signaling** theory is more limited, and consists of findings suggesting that tight synchronization may communicate high group quality . For example, in one study, participants rated highly synchronized musical performances as being of better quality, and rated better performances as suggestive of higher coalition quality (Hagen & Byrant, 2003). Similar correlations have been found beyond the context of music, as the formidability (size and muscularity) of a group was judged to be higher for audio-tracks of highly synchronized footsteps compared to audio-tracks of less synchronized footsteps (Fessler & Holbrook, 2016). Several experiments have also revealed that entitativity is judged to be higher when individuals move in synchrony (Edelman & Harring, 2014; Lakens, 2010; Lakens & Stel, 2011).

According to the **coalition signaling** theory, the crucial socio-ecological pressures that triggered the emergence of proto-musical behaviors are linked to two ancestral traits of *Homo* species during the Middle Pleistocene: primate territoriality and social carnivory. Specifically, the coalition signaling function of musicality would be rooted in the territorial displays of ancestral humans (Hagen & Hammerstein, 2009). The communal territorial displays of present day social carnivores (e.g., lions, wolves) and nonhuman primates (e.g., chimpanzees) are postulated as *analogous* and *homologous* precursors to the putative territorial advertisement of ancestral humans, based on socio-ecological convergence and shared ancestry (e.g., Hagen & Hammerstein, 2009). On the one hand, during the Middle Pleistocene, our ancestors occupied a similar niche to that of large social carnivores. On the other hand, territorial group displaying is likely an ancestral primate trait, which in monogamous primate species may take on the form of duetting based on song-like vocalizations (e.g., Hagen & Hammerstein, 2009).

2.4. Mother-infant interaction as the primordial context of proto-musical behaviors

Theories of mother-infant interaction as the primordial context of musicality focus primarily on vocal behavior, and emphasize two putative adaptive functions that may explain why hominin vocalizations acquired music-like features: **mother-infant bonding** (Dissanayake, 2000, 2009a; Mithen, 2005) and **credible signaling of parental attention** (Mehr & Krasnow, 2017; Mehr et al., 2020).

According to the **mother-infant bonding** theory (e.g., Dissanayake 2000, 2009a), music-like vocalizing stems from affiliative signals that in mother-infant interaction were modified through a process of ritualization. Borrowed from the field of animal acoustic communication, the notion of ritualization refers to a process by which instrumental behaviors evolve into communicative signals, through gradual changes toward increased formalization, repetition, and exaggeration (Grammer & Eibl-Eibesfeldt, 1990; Watanabe & Smuts, 1999). Arguably, adult affiliative signals (e.g., head bobbing, body leaning toward, smiling, eye-brow flash, high-pitch voice modulations, etc.) were subjected to similar changes in the context of mother-infant interaction, thereby evolving into repetitive and exaggerated signals that were emotionally-evocative to infants. Given their emotional efficacy, music-like interactions would have then spread from the primordial context of ritualized parent-infant communication to ceremonial contexts, where they were further elaborated through ritualization (Dissanayake, 2000, 2009a).

The **credible signaling of parental attention** theory proposes that infant-directed song, which is universally present across human cultures (Mehr et al., 2019; Trehub, 1993), is rooted in increasingly elaborated contact calls that, in ancestral hominin species, had evolved to encode parental cues of attention to the offspring (Mehr & Krasnow, 2017; Mehr et al., 2021). Two putative selection pressures would have led to this development. On the one hand, it is argued that multiple immature offspring simultaneously competed for parental attention, given the longer postnatal period of brain development and the shorter interbirth intervals of humans compared to the other apes. On the other hand, bipedalism and loss of body hair prevented infants from clinging onto the parent's body (as the infants of other ape species do), thereby preventing constant mother-infant proximity. These shifts in parent-infant ecology purportedly prompted early Homo species to evolve song-like vocalizations that enabled foraging caregivers to continuously convey parental attention to their multiple offspring from some distance, without incurring foraging costs (Mehr & Krasnow, 2017).

As recently pointed out by Trehub (2021), the **credible signaling of parental attention** theory is challenged by childcare features among contemporary foraging cultures, where prolonged breastfeeding entails prolonged interbirth intervals and infants are continuously carried by their mothers. Moreover, cross-cultural data indicate that the most frequent strategies employed by mothers for soothing distressed children are picking up, rocking and talking to the infant rather than song (Bornstein et al., 2017; Richter & Ostovar, 2016).

Another potential challenge is that the evolutionary scenario presented by the **credible signaling of parental attention** theory is very similar to a scenario of speech evolution presented more than a decade earlier by Falk (2004). Specifically, Falk (2004) theorized that speech emerged from infant-directed signals used by foraging parents to communicate with physically distant infants. Just like Mehr & Krasnow (2017), Falk (2004) emphasized evolutionary pressures related to increased demands for parental attention, and identical adaptive solutions to deal with such demands, i.e., the elaboration of ancestral contact calls. The two theories, however, diverge with respect to the outcome of this evolutionary process: infant-directed speech versus infant-directed song. As current research shows, these two parent-infant communication systems are acoustically distinct, and differ with respect to many pitch, timbre, and rhythmic features (Moser et al., 2020). It is currently unclear how the two theories can be reconciled, since neither of them discusses when and what prompted the emergence of the specific features that differentiate infant-directed song from infant-directed speech. One potential route to reconciliation is through the 'musilanguage' theory reviewed in 2.2. Accordingly, infant-directed song and infant-directed speech could be viewed as resulting from an originally undifferentiated system of contact calls. This would, however, require that each of the two theories is complemented with additional assumptions concerning a putative differentiation stage and the adaptive demands that prompted it. In line with the theories reviewed in 2.2, other possibilities would be that one of the systems evolved from the other, or independently from one another but at different points in time. Both phenomena would then have survived because both have added to the survival of the species, or at least were connected to something that has.

Before concluding this section, it is important to note that one of the two proposals reviewed here – the **credible signaling of parental attention** theory – is primarily concerned with the evolutionary emergence of musical pitch and melodic modulation. The **mother-infant bonding** theory implicitly incorporates the evolution of both rhythm and melody, by emphasizing the emergence of rhythmicity and exaggerated pitch contours through the ritualization of pre-existing affiliative signals. The explanatory power of evolutionary proposals focused on mother-infant interaction as the putative primordial context of musicality might benefit from extending their focus beyond vocal behavior, to also encompass rhythmic moving, as reviewed above.

2.5. Therapeutic and ritual uses of musicality and rhythmicity

Self-report studies show that music is one of the most common affect-regulation strategies that humans use in everyday life to boost energy levels, or reduce tension (Saarikallio, 2011; Saarikallio & Erkkilä, 2007; Thayer et al. 1994). Music is also used in clinical settings, to alleviate stress-related disorders (Aalbers et al., 2017; Koelsch, 2009; Landis-Shack, 2017; Sihvonen et al., 2017; Thaut et al. 2015). Many studies show that the beneficial psychological effects of music are reflected in physiological measures, including heart rate, blood pressure and hormonal levels (for a recent review and meta-analysis, see de Witte et al., 2020). Similar positive effects on stress levels, assessed with both physiological and behavioral measures, have been documented in multiple nonhuman species, including apes, monkeys, dogs, hens and rats (for a review, see Alworth & Buerkle, 2013), although the presence and type of such effects vary as a function of species and type of musical stimuli.

It is currently unclear, however, which musical features – and thus mechanisms – underlie the clinical benefits of musicality, since relevant variables – such as tempo, rhythm, and melodic contours – are typically not dissociated in the design of intervention studies, or are inconsistently controlled in clinical and research practices (Aalbers et al., 2017; Leubner & Hinterberger, 2017; McPherson et al., 2019). Recent evidence, nonetheless, suggests that active interventions that implement rhythm-based activities (hand clapping, instrument playing, dancing) may have greater potential for improving physiological stress markers (McPherson et al., 2019). Additional findings show that greater stress relief is provided when experiencing live music, compared to prerecorded music (for a review, see Witte et al., 2020).

The use of music as a coping strategy has been attributed to the sense of pleasure that arises when humans synchronize to external rhythms (Dunbar, 2012; Koelsch, 2014; Launay et al., 2016; Salimpoor et al., 2015; Thaut et al., 2015; Trost et al., 2017; Vuust & Kringelbach, 2010), which in turn may be linked to anxiolytic and elating effects mediated by the release and circulation of dopamine, oxytocin, and endorphins (as detailed in section 3.4). Purportedly, these acted as reinforcers for the preservation and elaboration of rhythmic synchronization into rituals and, ultimately, into that which today is called music and dance (Brown, 2000a; Dissanayake, 2006, 2009b; Mithen, 2005). Plausibly, this reinforcement mechanism has protracted evolutionary origins, given the presence of entrainment in a type of chimpanzee behavior – the so-called ‘Conga line’ (Lameira et al., 2019). This behavior is a form of locomotion, whereby two or more individuals align in a row, either standing or sitting, moving forward with a swaying movement, while maintaining bodily contact. Lameira et al. (2019) described the Conga-line as a stereotypical behavior, and speculated that rhythmic synchronization may have emerged in the *Homo* lineage as a coping co-stereotypy under the pressure of increasing socio-ecological stressors, such as resource scarcity (due to climate change) and crowding (as a consequence of increased group size and cooperative breeding). In support of this theory, the authors also mention that chimpanzee rocking is commonly regarded as a kind of stereotypical behavior related to confinement in captivity (e.g., Chamove, 1989; Lopresti-Goodman et al., 2013). Conga-line variants, however, may occur in zoo-housed groups in contexts (e.g. social play) where they cannot be explained as stereotypical behaviors (personal observation). Moreover, the Conga-line behavior (called ‘snake dance’) has been observed in gorillas in wild populations (Schaller, 1963).

Other putative selective pressures that would have facilitated the recurrent use – and thus preservation – of musical behaviors as a coping mechanism may derive from enhanced cognitive abilities in the *Homo* lineage, in particular enhanced memory, foresight and causal reasoning abilities. Plausibly, this entailed an increased awareness of the uncertainty surrounding the outcome of future events, activities or natural phenomena, thereby leading to increased stress and anxiety. It has been proposed that apprehension about the future and the need to control uncertainty likely co-evolved with musicality – and the other artistic skills – as the motivators and the expression of what we currently define as religious practices (Brandt, 2009; Dissanayake, 2009a,b). In other words, proto-musical behaviors – which took the form of ritualized versions of ordinary, instrumental behaviors – came to be connected to situations that had a vital significance to the group, as practices for curbing future uncertainty and mitigating collective crises. In such contexts, joining with others in rhythmic behaviors had purported coping benefits, by lowering stress and anxiety, and instilling a sense of uncertainty-control in the participants (Brandt, 2009; Dissanayake, 2009b; Mithen, 2005).

Consistent with the view that musicality – and in particular musical rhythm – has evolved to cope with future uncertainty, musical stimuli engage several components of the so-called 'default mode neural network', which is commonly associated with remembering the past and simulating the future (Herdener et al., 2010; Toiviainen et al., 2020). In addition, musical stimuli improve memory functions and induce plasticity by increasing gray matter volume in a key structure of this network – the hippocampus (Herdener et al., 2010).

Strong support in favor of this view comes also from ethnographic records showing that musical behaviors are often linked to rituals that are associated with the supernatural, and have the function of invoking forces believed to affect vital aspects of group functioning, from hunting success and fertility, to protection from disease or warfare victory (Arom, 2000; Cross, 2003; Dissanayake, 2009b; Brandt, 2009; Nettl, 2015). As mentioned above, the widespread presence of synchronized displays in such ritualic practices has been attributed to the putative anxiolytic effects of rhythmic behaviors and, thus, to the sense of control that these may have evoked in situations in which ancestral humans had to deal with outcome uncertainty prior to crucial undertakings (Dissanayake, 2009b; Mithen, 2005). There are several examples of ritual practices that fit this scenario, such as the collective chanting and swaying of tribes in Papua New Guinea during violent storms (Malinowski, 1922; Mead, 1930, apud Dissanayake 2009a), and the widespread use of communal musical behaviors in healing ceremonies and for lamenting the passing of group members (Dissanayake, 2009b; Mithen, 2005). Whether the anxiolytic use of collectively enacted rhythms constituted a primordial function in the *evolution* of musicality is, however, far from clear. In contemporary traditional societies, the ritual use of musical behaviors extends to a variety of rituals that are less clearly connected to a need for stress relief. As Seeger (1987:7) wrote about the lowland regions of South America: “wherever music is heard, something important is happening”. In perhaps all cultures of the world, musical behaviors are used – and conceptualized – as an integral part of important ceremonies, and as means to achieve important social purposes.

3. Proximate mechanisms

Many mechanistic hypotheses on the evolution of musicality focus on audiomotor entrainment, i.e., on the accurate synchronization of bodily movements to an auditory rhythmic stimulus. This is perhaps understandable, since audiomotor entrainment is regarded as the most iconic expression of human musicality, and has long been considered absent in other species. There is general agreement that entrainment relies on an ability to extract the temporal periodicity of an acoustic signal, and to predict the subsequent occurrence of its units, which in turn enables individuals to synchronize their rhythmic movements with the acoustic signal. Mechanistic hypotheses on the evolution of other aspects of human rhythm cognition are generally lacking (as will be detailed in section 4), with the exception of recent attempts to extend the scope of entrainment hypotheses to relevant perceptual abilities, such as isochrony detection (as reviewed in 3.1 and 3.2).

3.1 Neural adaptations underpinning rhythmicity

The most prolific hypothesis advanced within the recent wave of comparative research on rhythmicity is the **vocal learning and rhythmic synchronization** hypothesis (in short, VLRS: Patel, 2006), with its subsequently revised versions – the **intrinsic reward and rhythmic synchronization** hypothesis (Takeya et al., 2017) and the **vocal learning as a preadaptation for human beat perception and synchronization** (Patel 2021; Rouse et al. 2021). In its original formulation, VLRS held that audiomotor entrainment is a by-product of neural adaptations evolved to support vocal learning (in short, VL), in particular tight audio-motor neural couplings in the forebrain circuitry. The VLRS thus predicted that audiomotor entrainment should be present only in *flexible* VL species, such as songbirds, parrots, hummingbirds, seals, whales, bats and elephants (e.g. Petkov & Jarvis, 2012).

Initially, the comparative data from experimental studies appeared consistent with this divide, as eight budgerigars (*Melopsittacus undulatus*) could learn to synchronize motor responses with a metronomic stimulus (Hasegawa et al., 2011), and two parrots could synchronize head bobs with complex musical stimuli at various tempi (Patel et al., 2009; Schachner et al., 2009). Moreover, a YouTube sampling study found that only individuals of VL species (14 parrot species, one elephant species) exhibited audiomotor entrainment, while individuals of non-VL species (including great apes) did not (Schachner et al., 2009). Initial experimental studies with rhesus macaques (*Macaca mulatta*) further suggested that nonhuman primates may be unable to synchronize motor output with an acoustic signal (Merchant & Honing, 2014).

Subsequent findings from experimental studies with species traditionally regarded as vocal non-learners appeared to contradict VLRS. Notably, a sea lion (*Zalophus californianus,* Cook et al., 2013), and several rhesus macaques (Gámez et al., 2018; Takeya et al., 2017) could be trained to synchronize a motor response with an exogenous auditory rhythm. To account for such findings, a revised version of VLRS – the **intrinsic reward and rhythmic synchronization** hypothesis (Takeya et al., 2017) – stated that non-VL species may be able to acquire audiomotor entrainment, but only *flexible* VL species are motivated to display it *spontaneously*. Yet, evidence of spontaneous audiomotor entrainment in chimpanzees (*Pan troglodytes*: Hattori et al., 2013; *Pan paniscus*: Large & Gray, 2015), who are not considered flexible vocal learners, may challenge this updated version of the VLRS. Evaluating the great ape data in relation to VLRS, however, is not entirely straightforward. On the one hand, there is increasing evidence that non-human great apes exhibit VL, both in the form of socially-driven adaptations of pre-existing vocalizations and in the more advanced form of acquiring entirely new vocalizations through social learning (as reviewed by Lameira et al., 2017). On the other hand, as Patel has argued (2021), the chimpanzee in Hattori et al. (2013), was only exposed to metronomic stimuli (as opposed to more complex beat-driven rhythm). Moreover, the chimpanzee did not generalize spontaneous entrainment to tempi that differed from her own spontaneous motor output. Thus, evidence of audiomotor entrainment in the common chimpanzee is limited to the most rudimentary form of audiomotor synchronization (following the conceptual model of Iversen & Balasubramaniam, 2016). Finally, visual feedback (i.e., the drumming behavior of the experimenter) may have aided drumming synchronization by the bonobo in Large & Gray (2015). Against this latter objection, however, a recent study has found that chimpanzees display similar levels of rhythmic coordination with a conspecific, regardless of whether the exogenous rhythm (i.e., the tapping rhythm of the conspecific) is provided in the auditory channel only or in both the auditory and visual channels (Yu & Tomonaga, 2018).

The most recent version of VLRS states that **VL is a necessary preadaptation for *beat* perception and synchronization** (Patel 2021). In this reformulation, it is clarified that audiomotor entrainment to metronomic stimuli is not within the scope of the VLRS. In contrast to such rudimentary forms of synchronization, perceiving and synchronizing with a *beat* entails the ability to extract the regular pulse in a temporally and/or acoustically varying rhythm (Honing, 2012). Accordingly, this newly updated VLRS predicts that, across species, degree of audiomotor connectivity – and thus VL ability – correlates with differences in rhythm perception and/or production. Specifically, and drawing on the VL taxonomy of Petkov & Jarvis (2012), only high VL species (i.e., humans and parrots) will exhibit spontaneous beat perception and synchronization. Rhythm perception and production in other species are predicted to co-vary with their VL capabilities, and thus to decrease from complex vocal learners (bats and songbirds) to moderate vocal learners (non-human apes), to limited vocal learners (mice), to vocal non-learners, such as chickens or lizards.

The VLRS has driven the comparative study of rhythm cognition more than any other theoretical proposal in the field. More research is, however, needed to consolidate its empirical basis. Research is, for example, needed on the rhythm perception abilities of animals, which should be examined, comparatively, across categories of VL species, using standardized methods. Likewise, the capability and readiness of animals to acquire rhythmic synchronization needs to be assessed across the VL continuum. Finally, the spontaneous predilection of parrots to synchronize with the beat of complex musical stimuli, as well as its ontogenetic emergence, needs to be systematically investigated. Interestingly, parrots exhibit unique patterns of brain organization compared to other avian VL species, as vocal motor pathways that are anatomically adjacent to the auditory pathways in songbirds, are more distant in parrots (Jarvis & Mello, 2000). Moreover, parrots possess an additional vocal system that also includes non-vocal motor pathways (Chakraborty et al., 2015), which are active during e.g., hopping and head bobbing movements (Feenders et al., 2008). Based on these findings and given evidence of spontaneous audiomotor entrainment in parrots, it is possible that, not only VL, but a predisposition for rhythmic gestural/bodily communication (especially if exhibiting a certain level of flexibility and voluntary control) represents a prerequisite trait for spontaneous rhythmic synchronization to emerge. In turn, this could explain why chimpanzees, which are less accomplished vocal learners than parrots but exhibit rhythmic gestures (as detailed in 4.3.), also exhibit spontaneous – albeit rudimentary – audiomotor entrainment.

An alternative hypothesis, which, like VLRS, focuses on neural audiomotor adaptations (see Table 2 for an overview of hypotheses on neural adaptations that support rhythm production and perception), is **the gradual audiomotor evolution** hypothesis (in short, GAE). According to GAE, a progressive chain of anatomical and functional brain changes has led to the evolution of a human-specific beat-based timing (or relative timing) mechanism, in addition to the interval-based timing mechanism present in, e.g., monkeys (Merchant & Honing, 2014). As such, in the primate lineage, audiomotor entrainment is predicted to co-vary with the complexity of auditory and motor circuits and the robustness of their coupling. More specifically, monkeys are predicted to lack an ability for audiomotor entrainment, and to exhibit very few audiomotor neural connections. In contrast, nonhuman apes are predicted to exhibit higher complexity of relevant audiomotor circuitry, as well as rhythmic abilities that are superior to those of monkeys, but inferior to those of humans. Initial support for GAE came from evidence of spontaneous – yet inflexible – audiomotor entrainment in chimpanzees (Hattori et al., 2013; Large & Gray, 2015 reviewed above), as well as evidence that macaque monkeys failed to acquire entrainment in spite of extensive training (Merchant & Honing, 2014).

Insert Table 2 about here

The GAE is to some extent challenged by recent data showing that macaque monkeys can be trained to synchronize a motor response to an isochronous metronome pulse (Gámez et al., 2018; Takeya et al., 2017). Moreover, they generalize their response to non-trained tempi (Takeya et al., 2018), adapt the tempo of their motor response to accelerating and decelerating metronomes (Gámez et al., 2018), with tempo adaptation being driven by the adjustment of the inter-onset interval (Donnet et al., 2014). Taken together, these results show that, when trained and tested with more suitable methods, monkeys can exhibit more flexible audiomotor synchronization than initially predicted by GAE.

Unlike the VLRS and GAE, other theories argue that entrainment is widespread in the animal kingdom. Cook et al. (2013), for example, propose that entrainment is enabled by **neural resonance** mechanisms that are largely conserved across taxa, and result from oscillatory interactions between neuronal assemblies in sensory and motor areas (e.g., Large & Gray, 2015;

Rouse et al., 2016). Specifically, in the case of audiomotor entrainment, oscillations in auditory neurons phase-lock on the frequency of an exogenous auditory stimulus. In turn, this causes neural oscillations in the motor system to phase-lock onto the same frequency, thereby leading to motor output that is entrained to the exogenous auditory input. Consistent with this proposal, evidence from studies with rats and macaques suggests that, similarly to what has been observed in humans (e.g., Rouse et al., 2016), exogenous rhythmic stimuli induce entrainment in nonhuman brains as well (Bartolo et al., 2014; Noda et al., 2017).

Besides oscillatory interactions (such as those described above), additional neural mechanisms, such as **dynamic attending** (e.g., Large & Jones, 1999) and **predictive coding** (e.g., Vuust et al., 2009) have been proposed to contribute to the emergence of rhythm percepts. The involvement of **dynamic attending** in rhythm processing is linked to the fact that, due to their regularity, rhythmic stimuli induce strong expectations about upcoming events. This causes attention to fluctuate in an oscillatory way, with most attentional resources being cyclically allocated to the most salient or most informative events in a sequence. Through dynamic attending, the deployment of highly energetic cognitive capacities is thus optimized to capture the most relevant stimuli. The process described above also entails that rhythmic stimuli are **coded predictively**, i.e., predictions are generated about event recurrence at regular points in time, which in turns enables optimal allocation of attention, as well as audiomotor entrainment (Vuust et al., 2009). Behaviorally, predictive coding is captured by the fact that bodily movements occur near the onset times of the auditory stimulus, with a slight anticipation of it (for reviews, see e.g., Fitch, 2013; Merchant & Honing, 2014; Merker et al., 2009; Repp & Su, 2013). Since **dynamic attending** and **predictive coding** provide broad adaptive advantages, by optimizing stimulus reception and minimizing processing load (Large & Jones, 1999), and are crucially involved in rhythm perception and production, it is plausible that rhythmic signals are widespread in animal communication, as such signals would be easier to decode and recall, while also allowing for temporal pattern variations and, thus, species specificity. This is corroborated by evidence of shared neural mechanisms of dynamic attending in humans and nonhuman primates. In particular, when macaque monkeys are presented with rhythmic stimuli, neural oscillations in the relevant sensory modality entrains to the stimuli, with momentary sensory response gain (i.e., increased attention) at expected points in time that reflect the attended rhythm (Lakatos et al., 2008).

Another proposal that argues for the widespread distribution of entrainment in the animal kingdom, and implicates neuroanatomical adaptations, is **the acoustic advantages** hypothesis (Larsson & Abbott, 2018). This hypothesis extends an earlier version, which proposed that human entrainment is a by-product of **incidental sounds of bipedal locomotion** (Larsson, 2013; Larsson et al., 2019). Since bipedal locomotion entails periodic symmetric gaits, it purportedly enabled our ancestors to match each other's steps, and, thus, to reduce footfall noise, which resulted in stealthier locomotion and better detection of relevant environmental sounds. The **acoustic advantages hypothesis** extends this reasoning to vertebrates in general, proposing that the audiomotor circuits and genes involved in entrainment have their origins in fish schooling behavior and locomotor-respiratory coupling. In tetrapod descendants, this then evolved into couplings between forepaw motor processing and vocal communication, which favored the synchronization of movement and vocalization (Larsson & Abbott, 2018). Synchronized behaviors, such as schooling guided by incidental sounds of locomotion, putatively provided the adaptive advantage of improved situational awareness due to extended windows of silence. To explain evidence of entrainment in vocal non-learners, the **acoustic advantage hypothesis** contemplates the relationship between entrainment and VL in reverse, i.e., the ability to entrain to external sounds is proposed to be a prerequisite for VL and not the other way around (Larsson & Abbott, 2018).

A related hypothesis is the **bipedal experience in utero** hypothesis (Larsson et al., 2019; Parncutt & Chuckrow, 2017), whereby maternal walking provides coupled multisensorial (auditory, vestibular, tactile) and motor isochronous experience to the fetus, thereby facilitating the development of neural sensorimotor couplings. As such, perceptual and affective biases for rhythmic stimuli are induced in the fetus, in all three sensory modalities, with the effects being predicted to be stronger when modalities are combined. Empirical research addressing evolutionary hypotheses of musicality connected to bipedalism is currently lacking. There are, however, some intriguing parallels between the optimal tempo of rhythmic synchronization (for a review see, e.g., Repp & Su, 2013) and the optimal walking tempo recorded in laboratory studies (MacDougall & Moore, 2005), as both are situated around 120 beats per minute (BPM). In addition, the tempo of popular dance music tends to cluster around 120-130 BPM (Leman et al., 2013). Humans, however, are able to synchronize with an auditory stimulus over a broader tempo range (50 - 230 BPM), that also exceeds the locomotor tempo range, which is between 75–190 BPM (Larsson et al., 2019).

3.2. Cognitive mechanisms of rhythm perception and production

Currently, there is agreement that, through evolutionary time, musicality has been shaped by species-specific **bio-cognitive constraints** (e.g., Fitch, 2015; Honing & Ploegger, 2012; Huron, 2012; Merker et al., 2015). Candidate cognitive mechanisms discussed in the literature comprise both low- and high-level features, which range from **perceptual biases** rooted in attentional, learning and memory processes to cognitive operations based on **relational processing**, such as **isochrony perception, hierarchical grouping, recursion** and **combinatorics** (Brown, 2000b; Fitch, 2013, 2015; Merker et al., 2015; Ravignani et al., 2016; ten Cate & Spierings, 2019).

The involvement of such cognitive mechanisms in shaping musical rhythm is reflected by quasi-universal structural features, including (1) the presence of an isochronous beat; (2) the hierarchical organization of beats in a metric structure of strong and weak beats; (3) the predilection for hierarchical grouping based on (multiples of) two; (4) but also three beats; (5) the presence of motivic patterns (e.g., riffs) grounded on the beat (Savage et al., 2015). Interestingly, this set of quasi-universal rhythm features has been experimentally reproduced in an iterated learning study in which participants allocated to several 'generations' in a transmission chain had to reproduce a sequence of percussive sounds for the next 'generation' of participants (Ravignani et al., 2016). At the end of the experiment, the random sequences seeded into the first 'generation' evolved into rhythmic patterns that exhibited all the quasi-universal features of musical rhythm listed above, although they also exhibited 'cultural' (i.e., transmission chain) specificity (Ravignani et al., 2016). Using a similar approach, couched in a task where participants had to tap or vocalize seeded sequences, Jacoby & McDermott (2017) found that both US and Amazonian participants favored rhythms with intervals characterized by durations based on small integer ratios (e.g., 1:1, 2:1), although the specific intervals exhibited by the two populations differed qualitatively.

Given the converging outcomes of cross-cultural and experimental research, it has been proposed that the rhythmic structure of music is shaped by mechanisms, processes and biases that generically constrain human cognition. For example, working memory constraints are arguably responsible for the emergence of features that make stimuli easier to process and recall, such as isochrony, motifs and few durational categories (Ravignani et al., 2016). Many of the cognitive traits that shape musical rhythm – and enable musicality – may have evolved for the purposes of **auditory scene analysis**, i.e., as mechanisms that enable parsing auditory input (Honing et al., 2015; Sperber, 1996; Trainor, 2018*)*. To exemplify, the tendency to assign the basic beat to the lowest-pitched instrument in an ensemble is likely rooted in inner ear adaptations that give low-pitch superiority for temporal processing (e.g., Hove et al., 2014; Nelson & Young, 2010). In contrast, it has been argued that rhythmic **isochrony**, **beat-based (i.e., relative) timing** and **entrainment** are traits that cannot be explained as non-musical adaptations (Trainor, 2018). The argumentation for such musicality-specific traits builds primarily on comparative data, which initially suggested the absence of isochrony detection, beat-based timing, and entrainment in other species. However, the evidence reviewed in 3.1., suggests that audiomotor **entrainment** can be acquired by animals through training, and that some species may even display it spontaneously.

Isochrony detection is considered to be 'the first cognitive step for beat perception' (Celma-Miralles & Toro, 2020), which, in humans, is postulated to rely on **relational processing**, i.e., the ability to extract and assess relations between items or events (ten Cate & Spierings, 2019). In the case of isochrony detection, this entails that structural patterns in the auditory stream are detected based on relative – rather than interval – timing (e.g., Merchant & Honing, 2014; Teki et al., 2011; ten Cate & Spierings, 2019). It is important to note that **relational processing** is not a music-specific ability, but can occur in a variety of domains, including theory of mind, language, and visual processing. Moreover, relational processing is not a monolithic ability, but varies in complexity from the encoding of basic relations (such the relation between a tool and its use), to the processing of highly abstract and hierarchically embedded relations. It is not yet clear if such abilities translate from one domain to the other, but, in humans, relational processing abilities, for example, have been found to correlate between the domains of language and theory of mind (Oesch & Dunbar, 2017). While the comparative study of relational processing in the domain of rhythm is in its infancy, there is suggestive evidence that starlings (Hulse et al., 1984), jackdaws (Reinert, 1965, as reported in ten Cate & Spierings, 2019), rhesus macaques (Honing et al., 2018), rats (Celma-Miralles & Toro, 2020) and zebra finches (Rouse et al., 2021) can be trained to discriminate between isochronous and non-isochronous sequences. Budgerigars, moreover, perform such discriminations spontaneously, in the absence of training (Hoeschele & Bowling, 2016). The mechanisms underlying such discrimination, however, may differ across species. For example, zebra finches and, to a lesser extent, rats seem to rely on relative timing, as they are able to generalize isochrony detection from trained to novel tempi (Celma-Miralles & Toro, 2020; Rouse et al., 2021). Rhesus macaques, on the other hand, seem to detect isochrony by attending to the absolute duration of inter-onset intervals (Honing et al., 2018; Merchant & Honing, 2014).

Beat perception and, broadly, the production and detection of metric structure arguably relies on even more complex forms of relational processing that entail **hierarchical grouping** and **recursion** (Fitch, 2013; Iversen & Balasubramaniam, 2016). Such abilities enable individuals to extract the hierarchical structure of meter, which is based on recurring patterns of strongly and weakly accented beats, with strong beats being relatively louder or longer than weak beats. For example, when exposed to rhythmic sequences where strong or weak beats are occasionally omitted, human adults and neonates react stronger to downbeat alterations (Bouwer et al., 2014; Winkler et al., 2009). So far, beat perception has not been demonstrated in animals.

For example, rhesus macaques exhibit undifferentiated reactions to occasional omissions of strong and weak beats, regardless of their position (Honing et al., 2018). Rhesus monkeys can, however, detect rhythmic groupings, as they show a differential – and stronger – reaction to omissions at the onset of a rhythmic group (Selezneva et al., 2013).

Another cognitive mechanism suggested to enable meter perception is **perceptual grouping**, i.e., the propensity to spontaneously group stimuli into units based on, e.g., their proximity or similarity (Fitch, 2016; Patel, 2006). In humans, a common form of perceptual grouping in the auditory domain is the so-called 'iambic-trochaic law', whereby sound sequences with contrasting duration or intensity/pitch are spontaneously perceived as binary groups with two distinct patterns. When the contrast is durational, humans tend to perceive an iambic structure, i.e., a rhythmic pattern accented on the second sound. When the contrast regards pitch (high-low) or intensity (loud-soft), humans tend to perceive the reversed (i.e., trochaic) pattern, with emphasis on the first element (e.g., Hay & Diehl, 2007). These propensities appear to be innate in humans (Abboub et al., 2016), although the iambic bias is susceptible to influences from the linguistic background of the individual, while the trochaic bias is robust regardless of linguistic background (Iversen et al., 2008; Molnar et al, 2016).

Research shows that animals (e.g., budgerigars: Hoeschele & Fitch, 2016; rats: de la Mora et al., 2013; Toro & Hoeschele, 2017; Toro & Nespor, 2015; zebra finches: Spierings et al. 2016) can (be trained to) discriminate both trochaic and iambic patterns, and generalize this learned discrimination to novel test stimuli. Similarly to humans, the trochaic bias appears to be more robust in other species, and the iambic bias more dependent on relevant acoustic experience (de la Mora et al., 2013; Spierings et al., 2017; Toro & Nespor, 2015). It has thus been proposed that the trochaic (but not the iambic) bias may be an evolutionarily ancient principle for sound grouping, and that the two forms of sound grouping (trochaic, iambic) may depend on distinct mechanisms (Spierings et al., 2017; Toro, 2016). In addition, there seem to be cross-species differences with respect to pattern representations that are deployed in perceptual grouping tasks. As such, when test sequences are degraded by removing one or more salient acoustic features (e.g., pitch, duration), rats are no longer able to perform a trained discrimintation (Toro & Hoeschele, 2017). In contrast, budgerigars continue to perform well when only one acoustic feature is removed (Hoeschele & Fitch, 2016). For comparison, humans are able to recognize iambic and trochaic patterns even with very degraded stimuli (Hoeschele & Fitch, 2016). These differences may reflect cross-species differences with respect to relational processing, in particular abilities related to degrees of abstraction 'power' and top-down processing, as argued in research on the recognition of severely degraded words. Interestingly, in such tasks, language-trained chimpanzees may attain human-level performance in certain conditions (Heimbauer et al., 2021), thereby suggesting a common foundation for pattern discrimination in humans and nonhuman apes.

In addition to the research reviewed in the present section (3.2), the evidence of entrainment reviewed in 3.1. implies that animals that are capable of audiomotor entrainment also perceive isochrony, since isochrony detection is assumed to be a prerequisite for synchronizing bodily responses to an auditory stimulus. Intriguingly, the only species that have so far shown spontaneous entrainment – parrots and chimpanzees – are also among the few species that display an ability to map higher-level abstract relations, thereby evidencing some **recursive** abilities. Indeed, evidence of second order relational processing in animals is currently limited to great apes (Hribar et al., 2011 and references therein), corvids (Smirnova et al., 2015) and parrots (Obozova et al., 2015). The great apes also exhibit some recursive abilities in the theory of mind domain, as they show an awareness of how others monitor their own attentional states (Hall et al., 2016) and a sensitivity to others' false beliefs (e.g., Krupenye et al., 2016). Finally, chimpanzees are capable of cross-modal audio-visual structural mapping, being able to map an abstract relationship from the visual to the auditory domain (Ravignani & Sonnweber, 2017), and can also represent distant dependencies between visual stimuli, e.g., by detecting the regularity of non-adjacent elements (Sonnweber et al., 2015). A sensitivity to distant dependencies has been also demonstrated in other primate species, including cotton-top tamarins, squirrel monkeys and common marmosets (Ravignani et al., 2013; Reber et al., 2019, and references therein). These findings suggest that sensitivity to distant dependencies, which has been argued to underlie the structuring of musical pitch and rhythm (e.g., Fitch, 2013; Patel, 2003), is an ancestral trait that probably emerged in the common ancestor of anthropoid primates.

Summing up, there is widespread agreement that both low- and high-level cognitive abilities are implicated in the perception and production of musical rhythm (see Table 3 for an overview of these and key references). Extant evidence suggests that some of these abilities support auditory scene analysis, and are shared with many vertebrates. Other relevant cognitive abilities, such as second order relational processing, have been attested only in, e.g., great apes, corvids and parrots. Currently, it is unclear which cognitive mechanisms – if any – are specialized for musical rhythm production and perception, considering that all the mechanisms reviewed in this section serve general cognitive abilities that span several domains. Much work remains to be done in order to answer this question, including research aimed at clarifying the role of relational processing in enabling musical rhythm, as well as research mapping the presence of rhythmic abilities in nonhuman species (as detailed in section 4).

Insert Table 3 about here

3.3. Socio-cognitive and motivational mechanisms

According to the most recent versions of VLRS (reviewed in 3.1), VL abilities affect both the capacity and the *motivation* to engage in rhythmic synchronization (Takeya et al., 2017; Patel, 2021). The ensuing prediction is that non-VL species can learn to entrain after extensive training, whereas high VL species will show an intrinsic motivation to engage in it. Neural specializations in the striatum are proposed as a hypothetical neural mechanism underlying this divide, considering that, in humans, this brain region is implicated in both beat perception and reward (Patel, 2021; see also section 3.4 for additional details). As such, the predilection to spontaneously engage in beat synchronization is rooted in the intrinsic rewards that high VL species experience when predicting the temporal structure of complex auditory stimuli (Patel, 2021). Consistent with this hypothesis (as reviewed in 3.1), flexible entrainment to musical stimuli has been only demonstrated in high VL species – humans and parrots (Patel et al., 2009; Schachner et al., 2009). However, spontaneous audiomotor entrainment, albeit in more rudimentary forms, has been also demonstrated in chimpanzees (Hattori et al., 2013; Large & Gray, 2015). Given the unclear VL status of chimpanzees, it is currently difficult to determine if such spontaneous rhythmic synchronization is related to VL, to other socio-cognitive and motivational traits that chimpanzees may share with humans (as reviewed further down, and also summarized in Table 4), or to both.

Another motivational mechanism potentially implicated in rhythmic synchronization is **emotional contagion**, whereby emotional cues exhibited by an individual trigger similar behaviors and corresponding states in other individuals. For example, Fritz & Koelsch (2013) argued that **acoustically mediated emotional contagion** is a homologous trait that, alongside song and drumming, has been conducive to the emergence of musicality in the *Homo* lineage. By enabling the synchronization of mood and motivations in large-sized groups of ancestral hominins, acoustically mediated emotional contagion has purportedly played a role in facilitating vital group functions (cooperation, communication, cohesion), thereby contributing to the development and establishment of rituals. As such, **acoustically mediated emotional contagion** is a mechanism consistent with **social functional theories** on the evolution of musical rhythm and musicality (as reviewed in 2.3).

Insert Table 4 about here

As discussed by Fritz & Koelsch (2013), the chimpanzees exhibit two relevant types of vocalizations that spread at group level through emotional contagion and display acoustic features reminiscent of musical structure: the play pant and the pant hoot. The play-pant is homologous to human laughter, and, like laughter, facilitates social interaction (Davila-Ross et al., 2011; Matsusaka, 2004). While human laughter has contagious and prosocial effects at the group level, great ape play-panting has, thus far, primarily been documented at a dyadic level. The pant hoot has been documented in various contexts, both affiliative and agonistic (e.g., discovering food, connecting with distant group members, patrolling, displaying). Typically, this vocalization spreads within the group, potentially amplifying arousal in a social – and indeed interactive – fashion. As argued by Fritz & Koelsch (2013), the outputs of musicality (i.e., music and dance) differ from such chimpanzee vocalizations in that they mediate a broader spectrum of emotional experiences (Juslin & Västfjäll, 2008; Koelsch, 2010; Koelsch et al., 2010), which are transmitted to groups far larger than the typical gatherings of nonhuman apes (Dunbar, 2012; Fritz & Koelsch, 2013).

With respect to relevant structural features, both these vocalizations build on the repetition of a motif, which hypothetically contributes to increasing their emotional contagiousness (Fritz & Koelsch, 2013). This kind of structural organization has been highlighted in ritualized communication (in humans and other animals, as reviewed in 2.1, 2.2 and 2.4), as well as in the strong formal structure of traditional music, which is driven by a few simple rhythms (Savage et al., 2015).

A similar mechanism has been proposed by Brown (2007) in his **contagious heterophony** theory, whereby the unsynchronized group calls of social species are a common precursor for musicality and speech. Unlike the **acoustically mediated emotional contagion** theory, the **contagious heterophony** theory sees the evolution of human musicality as an analogous rather than homologous phenomenon. According to the **contagious heterophony** theory, a crucial point for the emergence of proto-musical rhythm was the transition from reflexive vocal contagion (which is considered common in animals) to intentionally cooperative and synchronized chorusing, which is postulated as uniquely human (Brown, 2007; Bryant, 2013). Evidence suggests, however, that some nonhuman species exhibit intentionality during chorusing. For example, chimpanzees flexibly adjust the duration of pant hoot elements in a way that favors chorusing (Fedurek et al. 2013; Mitani & Gros-Luis, 1998), and actively modify their pant-hoots to create community-specific calls (Crockford et al., 2004). Similar tempo adjustments have been documented in the coordinated calls that Indo-Pacific bottlenose dolphins emit when herding or defending mates (Moore et al., 2020).

Higher-level motivational mechanisms theorized to be implicated in the evolution of in human entrainment are the **propensity for gathering to display** as a group (e.g., dance, sing: Merker et al., 2015) and **shared intentionality** (Kirschner & Tomasello, 2009). Owing to observations that chimpanzees in the wild exhibit excited communal displays, in which loud calls co-occur with drumming and exaggerated bodily movements (Geissman, 2000), Merker and colleagues (e.g., Merker et al., 2015) argued that the motivation to gather and display in large groups likely constitutes a primitive trait of musicality, which was present in the last common ancestor of humans and chimpanzees. In the human lineage, however, such displays have arguably undergone a process of refinement and elaboration, whereby the pace of vocal and bodily displays became increasingly regular, thereby enabling rhythmic synchronization.

Shared intentionality captures the purportedly human-specific motivation to share – and co-represent – mental states with others, which in turn enables individuals to jointly attend to and coordinate actions towards common goals. This kind of motivation is theorized to have emerged late in human evolution, as recently as 400 000 years ago, with the onset of obligate, collaborative foraging, which required increased and frequent joint coordinated action (e.g., Tomasello et al., 2012). The contribution of shared intentionality to the emergence of human musicality remains an understudied topic. Extant evidence – although very limited – suggests that synchronization-induced prosociality (as reviewed in 2.3) is the result of a co-dependence between shared intentionality and rhythmic synchronization, rather than being the effect of synchronization alone. As such, there is evidence that synchronization-induced prosociality selectively emerges in contexts that prompt the representation of shared goals, i.e., when rhythmic synchronization is the result of a joint intention to synchronize, as opposed to when synchronization is the by-product of individuals separately synchronizing with the same beat

(Reddish et al., 2013). Conversely, across situations that impose shared goal representation, prosocial effects *only* emerge in the presence of rhythmic synchronization. For example, participants cooperate more when instructed to work together to synchronize with respect to phase and tempo, as opposed to when instructed to work together to keep the same tempo, but by delivering sequential, rather than simultaneous, contributions to rhythm keeping (Reddish et al., 2013). Thus, it has been suggested that a **reinforcement of cooperation model** may explain the persistence of protomusical behaviors across evolutionary time. As such, perceived synchrony during the pursuit of shared goals would give an immediate feedback for achieving successful cooperation, thereby reinforcing the group's cooperative tendencies.

Developmental evidence suggests that **shared intentionality** facilitates the expression of musicality – in particular rhythmic synchronization – in toddlers. Indeed, the age at which children exhibit flexible and accurate entrainment is lowered from 4,5 to 2,5 years of age if the entrainment task (e.g., drumming) is set in a social context (Kirschner & Tomasello, 2009). The presence of a social partner supposedly prompts a shared representation of the task as a joint action, thus engaging the motivation to co-represent the task and synchronize movements in time. Fundamental components of shared intentionality, such as joint attention and engagement in imitation games, emerge already at 6 months of age in humans (Sauciuc et al., 2020 and references therein). Thus, it is also likely that shared intentionality mediates the prosocial effects of rhythmic synchronization documented in 12- and 14-month old infants (as reviewed in 2.3), since such effects are specific to social stimuli. A caveat, however, could be that in the studies reviewed in 2.3 the infants' rhythmic movement was not intentionally produced by the infants. Instead infants were rocked in a car seat or bounced by an adult. Nevertheless, this does not preclude that the infants recognized the intention of the synchronizing interaction partner to match, and thus share, their state. If the evolution of human musicality can be explained by its cooperative effects, the key mediating mechanism is likely shared intentionality.

Shared intentionality has been theorized to be unique to humans, but recent evidence shows that all great ape species exhibit behaviors indicative of it. For example, in the context of social play, all great ape species engage in joint attention and use communicative gestures to re-engage reluctant partners, which in turn suggests an understanding of collaborative roles and shared motivations (chimpanzees, bonobos: MacLean & Hare, 2013; gorillas: Tanner & Byrne, 2010; orangutans: Gruber, 2013). Moreover, when being imitated, apes show signs of enjoyment and playfulness, as expressed by imitation games, play face, and laughter (Persson et al., 2018). Rudimentary shared intentionality, such as it is attested in human infants and nonhuman great apes, may thus underlie spontaneous rhythmic engagement in these populations (for more details see 4.1 and 4.3). It is also plausible that the maturation of shared intentionality enables more precise forms of coordination in human ontogeny, including inter-personal synchronization in the auditory and visual domain. This does not entail, however, that the maturation of shared intentionality is solely responsible for the emergence of accurate and flexible entrainment in children, nor that advanced shared intentionality is solely responsible for the emergence of advanced rhythm cognition in human evolution. Evidence shows that rhythm cognition co-develops in children with a whole suite of motoric and (socio) cognitive abilities that are underpinned by several neural, neurochemical and genetic developmental transitions (e.g., Hannon et al., 2021).

3.4. Neurohormonal mechanisms of musical behaviors

Research conducted within the last two decades suggests that neurotransmitters, such as dopamine, endorphin, oxytocin, are implicated in the production of musical behaviors, as well as in mediating their positive effects. Several of these findings relate specifically to rhythm perception or rhythmic synchronization, and may account for mechanisms which, evolutionarily, promoted a predilection for rhythmic communication and facilitated the cross-generational transmission – and thus preservation – of musical behaviors (see 2.3, 2.5 for functional theories featuring such arguments). Indeed, all of these neurotransmitters are known to promote psychological well-being through interactions with the hypothalamic-pituitary-adrenal (HPA) axis (for reviews, see, e.g., Carson et al., 2013; Pilozzi et al., 2020; Stanwood, 2019). These interactions are thought to mediate prosociality during communal performance, thus being crucial for achieving adaptive benefits such as those reviewed in section 2.3, 2.4 and 2.5.

Endorphins are endogenous opioid neuropeptides produced in the pituitary gland, which are released in response to pain, and have an analgesic effect. Endorphin release has been linked to euphoric effects induced by physical exertion and rewarding activities such as sex, laughter, and eating (Chaudhry & Bhimji, 2018). The role of the opioid system in the evolution of musicality – in particular the role of *endorphins* – has been emphasized by proponents of the **social bonding** theory, based on evidence that rhythmically synchronized behavior seemingly implicates this neurochemical mechanism (Cohen et al., 2009; Dunbar et al., 2012; Tarr et al., 2015; Tarr et al., 2016). As such, active musical behaviors (e.g., communal singing, chanting, dancing and drumming) trigger endorphin release (Dunbar et al., 2012; Tarr et al., 2015; Tarr et al., 2016), while passive listening to music does not (e.g., Dunbar et al., 2012). Rhythmic interpersonal synchronization outside of a musical context appears to have comparable effects, as, e.g., synchronized rowing produces higher levels of endorphins than solo rowing (Cohen et al., 2009).

The anxiolytic function of rhythm has also been invoked by evolutionary theories that stress the **therapeutic functions of musicality** (reviewed in 2.5), linking proto-musical motor behaviors to animal stereotypic movement (Dissanayake, 2009a; Lameira et al., 2019), which is known to result in reduced stress (Chamove, 1989). Brain imaging evidence suggests that music exerts a strong influence on several amygdala clusters that contribute to emotion regulation through the release of endorphins (Koelsch, 2014). It is, thus, plausible that the endorphin system is implicated in the evolution of proto-musical rhythmic displays in the primate lineage. Indeed, in the primate lineage the endorphin system is involved in regulating social bonding beyond kin and reproductive relationships (Chang et al., 2013; Launay et al., 2016), and spontaneous rhythmic synchronization with a potential anxiolytic function has been documented in chimpanzees (Lameira et al., 2019). However, it remains to be established if the same neurochemical mechanism mediates the anxiolytic effects of musical and stereotypic behaviors.

It has long been theorized that **dopamine** is involved in mediating the processing and production of musical rhythm, and the feeling of pleasure elicited by engaging with music (Ferreri et al., 2019; Salimpoor et al., 2015). Dopamine is generally known as the main neurotransmitter involved in reward and motivation processing, although distinct dopaminergic pathways are also implicated in learning, executive function, motor function and neuroendocrine control (Alcaro et al., 2007). Several cognitive computations that are dopamine-dependent have been proposed to account for the role of dopamine in musical behaviors, such as expectations regarding rhythmic structure and the violation of these expectations (Salimpoor et al., 2015; Vuust & Kringelbach, 2010; Zatorre & Salimpoor, 2013), associative or episodic memory (Janata, 2009; Panksepp & Bernatzky, 2002), and temporal processing in the millisecond range (e.g., Merchant et al., 2013). Starting with the seminal study of Blood & Zatorre (2001), brain imaging research has repeatedly established increased activity in dopamine-rich areas, such as the striatum, when listening to pleasurable music (for reviews see Koelsch, 2014; Zatorre, 2015). Direct evidence for the role of dopamine in rhythmic motor control and synchronization has only recently begun to emerge (Brodal et al., 2017; Koshimori et al., 2019; Miller et al., 2013).

Freeman (2000) suggested **oxytocin** as an underlying mechanism of the social bonding function of musicality, while Feldman (2007) suggested that interactional synchronization triggers the release of oxytocin. Oxytocin is an evolutionarily conserved neuropeptide that, in mammals, is implicated in sexual and parental behaviors through effects exerted on subcortical structures in the forebrain and amygdala (for a review see, e.g., Rogers et al., 2018). In humans and other primates, oxytocin and vasopressin have also been linked to social cognition and affiliation beyond sexual- or kin relationships (Crockford et al., 2013; Festante et al., 2021; Samuni et al., 2018; Webb et al., 2019), owing to effects exerted on cortical regions relevant to social reward, emotional recognition and empathy.

Recent empirical research on the involvement of oxytocin and vasopressin in musical behaviors appears to confirm the early suggestions of Freeman (2000) and Feldman (2007), as intranasally administered oxytocin was shown to increase interpersonal rhythmic synchronization (Gebauer et al., 2014). Conversely, singing in group (Grape et al., 2002) and passive music listening (Nilsson, 2009) may lead to an increase in peripheral oxytocin, especially for slow-paced rhythms (Ooishi et al., 2017). Moreover, therapeutic interventions consisting of group drumming sessions were found to result in increased oxytocin levels and well-being in maltreated children (aged 8-12 years, Yuhi et al., 2017). Other studies, however, have reported a decrease in peripheral oxytocin after solo and/ or group singing (Bowling et al., 2022 and references therein). Such discrepant findings are most likely explained by methodological differences related to procedural details and sample characteristics. As discussed by Bowling et al. (2022), contextual factors, such as stress levels, the relationships between participants and the level of intimacy implicated by the procedure may interact with – and affect – the dynamics of oxytocin expression. The current literature further suggests that rhythmic tempo (Ooishi et al., 2017) and rhythmic movement (Gebauer et al., 2014; Yuhi et al., 2017) are relevant variables that should be examined by future research.

Insert Table 5 about here

As reviewed above, recent evidence suggests that several neurotransmitters are implicated in the production and/or perception of musical rhythm (see Table 5 for an overview and key references). Since these neurotransmitters are ancestral mechanisms for promoting well-being, the evidence reviewed in this section is consistent with the therapeutic function of musical rhythm (and musicality), as outlined in 2.5. Moreover, it is well-established that these neurotransmitters are implicated in regulating a wide range of socio-behavioral traits in humans, as well as social bonding in primates. Thus, the evidence reviewed here is also consistent with the **social bonding** theory on the evolution of musical rhythm (as reviewed in 2.3). Indeed, it has been argued that human musical behaviors are a recently evolved form of social bonding that has appropriated the neurochemical mechanisms of ancestral bonding behaviors existent in primates (e.g., Dunbar et al, 2012). Additional studies are, however, required to pinpoint the specific role that each of these neurotransmitters plays in the production and appreciation of musical rhythm. Indeed, in the formation and expression of socio-affiliative behaviors, oxytocin and vasopressin interact with dopamine, which is believed to exert a reinforcing influence on behaviors conducive to, e.g., affiliation (Baskerville & Douglas, 2010; Carson et al. 2013; Riedl & Javor, 2012). Similar interactions are likely to characterize the production and appreciation of musical rhythm, considering the bidirectional relationships that exist between rhythmic synchronization and social bonding (as reviewed in 2.3). Moreover, given the co-localization of the dopaminergic and opioid systems, it is likely that the hedonic effects of music-related behaviors are due to an interaction between the two systems, and, possibly, also with the oxytocinergic system.

Before concluding this section, we will mention relevant findings from research on the neurochemical mechanisms of birdsong, which is so often invoked in discussions on the evolutionary foundation of human musicality. This research reveals that birdsong is highly dependent on steroid sex hormone release and hormone-neurotransmitter interactions, both during ontogenetic development and throughout the reproductive cycle. Songbirds appear to be, in fact, unique among vertebrates, in that they display considerable amounts of sex steroid receptors in non-limbic regions of the forebrain (Alward et al., 2017; Schlinger & Brenowitz, 2002). Such receptors abound in neural assemblies that are implicated in song production, thus suggesting that sex steroids are directly involved in controlling birdsong. This high dependency on the release and fluctuations of sex steroids is not seen in other bird species, not even in other vocal learning birds, such as parrots (Schlinger & Brenowitz, 2002). The prominence of sex hormones in birdsong control might explain the functional inflexibility of birdsong (as compared to human music), with birdsong being typically linked to either mating or territorial defense.

The role of sex steroid hormones in the evolution of human musicality – and musical rhythm, in particular – has been minimally investigated, with preliminary results being inconclusive. On the one hand, a recent exploratory analysis found no relationship between singing and testosterone levels in males (Bowling et al., 2022). On the other hand, a single-subject study found evidence that sexual arousal in males improves rhythmic synchronization, thus suggesting that rhythmic synchronization may be mediated by sex hormones (Miani, 2016). Sexual arousal, however, is also mediated by vasopressin – a neuropeptide closely related to oxytocin. Just like oxytocin, vasopressin is implicated in regulating social behavior and binds to receptors in the basal ganglia (reviewed by Baribeau & Agnastou, 2015) – a neural structure whose involvement in rhythmic processing is well-established (Grahn & Brett, 2007; Nozaradan et al., 2017). Evidence shows that sex steroid hormones affect the expression of oxytocin and vasopressin in animal models, although in humans the relationships between sex steroid hormones and neuropeptides is not entirely clear, and, in certain studies even fails to materialize (for a review, see Baribeau & Agnastou, 2015). Altogether, these findings suggest that sex steroid hormones may exert an indirect influence on human rhythmic synchronization (and other musical behaviors), given the implication of oxytocin in such behaviors and the potential influence of sex steroid hormones on oxytocin. As such, future research should more directly target the involvement of sex steroid hormones in rhythmic behaviors and synchronization in humans. Similar research with animals will help address if this potential neurochemical mechanism of rhythmic behavior is a trait shared with other species, thereby informing theories on the evolutionary foundations of human rhythm cognition and of musicality, in general.

3.5. The genetic basis of musical rhythm

The heritability of rhythmic abilities has been assessed in twin studies and with molecular genetic methods. The twin studies have yielded somewhat contrasting results. Ullén et al. (2014) found evidence of moderate (i.e., 50%) heritability for rhythm perception in a task that required participants to make similarity judgements for rhythmic strings. In contrast to these findings, Seesjärvi et al. (2016) found that rhythm perception, as measured by the ability to detect off-beat sounds, was best predicted by non-shared environment effects, as opposed to genetic or shared environment effects. These inconsistent findings may reflect distinct task requirements for measuring rhythm perception ability. As such, similarity judgments are likely to rely more on innate abilities for **auditory scene analysis** and even on mechanisms that do not necessarily require **beat-based timing** (e.g., ten Cate & Spierings, 2019). In contrast, off-beat detection is more directly linked to beat-based timing, although it is also likely to capitalize on experience and acquired knowledge. Additional studies probing the heritability of the mechanisms that underpin the emotional processing of musical rhythm, in particular dopaminergic striatal functions and music-induced rewardness, indicate that these do not appear to be affected by genetic or shared environmental effects (Menne-Lothmann et al., 2012; Stokes et al., 2013).

Recent genome-wide association studies corroborate the findings above. As such, the heritability estimates for the ability to synchronize to a musical rhythm range between 13-16% (Niarchou et al., 2021). The heritability estimate for rhythmic discrimination skills is reported to be 21%, both by an initial small-sample study, and a subsequent large-scale replication (Oikkonen et al., 2015; Pulli et al., 2008).

Genomic analyses in humans reveal a similar pattern of findings, whereby genes implicated in the development of auditory pathways, sound production, as well as non-specialized cognitive processes (such as learning, memory and attention) are important for musical skill (Mosing et al., 2014; Oikkonen et al., 2015). Such molecular mechanisms that are implicated in musical skill, but are not necessarily specific to musicality, appear to be ancient and well-conserved in vertebrates (Oikkonen et al., 2016).

Genetic correlation analyses further indicate that the gene coding for the 1a receptor of the neuro-hormone arginine-vasopressin (which is widely implicated in regulating social behavior and socio-behavioral traits), is associated with traits related to musical rhythm, such as creative dance (Bachner et al., 2005), rhythm perception and rhythm memory (Granot et al., 2007; Ukkola et al., 2009). A study, which more directly targeted the ability to synchronize with musical rhythm, found that rhythmic synchronization is a highly polygenic trait that shares molecular architecture with traits relevant to emotion, cognition, neurodevelopment and health (Niarchou et al., 2021). As such, better rhythmic synchronization correlates with better lung function, faster processing speed in matching tasks, stronger hand grip, and the evening chronotype (i.e., people who prefer activity late in the day). Moreover, genomic loci related to neural structures that are crucially implicated in rhythm cognition (cerebellum, basal ganglia and cortex) contribute to the heritability of rhythmic synchronization.

Recently, the genetic signature of human rhythmic synchronization (as outlined in Niarchou et al., 2021) has been compared to several zebra finch gene sets that reflect genes expressed in brain structures known to be implicated in birdsong (Gordon et al., 2021). This comparison revealed that the genetic architecture of human rhythmic synchronization is, to some extent, shared with that of birdsong, as the former was significantly enriched for the avian gene sets expressed in Area X – the avian analogue of mammalian basal ganglia. This shared genetic substrate between human rhythmic synchronization and learned vocalizations in zebra finch provides some support to VLRS (reviewed in 3.1). A first caveat, however, to this interpretation, as this shared genetic architecture may represent a homologous – rather than a convergently evolved – trait (Gordon et al., 2021), especially considering that the mammalian basal ganglia are broadly associated with dopaminergic control, action selection and timing, motor learning, as well as value-based decision making (for a review, see, e.g., Grillner & Robertson, 2016). To address this caveat, future research should, for example, include vocal non-learning species (both avian and mammalian) in such comparisons. A second caveat is that the study did not find significant enrichment for the gene sets expressed in three other brain structures – the HVC (proper name), the lateral magnocellular nucleus of the anterior nidopallidum (LMAN) and the robust nucleus of the arcopallium (RA) – that are crucially involved in birdsong (Nieder & Mooney, 2019). The HVC-RA circuitry, in particular, is commonly referred to as the forebrain song control circuitry and is regarded as analogous to the human laryngeal motor cortex (LMC). The HVC-RA and the LMC circuitry share a feature considered to be a recent neural adaptation for advanced VL, i.e. they exhibit *direct* connections to vocal motor neurons (Nieder & Mooney, 2019; Patel, 2021; Vernes et al., 2021).

Currently, very little is known about the genetic underpinnings of traits related to musical rhythmicity, with the handful of studies conducted so far targeting primarily rhythm discrimination (see Table 6 for an overview of current findings and key references). Future studies will thus need to investigate the heritability of additional rhythmic abilities, and uncover their molecular underpinnings. Since a small inventory of rhythmic features are universally present in musical traditions across the world, and since these seem connected to specific perceptual and production skills, it is highly plausible that genetic factors play an important role in the variability of their expression. Genetic studies targeting each of these skills – as well as their interplay – will provide valuable data for research on the origins and evolution of human rhythmicity and musicality.

Insert Table 6 about here

4. A theoretical and empirical agenda for research on animal rhythmicity

As outlined above, the research on the evolutionary origins of musicality – and rhythm cognition, in particular – is currently characterized by an effervescence of theories and a mosaic of fragmentary evidence. Given this motley empirical background, researchers have launched repeated calls for systematic data collection, in which various animal taxa are emphasized as highly relevant for an evolutionary inquiry into the origins of musical rhythm and associated behaviors. Recurrent taxa in these calls are seals, whales, parrots and songbirds (Fitch, 2006; Kello et al., 2017; Kotz et al., 2018; ten Cate et al., 2016), thus implicitly suggesting underlying assumptions about convergent evolution. In other words, these calls imply that the rhythmic abilities involved in musicality may have repeatedly evolved in various species due to socio-environmental pressures similar to those that influenced their evolution in humans. Proposals favoring a homology-based perspective – i.e., that emphasize the phylogenetic continuity of musicality-relevant traits in the primate lineage – are less common. Nevertheless, given the current componential approach to the evolution of rhythmic abilities, homology and convergence need not be mutually exclusive. For example, the phenomenon of tempo adaptation during communal calling in a cooperative context has been described in both chimpanzee (Crockford et al., 2004; Fedurek et al., 2013; Mitani & Gros-Luis, 1998) and dolphins (Moore et al., 2020). This could suggest that, in humans, tempo adaptation during communal sound-making is a phylogenetic trait ostensibly present in the last common ancestor of humans and chimpanzees. At the same time, the trait is also shared with dolphins, a species in which it presumably evolved by analogy, given similar evolutionary constraints related to cooperation. Considering the current empirical status of the field, it is, however, premature to jump to any conclusions. In the next section, we will outline several research topics that need to be consolidated empirically – and in a systematic manner – if we are to make any major progress in disentangling the evolutionary foundations of human musicality.

4.1. Musical rhythm is not a monolithic ability – moving beyond audiomotor entrainment

The empirical evidence reviewed in sections 2 and 3 suggests that musical rhythm is not a monolithic property of music, and that synchronizing to a musical rhythm is not a monolithic ability. Instead, humans exhibit a suite of rhythmic abilities, which fall along a continuum of perceptual and motor sophistication. This point is well-illustrated by the protracted development of rhythmic abilities in human ontogeny, with perceptual rhythmic abilities generally preceding rhythmic production. With respect to perceptual abilities, experimental evidence suggests that neonates exhibit rudimentary beat perception (Winkler et al., 2009). Another milestone occurs at 3-5 months of age, when infants can learn to discriminate between audio-visual rhythmic patterns (Bahrick & Lickliter, 2000; Bahrick et al., 2002). Around 7 months of age, infants begin to categorize rhythms based on their underlying beat (Hannon & Johnson, 2005), and at 8 months of age they begin to detect metric asynchrony in complex audio-visual stimuli, an ability which is significantly consolidated at 12 months (Hannon et al., 2017).

With respect to production abilities, from 3 months of age, infants begin to sparingly show rhythmic engagement, i.e., they display rhythmic movement when exposed to musical stimuli, and adapt bodily responses to the faster or slower tempo of the stimulus (e.g., Fuji et al., 2014). Accurate audiomotor entrainment is only attained at 4 years of age, albeit in a rudimentary form, whereby entrainment is generally restricted to stimuli that match the children's spontaneous tempo. As reviewed in 3.3, the age when this ability is expressed can be lowered to 2,5-years if the exogenous stimulus is embodied by a social partner (Kirschner &

Tomasello, 2009). Adult-like proficiency with respect to timing accuracy and tempo flexibility is only attained at 8-10 years (as reviewed by Hannon et al., 2017; Hannon et al., 2021), an age when children have also formed stable representations of the metrical structures present in their culture (e.g., Hannon et al., 2012). The development of advanced abilities related to rhythm perception and production coincides with important milestones in the physical, hormonal, cognitive and socio-cognitive maturation of children during middle childhood (as reviewed by Hannon et al., 2021). In turn, this points to biomechanical, neural, neurochemical, genetic, (socio) cognitive and cultural mechanisms that need to be taken into account by evolutionary theories of musical rhythm.

In the light of this evidence, we argue that comparative and evolutionary research will benefit from considering the full set of quasi-universal features of musical rhythm (as outlined in 3.2.), as well as the full spectrum of rhythmic abilities as they unfold in human ontogeny. Taking into account developmental precursors to mature adult musical behaviors may provide unique insight into the functions served by particular musical features and behaviors, and may also explain certain aspects of cross-cultural variability (Hannon et al., 2021). As reviewed in 3.1, animal studies initially focused on entrainment, using methods that required animal subjects to produce a motor response (tapping, head bobbing, etc.), and synchronize it with an exogenous stimulus. Developmental data indicate that, in our species, accurate audiomotor entrainment is only achieved after years of practice and exposure to (auditory) rhythms. Thus, when probing animal rhythm cognition, a heavy focus on motor responses that need to be acquired relatively quickly (i.e., within the timespan of an experiment), most likely underestimates the rhythmic abilities of those species, in particular those related to rhythm perception (as suggested by the recent studies on isochrony detection reviewed in 3.2).

Observational studies, playback experiments and detailed analyses will be needed for gaining insights into the kind of rhythmic patterns and structures that different species produce and are likely to discriminate. The extensive body of research on insect rhythmic chorusing (as reviewed in 2.1) provides an excellent model, both with respect to methodological practices and the level of theoretical refinement that can be achieved based on data generated with such methods. So far, in vertebrates, the emphasis has been on the need for detailed analyses of multi-syllable, complex calls, that potentially exhibit a hierarchical structure, and that have been likened to music in the past, e.g., bird, whale or gibbon song. We maintain that the focus of such detailed analyses needs to be broadened to encompass vocalizations that we do not perceive as melodic, but which nevertheless exhibit a rhythmic structure, as well as non-vocal rhythmic behaviors. Such studies are fundamental for mapping the perception and production capabilities of animals with respect to rhythmic structures of variable complexity, and for assessing if structural features of the temporal organization of animal signals bear similarities to musical rhythm. *Rhythm production capabilities* that are regularly involved in species-specific communication can be directly extracted from observational data with the help of appropriate analytic tools. *Perception capabilities* can sometimes be inferred from observational data. For example, when the data contains sufficient signal variability with respect to given features (e.g., tempo, isochrony, timbre) and information concerning conspecific responses, it should be possible to determine which rhythmic features – if any – are likely to be salient to individuals of a given species, as they should induce a measurable response. Playback experiments that systematically manipulate specific features of species-specific signals would be most useful in this respect.

Acoustic analyses of rhythmic features have recently been conducted on the calls of several marine mammal species (humpback whales: Schneider & Mercado, 2019; killer whales: Kello et al., 2017; Northern elephant seals: Mathevon et al., 2017; harbor seal: Ravignani et al., 2019b), birds (e.g., Kello et al., 2017; Roeske et al., 2020) and primates (e.g., de Gregorio et al., 2021; Gamba et al., 2019; Schruth et al., 2019; Terleph et al., 2017). Comparative studies that include multiple species are also under way. For example, a comparison of the complex vocalizations of whale (humpback whale, killer whale) and bird (nightingale, hermit thrush) species with human speech and various music genres have revealed important differences (Kello et al., 2017). As such, hierarchical grouping and recursion were found to be specific to human music, in which a small inventory of notes is typically combined into repetitive structures with a nested clustering of subunits. Bird and humpback whale song, on the other hand, relied on adjacency relationships, whereby only adjacent phrases shared sound units, but distant ones did not. Killer whale vocalizations had features more similar to the structure of human conversation, in that they exhibited features of turn-taking coordination of offsets, which is comparable to findings from a study on duet coordination in gibbon song (Terleph et al., 2017).

Data from two recent studies have uncovered that the song of thrush nightingales (*Luscinia luscinia*) and indri lemurs (*Indri indri*) exhibits a number of similarities to human musical rhythm at the lowest level of rhythmic organization, i.e., when comparing two successive inter-onset intervals in a sound sequence (de Gregorio et al., 2021; Roeske et al., 2020). As such, both indri and thrush song exhibited isochronous rhythms that favored small integer ratios, although the tendency for 1:2 ratio was significant only in the indris. Thrush song was also compared to human music, revealing that both exhibited comparable amounts of isochrony (Roeske et al., 2020). High-ratio rhythms, on the other hand, were very frequent in thrush song, as opposed to human music where these typically function only as ornaments.

Finally, in both thrushes and human music the number of rhythm categories decreased with faster tempo*,* to ultimately transition into stereotyped isochrony. However, the threshold of this transition was much lower in thrush song (135 ms) compared to human music (265-410 ms), thus suggesting that only thrushes produced variable rhythms at tempi that equated the limit imposed by species-specific biomechanical constraints. In contrast, in human music, isochrony was present at tempi that were much slower than such constraints.

Worth of mention are also the findings of a comparative study that assessed the musical quality of the calls in 45 primate species (Schruth et al., 2019). Although rhythm was one of six musical variables scored in this work, this variable was subsequently incorporated into a composite index of "acoustic reappearance diversity", which the authors used as a proxy for musical quality. This index was intended to capture the balance of redundancy (internal repetition) and diversity considered to be characteristic of human music (Brown & Jordania, 2013; Nettl 2015), by quantifying the number of distinct syllables in a call, as well as their reappearance. When pitted against ecological and social variables, this composite index was found to be higher for species with an arboreal lifestyle, as well as for species that live in monogamy or small groups (2-6 individuals). These results are intriguing, as they run against several of the theories presented in section 2 and 3. Most notably, the **social bonding** theory (section 2.3) attributes the emergence of musicality to selective pressures imposed by increased group size (50+) in the human lineage. Other theoretical proposals that invoke increased group size as the selection pressure implicated in the emergence of human musicality are the **coalition signaling** theory (section 2.3), the **credible signaling of parental attention** theory (section 2.4), and certain **therapeutic function** theories, such as the hypothesis that rhythmic synchronization may have emerged as a **coping co-stereotypy** (section 2.5). Mechanistic theories that relate the emergence of musical rhythm to bipedalism, such as the **incidental sounds of locomotion** theory (section 3.1) and the **coping co-stereotypy** hypothesis (section 2.5) are also contradicted by the findings of Schruth et al. (2019), since the emergence of bipedalism is related to terrestrial life (section 3.1). All these theories need, however, not be dismissed given their more specific focus on rhythmic cognition as opposed to musicality in general. In contrast, Schruth et al. (2019) used a composite index of musicality, and thus no separate correlation analyses were conducted between socio-ecological variables and distinct components of musicality. Since this composite index stresses complexity as a fundamental feature of human music, it would also fail to capture several forms of human music in which simplicity predominates. Moreover, there are reasons to assume that acoustic structural complexity was not one of the foundational features in the emergence of human musicality, but likely it was a culturally evolved one (Nettl, 2000). Thus, theories that run against the findings of Schruth et al. (2019) may still offer good accounts of putative selection pressures involved in the emergence of human rhythmic cognition. Additional research along the lines suggested throughout this section will be instrumental in clarifying such issues.

The mere presence of certain rhythmic features in the vocalizations (and movements) of other species does not necessarily imply that individuals of these species possess rhythmic abilities that are relevant to the evolution of musicality. In other words, rhythmic structure in animal signals does not entail that conspecifics *perceive* rhythmic structures in these behaviors in the same way as humans can, nor that they will be able to detect rhythmic structure in arbitrary stimuli. Nevertheless, sufficient and systematic data on consistent variations as a function of relevant variables (e.g., contexts, social consequences, sex, potential physiological correlates) can provide valuable insights into the evolution of rhythmic abilities. Data on such variables can, in turn, inform playback experiments that will allow us to determine if animals attend to the temporal structure of rhythmic behaviors, and whether their perception of rhythmic signals relies on interval or relative timing. For example, available playback studies indicate that both song- and non-songbird species may be sensitive to the temporal structure of vocalizations, but also that some songbird species are indifferent to it (Fishbein et al., 2019; Slabbekoorn & ten Cate, 1999 and references therein).

A recent playback study with Northern elephant seals (*Mirounga angustirostris*) demonstrates the importance of studying non-melodic calls (Mathevon et al., 2017). Throughout the breeding season, after the male hierarchy is established through fighting, mature males of this species emit threat calls that consist of a rhythmic series of pulses with individual-specific tempo and timbre signatures. By playing back calls recorded from dominant males, as well as altered versions of these calls, in which tempo, timbre and rhythmic structure were modified, Mathevon et al. (2017) showed that Northern elephant seals can discriminate a variety of tempi and rhythmic patterning. Indeed, the navigational decisions of *beta* males were influenced by stimulus manipulations, in that they would inhibit approach when call features were within the range of more dominant individuals. This finding indicates that Northern elephant seal males memorize and recognize the individual rhythmic signatures of their rivals, and use this information for navigating social life in the colony.

Finally, the usefulness of investigating non-vocal rhythmic behaviors beyond entrainment is illustrated by two recent primate studies: an observational study of cao vit gibbons (*Nomascus nasutus*, Fan et al., 2016) and an experimental study with chimpanzees (Hattori & Tomonaga, 2019). Female cao vit gibbons were observed making sequential movements of single body parts that have been likened to human so-called ‘robot dance’ and described as rhythmic. Directed towards males, these displays seemingly function as a form of sexual solicitation. The chimpanzee study, on the other hand, brings evidence of rhythmic engagement with an exogenous auditory rhythm, which was documented in both males and females. Moreover, in a follow-up single-subject experiment, the involved chimpanzee flexibly adapted the tempo of his movement to the tempo of musical stimuli (Hattori & Tomonaga, 2019). Intriguingly, however, the chimpanzee also exhibited rhythmic engagement when exposed to random – as opposed to periodic – rhythm. This may suggest important differences between chimpanzees and human infants with respect to rhythmic engagement, with the caveat that tests of rhythmic engagement in human infants have not included random musical stimuli as a control condition, although they contrasted infants' responses to musical rhythm with response to, e.g., infant-directed speech, which is characterized by an irregular rhythm (Fuji et al., 2017 and references therein). The two primate studies mentioned here provide examples of spontaneous bodily displays with rhythmic structure, as opposed to the trained bodily movements featured by the experimental studies on entrainment reviewed in section 3.1. Non-vocal rhythmic displays have been also documented in palm cockatoos (*Probosciger aterrimus*), in the form of courtship drumming displays that are directed by males towards females (Heinsohn et al., 2017).

Insert Figure 1 about here

In conclusion, after mostly focusing on audiomotor entrainment, comparative research is now beginning to direct attention to other rhythmic abilities. Recent studies have, for example, targeted isochrony detection and beat perception, have attempted to map and compare the rhythmic structure of animal calls to that of musical rhythms, and have begun to examine non-vocal rhythmic behaviors. We argue that this research can further be expanded by systematically targeting rhythmic abilities that, in human ontogeny precede the emergence of audiomotor entrainment (as reviewed above), and structural features of musical rhythm that are cross-culturally universal (see Figure 1 for a synopsis of future research directions outlined in this section). Systematic research targeting such perception and production abilities separately will put us on a better track for determining ancestral traits that are shared with other species, as well as potential innovations that have emerged in the *Homo* lineage.

4.2. Rhythmicity is not modality specific – moving beyond the auditory domain

Rhythmic synchronization is epitomized by the ability to align behaviors with a musical beat. In humans, however, this ability extends beyond the realm of music and dance, and beyond a sensitivity to rhythmic stimuli in the auditory modality. It enables a great variety of activities that rely on precisely timed teamwork through jointly synchronizing to a rhythm, from rowing in a team to military marches. It is highly likely that, throughout human history, rhythmic synchronization has been frequent in coordinating work during routine chores, such as grain pounding, rope pulling, tree chopping, group paddling, etc. (e.g., Dissanayake, 2009a).

Recent experimental evidence suggests that, rather than being modality specific, human rhythmic abilities are susceptible to multi-modal stimulation (as reviewed by Iversen & Balasabrumaniam, 2016). Initially, evidence from several studies showed that humans synchronized better with auditory than visual stimuli (for a review see Repp & Su, 2013). These findings appeared to conform to a broader pattern, whereby, unlike other primates, humans would have superior short-term and recognition memory in the auditory modality compared to other modalities. Inferior memory performance in the auditory modality compared to the visual or tactile modality has been documented in Old World and New World monkeys, as well as one chimpanzee (as reviewed by Bigelow & Poremba, 2014). However, when tested empirically, the assumption of human superiority in the auditory modality did not hold. Human participants, similarly to other primates, showed inferior memory for auditory stimuli compared to both visual and tactile stimuli (Bigelow and Poremba 2014).

Research also suggests that the auditory advantage, which humans appear to exhibit in the context of rhythmic synchronization, could be stimulus- and experience-driven (for a review see, e.g., Iversen & Balasubramaniam, 2016). Indeed, the auditory advantage disappears when humans synchronize to dynamic visual stimuli (e.g., bouncing balls) as opposed to static ones (e.g., blinking flashes), which were the customary visual stimulus in early synchronization studies (e.g., Gan et al., 2015; Iversen et al., 2015). Moreover, just like audiomotor entrainment (e.g., Madison, 2014), visuomotor entrainment to complex stimuli improves in the presence of metrical subdivisions (Su, 2016), thus extending the evidence on the amodal nature of rhythmic synchronization even to the higher-level of meter processing. There is also evidence that entraining to vibrotactile rhythms can attain comparable accuracy to audiomotor entrainment, although entrainment to nonisochronous, complex rhythms is superior when rhythms are presented in the auditory – compared to the somatosensory – modality (Gilmore & Russo, 2021 and references therein). Additional research shows that, contrary to what might be expected, visual – as opposed to auditory – information has a greater impact on people's ratings of musical performance quality (Griffiths & Reay, 2018; Tsay, 2013). For example, humans are significantly more successful at identifying winners of musical competitions using visual information of their performance only, compared to using auditory information only (Tsay, 2013). Moreover, when presented with manipulated audiovisual recordings of a classical piece in which the audio recording of a professional performer is paired with the visual recording of an amateur performer and vice-versa, both musicians and non-musicians rate the latter (i.e., the visual recording of the professional performer plus audio recording of the amateur performer) as being of higher musical quality (Griffiths & Reay, 2018).

Consistent with these findings, brain imaging evidence shows that auditory, visual and vibrotactile rhythms activate a common neural network for beat detection, which is mostly supramodal, thus transcending sensory modalities (Araneda et al., 2016). In fact, this might extend to all vocal communication (including speech), considering the multisensorial nature of most neocortical areas, including the primary auditory, somatosensory and visual cortices, even though these are, typically, regarded as the foundation of unimodal processing (Ghazanfar, 2009; Ghazanfar & Schroeder, 2006). Research shows, for example, that the auditory and visual components of speech are processed simultaneously at the earliest level in the neural circuitry, rather than unimodally and serially (Chandrasekaran et al., 2009: Rosenblum, 2008).

Evidence from early human infancy or prenatal development further supports the hypothesis that human rhythmic abilities are not modality specific. Research shows that fetuses are able to discriminate between vibrotactile rhythms (e.g., the mother's walking vs. swaying vs. rocking), as evidenced by changes in heart-rate response patterns (Lecanuet & Jacquet, 2002). Such vibrotactile stimulation appears to be crucial for normal fetal development, which is also evidenced by the benefits of rocking in promoting typical development in prematurely born infants (as reviewed by Provasi et al., 2014), as well as by the worldwide use of rocking to calm young infants (Richter & Ostovar, 2016). In both fetuses and young infants, cross-modal (e.g., vibro-acoustic, audio-visual) presentation of a rhythm provides a processing advantage compared to unimodal presentation of an auditory rhythm. As such, fetal responses are higher when fetuses are exposed to vibro-acoustic stimuli compared to unimodal stimuli (Kisilevsky & Muir, 1991; Kisilevsky et al., 1992). In addition, 3- and 5-month old infants can learn to discriminate a rhythmic pattern when the pattern is presented synchronously in the visual and auditory modalities, but fail to learn the rhythm when it is presented unimodally, in either the visual or auditory modality (Bahrick & Lickliter, 2000; Bahrick et al., 2002).

Taken together, the findings presented in this section suggest that unimodal theories likely provide an incomplete account of musicality in general, and of rhythmic abilities in particular. Examining the rhythmic behaviors of other species beyond vocal production and auditory reception will likely provide further insights into the evolution of human musicality. It is worth noting that a similar need exists in human developmental research, where the overwhelming majority of studies focus on rhythms in vocal communication and auditory perception, while the perception and production of *bodily* rhythms remain relatively underexplored (Provasi et al., 2014). For the purposes of comparative research, relevant (non-vocal or cross-modal) behaviors are those that fulfill a communicative function in other species, thus excluding biological rhythms such as breathing or locomotion. Observational and playback experiments similar to those mentioned in section 4.1. would be instrumental in this respect, but with a focus informed by research questions pertaining to modality. Investigating rhythmic communicative behaviors that take the form of bodily displays may reveal behavioral patterns and rhythmic features that are comparable to e.g., human dance and musical rhythm. For example, the detailed analysis of rhythmic co-locomotion in chimpanzees (i.e., the so-called ‘Conga line’) has recently led to the proposal that human proto-dance may have been rooted in stress-relieving behaviors that initially did not involve any sound (Lameira et al., 2019). Our own observations show that, generally, chimpanzees exhibit several rhythmic behaviors that seem to accomplish a communicative function, as these behaviors are primarily exhibited in a social context (Sauciuc et al., 2019a), and exert a social influence on conspecifics (Sauciuc et al., 2019b). Many parrot species also show spontaneous bodily movements (e.g., head bobbing) that seem to have a communicative function.

Non-communicative rhythmic behaviors (e.g., respiration, locomotion) are not completely without relevance for the evolutionary study of musical rhythm (see e.g., Pouw et al., 2021 for a similar point). Comparing e.g., respiratory and/or locomotor rates with the output rates of communicative rhythmic behaviors in various modalities will provide valuable data to determine if there are biological couplings across modalities and behaviors. In turn such findings might point to underlying mechanisms in the form of, e.g., species-specific internal-clock like mechanisms. This could also be informative for understanding how and why optimal rhythmic rates emerge, and for explaining variations in such rates. For example, the optimal finger tapping rate in human adults and common chimpanzees (Hattori et al., 2013; Repp & Su, 2013) appears to be similar (around 500-600 ms inter-onset-interval), while the rate of drum striking in bonobos and human children (Large & Gray, 2015; Repp & Su, 2013) are higher (around 300-450 ms inter-onset interval). Such data will also be helpful for assessing the predictions of the **acoustic advantages** hypothesis (Larsson & Abbott, 2018), which links the emergence of rhythmic synchronization to locomotor-respiratory couplings present in fish, and therefrom to audiomotor couplings between forepaw processing and vocalizing in tetrapods. Biological constraints grounded on such couplings appear to drive the rhythm of isochronous echolocation calls in several bat species, where such calls are coupled with wingbeat and breathing cycles (for a review see Burchardt et al., 2019). Recent evidence further shows that in the Greater sac-winged bat (*Saccopteryx bilineata*), a single frequency dominates the rhythm of both echolocation and social calls, such as pup isolation calls and male territorial songs (Burchardt et al., 2019). It is intriguing that these types of calls share a similar frequency (which is also shared with non-communicative rhythms), considering that isolation calls are innate (and later in life repurposed as appeasement calls in males) while territorial songs, in contrast, are learned (Eckenweber & Knörnschild, 2013).

Systematically mapping the biomechanics of animal gaits and the rhythmic patterns they produce will allow testing hypotheses that emphasize **bipedalism** as a crucial trait for the emergence of human rhythmic abilities. As reviewed in section 3.1, it has been proposed that bipedal walking more readily affords rhythmic synchronization (compared with quadrupedal locomotion), by virtue of entailing biphasic isochronous movement (see Larsson, 2013; Larsson et al., 2019). Testing this hypothesis requires the study of locomotion rates and perceptual rhythmic abilities in species in which locomotion is accomplished through biphasic periodic movements, such as flight and walking in birds, or stotting (i.e., springing with all four legs in the air) in mammalian species. Also, it remains to be investigated if other gait patterns have an influence on rhythm *perception*, for example, if quadrupeds are primed by quadrupedal gait patterns, thus showing heightened sensitivity to or a preference for non-binary rhythms.

Finally, experimental studies with animals should also compare rhythmic production and perception across modalities, as well as compare the perception of rhythmic and non-rhythmic stimuli in different modalities. Such experimental data will be helpful in assessing hypotheses that rhythmic signaling entails transmission advantages and patterning flexibility to facilitate the recognition of species-specific calls (as reviewed in section 2.1). Recognition and mnemonic advantages are also predicted by **the dynamic attending theory** (as reviewed in 3.1). While memory for auditory stimuli poses a challenge for many species, including humans (Bigelow &

Poremba, 2014 and references therein), this challenge may be lessened by rhythmic patterning, as e.g., suggested by several of the theoretical proposals reviewed in sections 2 and 3.

Insert Figure 2 about here

Summing up, our call to expand the focus of research on the rhythmic abilities of animals beyond auditory perception and vocal production encompasses behaviors with – as well as without – a *communicative* function (as also shown in Figure 2). Such research should examine the presence, range and function of rhythmic behaviors across modalities of perception and production, the flexibility of such behaviors, as well as the degree of intentional control over them. Within-species comparisons of rhythmic abilities exhibited across modalities and cross-species comparisons focusing on single modalities will allow us to test long-held assumptions (e.g., human auditory superiority) and more recent hypotheses (e.g., the **beacon effect** hypothesis, the **signal conservation** hypothesis, the **incidental sounds of locomotion** hypothesis, the **acoustic advantage** hypothesis) on the nature of musical rhythms and its underlying mechanisms.

4.3. A diversity of functional contexts – moving beyond courtship and territorial defense

As reviewed in section 2, rhythmic behavior has been associated with both individual and social benefits in humans, yet it remains unclear which of these have an ancestral relevance for the evolution of musicality. Interestingly, human music appears to exhibit cross-cultural form-function regularities. Independent of their cultural background, people reliably match lullabies, dance, love, and healing songs from across the world with their functional context (Mehr et al., 2018). Such functional categorization is partly driven by rhythmic features related to emotional valence, or levels of arousal conveyed by a given musical piece. For example, dance songs are characterized by higher rhythmic complexity, a steadier beat and a higher level of arousal and pleasantness compared to lullabies. This suggests the presence of a systematic relationship between musical form, its affective effects and functional uses, with potentially ancestral origins that are rooted in behavioral (vocal, gestural) expressions of emotion.

The comparative literature reviewed above suggests that rhythmic structure may be common in animal signals. Preliminary evidence further shows that the mammalian brain automatically attunes its oscillations to exogenous rhythms (e.g., Bartolo et al., 2014; Noda et al., 2017) and detects on-beat emphasis (Rajendran et al., 2020), which suggests a preparedness for rhythm detection, whose functional relevance, however, remains elusive. A systematic study of the rhythmic behaviors exhibited by a diversity of species, and the functional contexts in which they occur (as summarized in Figure 3), is thus likely to provide valuable data both for understanding the evolution of human musicality, as well as the species-specific functions of various forms of rhythmic behaviors, and whether functional trends can be delineated across species. This research should also consider modalities of perception (e.g., auditory, visual, as outlined in 4.2.) and the full spectrum of rhythmic abilities (as outlined in 4.1), and provide detailed descriptions of the structure of rhythmic behavior, as a crucial first step in understanding their ecology and evolution (for a similar point see Caselli et al., 2014).

Generally, there is a lack of studies focused on the potential significance of rhythmic signals beyond the contexts of mating and territorial defense (as reviewed in 2.1), as well as a bias towards studying rhythmically structured behavior in the vocal channel only (as reviewed in 4.2). A handful of studies suggests, however, that rhythmic signals may occur in other contexts as well, although these focus on vocal behaviors. For example, rhythmic vocalizations with a socialization function have been described in gelada monkeys (*Theropithecus gelada*: Richman, 1978) and in long-finned pilot whales (*Globicephala melas:* Zwamborn & Whitehead, 2016). The geladas appear to be able to synchronize the tempo and rhythm of their own vocalizations with the vocalizations emitted by a conspecific, while the pilot whales emit their rhythmic vocalizations primarily when gathered in larger groups at sea surface.

Insert Figure 3 about here

Rhythmic vocalizations in aggressive contexts have been described, for example, in bats and dolphins. In the greater sac-winged bat, these take the form of male-emitted appeasement signals (Fernandez & Knörnschild, 2017). In Atlantic spotted dolphins (*Stenella frontalis*) and bottlenose dolphins (*Tursiops truncatus*) two respectively three, distinct rhythmic vocalizations have been documented in such contexts. Dolphins emit these calls communally, during both intra- and interspecific aggressive encounters, where they also perform bodily displays, although the temporal structure of these displays remains to be studied in detail (Herzing & Dolphin, 2015).

Recently, our research group set out to examine the presence, range, form and functional significance of rhythmic behaviors in chimpanzees. Our preliminary reports suggest that rhythmic behaviors are frequent in chimpanzees, and may occur in a variety of contexts, including initiation of social interaction, social play, solitary play, dominance display, territorial group display, affiliative interaction, and courtship (Sauciuc et al., 2019a). We documented considerable flexibility with respect to form and function, as these behaviors can be vocal or gestural, can take on a diversity of forms, involving various effectors (whole body, head, hand, arm, foot, or combinations of these), and, generally, are not rigidly bound to a specific context (Sauciuc et al., 2019b). Rhythmic bouncing, for example, can occur as a signal for initiation of social interaction, as well as during social play, courtship or as an appeasement gesture. In addition, rhythmic synchronization has been also documented for instrumental behaviors, in a social learning context, where a chimpanzee infant mimicked the hammering movements of more experienced individuals (Fuhrmann et al., 2014). To date, such diversity and functional flexibility of rhythmic behaviors has only been documented in one other species: humans. It is not clear, however, if chimpanzees are indeed exceptional in this respect, or if this is the result of a comprehensive lack of data from other species. This further underscores the significance of the point raised in this section, i.e., the need for the systematic study of the rhythmic behaviors exhibited by a diversity of species in the functional contexts in which they occur.

In chimpanzees, rhythmic behaviors have previously been associated with stress-induced stereotypical movement, and, thus, with high levels of arousal (Chamove, 1989; Lopresti-Goodman et al., 2013). In the majority of the cases that we documented, piloerection was absent, thus suggesting that these rhythmic behaviors were not a mere expression of arousal. In addition, rhythmic behaviors co-occurred frequently with the relaxed open-mouth face, which is a correlate of states of relaxation and positive emotion (Parr & Waller, 2006). In chimpanzees, both males and females engaged in such rhythmic behaviors, while for the great majority of other species where rhythmic output has been described – almost exclusively in the form of vocalizations – this phenomenon is typically restricted to males (as reviewed in sections 2.1. and 4.1). Based on extant data, the vocalizations of some primate species – such as titi monkeys, tarsiers and indris – could be an exception to this male bias (Caselli et al., 2014; Clink et al., 2020; de Gregorio et al., 2021; Gamba et al., 2019). While we found that male chimpanzees performed rhythmic behaviors more often than female chimpanzees, the social efficacy of these behaviors did not seem to differ between sexes, displaying similar levels of social influence on conspecifics (Sauciuc et al., 2019b).

5. Conclusions: The fallacy of "one and only-theories"

During the last two decades, there has been an explosion of theoretical proposals aimed at explaining why and how humans have evolved into musical beings. Most often, these proposals focus on pinpointing *the one* function or *the one* functional context that had precedence in ancestral times, thus being the primary driving force in the evolution of musicality. Such endeavors, however, are invariably predestined to result in a chicken-or-the-egg debate, as complex traits, such as musicality, are inherently componential, and thus, their evolution is irreducible to the evolution of a single component with a single adaptive function. Even when looking at single components, however, multiple factors combine into making something adaptive, in the sense that traits and behaviors emerge ontogenetically from gene-brain-body-environment feedback loops, and it is therefore impossible to propose a 'one-and-only' origin story behind any evolved thing. This is perhaps best resolved by not asking "what has musicality evolved for", which easily leads to the idea of a singular reason, but "what makes and has made each component of musicality adaptive", which can be many things.

The review of proximate mechanisms hypothesized to be involved in the perception and production of musical rhythm demonstrates this point, suggesting that musicality has emerged in concert with, and not independently of, other communication systems as well as a plethora of socio-cognitive abilities. Some of these appear to be widespread in the animal kingdom (e.g., predictive coding, neural resonance, emotional contagion), others appear to have evolved repeatedly in phylogenetically distant species (e.g., vocal learning), while others appear to be human- (or Hominid-) specific (e.g., shared intentionality) or present in a few highly intelligent species (e.g., episodic foresight). How these prerequisites for evolving music-like rhythmic behaviors relate to one another to contribute to the emergence of musical rhythm is still to be investigated. The findings accumulated up to date (many of which reviewed in this paper) are far from providing a coherent picture of the evolution of musicality and, in particular, the rhythmical abilities that make it possible. They succeed, nevertheless, at revealing that even the most focused and seemingly straightforward questions beg highly complex answers. Since the evolutionary study of musicality is inherently multidisciplinary, seeking answers to these questions will require interdisciplinary collaboration, and research that will both *zoom in* on specific mechanisms, contexts and species, and *zoom out* from detailed analyses to broader views.

In this paper, we proposed a research agenda centered on topics that have been so far neglected in the comparative study of rhythmic abilities, but that are highly relevant for understanding where musicality comes from. In particular, we call for a widened focus of research on rhythmic cognition in nonhuman species, so as to include additional rhythmic abilities besides entrainment (4.1), other modalities of rhythmic expression besides the vocal modality (4.2), and a systematic focus on the functional contexts in which rhythmic abilities are spontaneously employed (4.3). Broadening the scope of comparative research to include rhythmic abilities (e.g., rhythmic engagement, isochrony detection, beat perception) that are considered precursors to accurate entrainment (at least in human ontogeny) is necessary if the purpose is to track the evolutionary *origins* of advanced rhythmic abilities. Expanding research to other modalities than auditory perception and vocal communication is also important. In particular, there is a need for systematic research on the perception and production of bodily rhythms, since rhythmic movement is an essential feature of musicality, and given suggestions that musicality may have originated as bodily-coupled entrainment in the absence of an auditory stimulus. Finally, systematically mapping the contexts in which animal rhythmic signals occur spontaneously is crucial to understanding the functional significance of rhythmic communication, its efficacy and the factors that underlie and influence it. The insights gained from such research will, in turn, inform theories about the likely adaptive significance of musical rhythm. In this context, special attention should be given to rhythmic communication through bodily movement, as this is conducive to rhythmic engagement, which likely promotes and reaffirms social affiliation.

Extant comparative data suggests that most of the mechanisms that enable musical rhythm may not be musicality specific. Some of these mechanisms appear to be widespread across species, and, thus, phylogenetically old. Yet, detailed analyses and cross-species comparisons are needed to determine the extent to which the neural networks and psychobiological mechanisms that mediate the perception, production and appreciation of musical rhythm in humans have similar counterparts in other species. A few high-level cognitive and socio-cognitive abilities that are relevant to the evolution of musical rhythm (e.g., relational processing, causal reasoning, enhanced memory and foresight, shared intentionality, imitation and pantomime) appear to be rare in the animal kingdom, being exclusive to highly intelligent species, such as corvids and hominids, or restricted to hominids only. The emergence of these high-level abilities is generally related to adaptive pressures posed by complex social organization and complex foraging strategies. Seemingly, however, at some point in the evolution of our species, these abilities have been co-opted into rhythmical, communal displays. Additional traits that, so far, do not seem to have a counterpart in the animal kingdom, and which are related to human creativity, narrativity, aesthetics and linguistic communication, have further shaped such behaviors into what we today call music and dance.

Author contributions

Gabriela-Alina Sauciuc conceived and drafted the paper. Tomas Persson and Elainie Madsen have critically reviewed it.

Acknowledgments

This work has been supported by grant 2017:0942:1 from Riksbankens Jubileumsfond (The Bank of Sweden Tercentenary Foundation). We are grateful to the editor and the reviewers for their constructive criticism and inspiring comments.

Conflicts of interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical statement. Not applicable.

Data availability statement. Not applicable.

References

Aalbers, S., Fusar-Poli, L., Freeman, R., Spreen, M., Ket, J.C., Vink, A., Maratos, A., Crawford, M., Chen, X., & Gold, C. (2017). Music therapy for depression. *The Cochrane database of systematic reviews, 11*, CD004517. https://doi.org/10.1002/14651858.CD004517.pub3

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and language*, *162*, 46–59. https://doi.org/10.1016/j.bandl.2016.08.002

Akerlund, L., & Gramming, P. (1994). Average loudness level, mean fundamental frequency, and subglottal pressure: comparison between female singers and nonsingers. *Journal of voice: official journal of the Voice Foundation*, *8*(3), 263–270. https://doi.org/10.1016/s0892-1997(05)80298-x

Alcaro, A., Huber, R., & Panksepp, J. (2007). Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. *Brain research reviews*, *56*(2), 283–321. https://doi.org/10.1016/j.brainresrev.2007.07.014

Alward, B.A., Rouse, M.L., Balthazart, J., & Ball, G. (2017). Testosterone regulates birdsong in an anatomically specific manner. *Animal Behaviour, 124*, 291-298. https://doi.org/10.1016/j.anbehav.2016.09.013

Alworth, L. C., & Buerkle, S. C. (2013). The effects of music on animal physiology, behavior and welfare. *Lab animal*, *42*(2), 54–61. https://doi.org/10.1038/laban.162

Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism – how to carry out an exaptationist program. *The Behavioral and brain sciences*, *25*(4), 489–553. https://doi.org/10.1017/s0140525x02000092

Anshel, A., & Kipper, D. A. (1988). The influence of group singing on trust and cooperation. *Journal of Music Therapy, 25(3)*, 145–155. https://doi.org/10.1093/jmt/25.3.145.

Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., & De Volder, A.G. (2017). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *European Journal of Neuroscience, 45(11),* 1439-1450. https://doi.org/10.1111/ejn.13349

Arom, S. (2000). *Prolegomena to a biomusicology*. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of music* (pp. 27–29). Cambridge, Mass: The MIT Press.

Bachner-Melman, R., Dina, C., Zohar, A., Constantini, N., Lerer, E., Hoch, S., Sella, S., Nemanov, L., Gritsenko, I., Lichtenberg, P., Granot, R., & Ebstein, R. (2005). AVPR1a and SLC6A4 Gene Polymorphisms Are Associated with Creative Dance Performance. *PLoS Genetics, 1(3),* e1008135. https://doi.org/10.1371/journal.pgen.0010042

Bachorowski, J.-A., & Owren, M. J. (2003). *Sounds of Emotion: Production and Perception of Affect-Related Vocal Acoustics.*In P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds.), *Annals of the New York Academy of Sciences: Vol. 1000. Emotions inside out: 130 years after Darwin's: The expression of the emotions in man and animals* (p. 244–265). New York Academy of Sciences. https://doi.org/10.1196/annals.1280.012

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental psychobiology*, *41*(4), 352–363. https://doi.org/10.1002/dev.10049

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190–201. https://doi.org/10.1037//0012-1649.36.2.190

Baribeau, D. A., & Anagnostou, E. (2015). Oxytocin and vasopressin: linking pituitary neuropeptides and their receptors to social neurocircuits. *Frontiers in neuroscience*, *9*, 335. https://doi.org/10.3389/fnins.2015.00335

Bartolo, R., Prado, L., & Merchant, H. (2014). Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping. *The Journal of* *neuroscience : the official journal of the Society for Neuroscience*, *34*(11), 3910–3923. https://doi.org/10.1523/JNEUROSCI.2679-13.2014

Baskerville, T. A., & Douglas, A. J. (2010). Dopamine and oxytocin interactions underlying behaviors: potential contributions to behavioral disorders. *CNS neuroscience & therapeutics*, *16*(3), e92–e123. https://doi.org/10.1111/j.1755-5949.2010.00154.x

Bigelow, J., & Poremba, A. (2014). Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. *PloS one*, *9*(2), e89914. https://doi.org/10.1371/journal.pone.0089914

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Bornstein, M. H., Putnick, D. L., Rigo, P., Esposito, G., Swain, J. E., Suwalsky, J., Su, X., Du, X., Zhang, K., Cote, L. R., De Pisapia, N., & Venuti, P. (2017). Neurobiology of culturally common maternal responses to infant cry. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(45), E9465–E9473. https://doi.org/10.1073/pnas.1712022114

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PloS one*, *9*(5), e97467. https://doi.org/10.1371/journal.pone.0097467

Bouwer, F. L., Nityananda, V., Rouse, A. A., & Ten Cate, C. (2021). Rhythmic abilities in humans and non-human animals: a review and recommendations from a methodological perspective. *Philosophical transactions of the Royal Society of London. Series B,* *Biological sciences*, *376*(1835), 20200335. https://doi.org/10.1098/rstb.2020.0335

Bowling, D.L., Gahr, J., Ancochea, P., Hoeschele, M., Canoine, V., Fusani, L., Fitch, W.T. (2022). Endogenous oxytocin, cortisol, and testosterone in response to group singing. *Hormones and Behaviour*, *139*, 105105. https://doi.org/10.1016/j.yhbeh.2021.105105

Brandt, P.A. (2009). *Music and how we became human: a view from cognitive semiotics – exploring imaginative hypotheses*. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 31–44). Oxford University Press.

Brodal, H. P., Osnes, B., & Specht, K. (2017). Listening to Rhythmic Music Reduces Connectivity within the Basal Ganglia and the Reward System. *Frontiers in neuroscience*, *11*, 153. https://doi.org/10.3389/fnins.2017.00153

Brown, S. (2000a). *Evolutionary models of music: From sexual selection to group selection*. In Tonneau, F., Thompson, N. S. (eds.), *Perspectives in ethology. 13: Behavior, evolution and culture* (pp. 231–81). New York: Plenum Publishers. https://doi.org/10.1007/978-1-4615-1221-9_9

Brown, S. (2000b). *The ’musilanguage’ model of music evolution*. In *The Origins of Music*, eds N. L. Wallin, B. Merker and S. Brown (Cambridge, MA: MIT Press), 271–300. https://doi.org/10.7551/mitpress/5190.001.0001

Brown, S. (2007). Contagious heterophony: A new theory about the origins of music. *Musicae Scientiae, 11(1)*, 26 - 3. https://doi.org/10.1177/102986490701100101

Brown S. (2017). A Joint Prosodic Origin of Language and Music. *Frontiers in psychology*, *8*, 1894. https://doi.org/10.3389/fpsyg.2017.01894

Brown, S., & Jordania, J. (2013). Universals in the world’s musics. *Psychology of Music, 41(2)*, 229 - 248. Https://doi.org/10.1177/0305735611425896

Brumm, H., & Slater, P. (2006). Ambient noise, motor fatigue, and serial redundancy in chaffinch song. *Behavioral Ecology and Sociobiology, 60*, 475-481. https://doi.org/10.1007/s00265-006-0188-y

Bryant G. A. (2013). Animal signals and emotion in music: coordinating affect across groups. *Frontiers in psychology*, *4*, 990. https://doi.org/10.3389/fpsyg.2013.00990

Burchardt, L. S., Norton, P., Behr, O., Scharff, C., & Knörnschild, M. (2019). General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*. *Royal Society open science*, *6*(1), 181076. https://doi.org/10.1098/rsos.181076

Carson, D. S., Guastella, A. J., Taylor, E. R., & McGregor, I. S. (2013). A brief history of oxytocin and its role in modulating psychostimulant effects. *Journal of psychopharmacology (Oxford, England)*, *27*(3), 231–247. https://doi.org/10.1177/0269881112473788

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., & Setz, E. Z. (2014). Vocal behavior of black-fronted titi monkeys (Callicebus nigrifrons): acoustic properties and behavioral contexts of loud calls. *American journal of primatology*, *76*(8), 788–800. https://doi.org/10.1002/ajp.22270

Celma-Miralles, A., & Toro, J. M. (2020). Discrimination of temporal regularity in rats (Rattus norvegicus) and humans (Homo sapiens). *Journal of comparative psychology (Washington, D.C. : 1983)*, *134*(1), 3–10. https://doi.org/10.1037/com0000202

Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E. E., Dabelsteen, T., Pakkenberg, B., Bertelsen, M. F., Dorrestein, G. M., Brauth, S. E., Durand, S. E., & Jarvis, E. D. (2015). Core and Shell Song Systems Unique to the Parrot Brain. *PloS one*, *10*(6), e0118496.

https://doi.org/10.1371/journal.pone.0118496

Chamove, A. (1989). Environmental Enrichment: A Review. *Animal Technology*, *40*, 3.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436.

Chang, S. W., Brent, L. J., Adams, G. K., Klein, J. T., Pearson, J. M., Watson, K. K., & Platt, M. L. (2013). Neuroethology of primate social behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*(Suppl 2), 10387–10394. https://doi.org/10.1073/pnas.1301213110

Chaudhry S.R., Bhimji S.S. (2018). Biochemistry, Endorphin. StatPearls. StatPearls Publishing. PMID 29262177. Retrieved 20 February 2019.

Chu, K., & Harcourt, P. (1986). Behavioral correlations with aberrant patterns in humpback whale songs. *Behavioral Ecology and Sociobiology, 19*(5), 309–312. https://doi.org/10.1007/BF00295703

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental science*, *17*(6), 1003–1011. https://doi.org/10.1111/desc.12193

Cirelli, L. K., Wan, S. J., Spinelli, C., & Trainor, L. J. (2017). Effects of interpersonal movement synchrony on infant helping behaviors: Is music necessary? *Music Perception, 34*(3), 319–326. https://doi.org/10.1525/mp.2017.34.3.319

Clink, D. J., Tasirin, J. S., & Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Current zoology*, *66*(2), 173–186. https://doi.org/10.1093/cz/zoz035

Cohen, E. E., Ejsmond-Frey, R., Knight, N., & Dunbar, R. I. (2010). Rowers' high: behavioural synchrony is correlated with elevated pain thresholds. *Biology letters*, *6*(1), 106–108. https://doi.org/10.1098/rsbl.2009.0670

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology, 127*(4), 412–427. https://doi.org/10.1037/a0032345

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild Chimpanzees Produce Group-Specific Calls: A Case for Vocal Learning? *Ethology, 110*(3), 221–243. https://doi.org/10.1111/j.1439-0310.2004.00968.x

Crockford, C., Wittig, R. M., Langergraber, K., Ziegler, T. E., Zuberbühler, K., & Deschner, T. (2013). Urinary oxytocin and social bonding in related and unrelated wild chimpanzees. *Proceedings. Biological sciences*, *280*(1755), 20122765. https://doi.org/10.1098/rspb.2012.2765

Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization draws others' attention: a playback experiment with wild chimpanzees. *Animal cognition*, *18*(3), 581–591. https://doi.org/10.1007/s10071-014-0827-z

Cross I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*, *930*, 28–42. https://doi.org/10.1111/j.1749-6632.2001.tb05723.x

Cross, I. (2003). *Music, cognition, culture and evolution*. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 42–56). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525202.003.0004

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: J. Murray.

Davila-Ross, M., Allcock, B., Thomas, C., & Bard, K. A. (2011). Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion (Washington, D.C.)*, *11*(5), 1013–1020. https://doi.org/10.1037/a0022594

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current biology : CB*, *31*(20), R1379–R1380. https://doi.org/10.1016/j.cub.2021.09.032

de la Mora, D. M., Nespor, M., & Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic-trochaic law?. *Attention, perception & psychophysics*, *75*(1), 92–100. https://doi.org/10.3758/s13414-012-0371-3

de Witte, M., Spruit, A., van Hooren, S., Moonen, X., & Stams, G. J. (2020). Effects of music interventions on stress-related outcomes: a systematic review and two meta-analyses. *Health psychology review*, *14*(2), 294–324. https://doi.org/10.1080/17437199.2019.1627897

Dissanayake, E. (2000). *Antecedents of the temporal arts in early mother–infant interaction.* In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 389–410). The MIT Press.

Dissanayake, E. (2009a). *Root, leaf, blossom, or bole: Concerning the origin and adaptive function of music.* In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (p. 17–30). Oxford University Press.

Dissanayake, E. (2009b). *Bodies swayed to music: The temporal arts as integral to ceremonial ritual.* In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (p. 533–544). Oxford University Press.

Donnet, S., Bartolo, R., Fernandes, J. M., Cunha, J. P., Prado, L., & Merchant, H. (2014).

Monkeys time their pauses of movement and not their movement-kinematics during a synchronization-continuation rhythmic task. *Journal of neurophysiology*, *111*(10), 2138–2149. https://doi.org/10.1152/jn.00802.2013

Dunbar, R. I. M. (2004). *Language, music, and laughter in evolutionary perspective*. In: Oller, D. K. & Griebel, U. (eds.) *Evolution of communication systems: a comparative approach*. (pp. 257–274). MIT Press.

Dunbar, R. I. M. (2012). *On the evolutionary function of song and dance*. In Bannan, N. (ed.) *Music, Language and Human Evolution*. (pp. 201–214). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199227341.003.0008.

Dunbar, R., Kaskatis, K., Macdonald, I., & Barra, V. (2012). Performance of Music Elevates Pain Threshold and Positive Affect: Implications for the Evolutionary Function of Music. *Evolutionary Psychology, 10(4)*, 688-702. https://doi.org/10.1177/147470491201000403

Eckenweber, M., & Knörnschild, M. (2013). Social influences on territorial signaling in male greater sac-winged bats. *Behavioral Ecology and Sociobiology, 67*, 639-648. Https://doi.org/10.1007/s00265-013-1483-z

Edelman, L. L., & Harring, K. E. (2015). Music and social bonding: The role of non-diegetic music and synchrony on perceptions of videotaped walkers. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 34*(4), 613–620. https://doi.org/10.1007/s12144-014-9273-y

Eriksen, N., Miller, L., Tougaard, J., & Helweg, D. (2005). Cultural change in the songs of humpback whales (Megaptera novaeangliae) from Tonga. *Behaviour, 142(3)*, 305-328. https://doi.org/10.1163/1568539053778283

Falk D. (2004). Prelinguistic evolution in early hominins: whence motherese?. *The Behavioral* *and brain sciences*, *27*(4), 491–583. https://doi.org/10.1017/s0140525x04000111

Fan, P. F., Ma, C. Y., Garber, P. A., Zhang, W., Fei, H. L., & Xiao, W. (2016). Rhythmic displays of female gibbons offer insight into the origin of dance. *Scientific reports*, *6*, 34606. https://doi.org/10.1038/srep34606

Fedurek, P., Machanda, Z. P., Schel, A. M., & Slocombe, K. E. (2013). Pant hoot chorusing and social bonds in male chimpanzees. *Animal Behaviour, 86*(1), 189–196. https://doi.org/10.1016/j.anbehav.2013.05.010

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H., & Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PloS one*, *3*(3), e1768. https://doi.org/10.1371/journal.pone.0001768

Feldman, R. (2007). Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes *Current Directions in Psychological Science, 16(6)*, 340 - 345. https://doi.org/10.1111/j.1467-8721.2007.00532.x

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R. M., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(9), 3793–3798. https://doi.org/10.1073/pnas.1811878116

Fessler, D. M. T., & Holbrook, C. (2016). Synchronized behavior increases assessments of the formidability and cohesion of coalitions. *Evolution and Human Behavior, 37*(6), 502–509. https://doi.org/10.1016/j.evolhumbehav.2016.05.003

Festante, F., Rayson, H., Paukner, A., Kaburu, S., Toschi, G., Fox, N. A., & Ferrari, P. F. (2021).

Oxytocin promotes prosocial behavior and related neural responses in infant macaques at-risk for compromised social development. *Developmental cognitive neuroscience*, *48*, 100950. https://doi.org/10.1016/j.dcn.2021.100950

Fernandez, A.A., & Knörnschild, M. (2017). Isolation calls of the bat *Saccopteryx bilineata* encode multiple messages. *Animal Behavior and Cognition, 4(2)*, 169-186. https://doi.org/10.12966/abc.04.05.2017

Fishbein, A. R., Idsardi, W. J., Ball, G. F., & Dooling, R. J. (2020). Sound sequences in birdsong: how much do birds really care?. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *375*(1789), 20190044. https://doi.org/10.1098/rstb.2019.0044

Fitch W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition*, *100*(1), 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Fitch, W.T. 2011. *The biology and evolution of rhythm: unravelling a paradox*. In, P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (p. 73-95). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553426.003.0009

Fitch, W.T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience, 7,* 68. https://doi.org/10.3389/fnsys.2013.00068

Fitch, W.T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1664),* 20140091. https://doi.org/10.1098/rstb.2014.0091

Fitch, W.T. (2016). Dance, Music, Meter and Groove: A Forgotten Partnership. *Frontiers in*

Human Neuroscience, 10, 64. https://doi.org/10.3389/fnhum.2016.00064

Frazer, L.N., & Mercado, E. (2000). A sonar model for humpback whale song. *IEEE Journal of Oceanic Engineering, 25*, 160-182. https://doi.org/10.1109/48.820748

Freeman, W. (2000). *A neurobiological role of music in social bonding.* In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 411–424). The MIT Press.

Fritz, T., & Koelsch, S. (2013). *Acoustically mediated emotional contagion as a cross-species homology underlying music processing*. In Altenmüller, E., Schmidt, S., & Zimmermann, E. (Eds.). *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (p. 300-312). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.003.0018

Fuhrmann, D., Ravignani, A., Marshall-Pescini, S., & Whiten, A. (2014). Synchrony and motor mimicking in chimpanzee observational learning. *Scientific reports*, *4*, 5283. https://doi.org/10.1038/srep05283

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PloS one*, *9*(5), e97680. https://doi.org/10.1371/journal.pone.0097680

Gamba, M., Valente, D., Gregorio, C.D., Zanoli, A., Randrianarison, R., Torti, V., & Giacoma, C. (2019). Song structure and sex specific features in the indris. *Journal of the Acoustical Society of America, 145*, 1919-1919. https://doi.org/10.1121/2.0001021

Gámez, J., Yc, K., Ayala, Y. A., Dotov, D., Prado, L., & Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Annals of the New York Academy of Sciences*, 10.1111/nyas.13671. Advance online publication. https://doi.org/10.1111/nyas.13671

Gan, L., Huang, Y., Zhou, L., Qian, C., & Wu, X. (2015). Synchronization to a bouncing ball with a realistic motion trajectory. *Scientific reports*, *5*, 11974. https://doi.org/10.1038/srep11974

Gebauer, L., Witek, M., Hansen, N. C., Thomas, J., Konvalinka, I., & Vuust, P. (2016). Oxytocin improves synchronisation in leader-follower interaction. *Scientific reports*, *6*, 38416. https://doi.org/10.1038/srep38416

Geissmann, T. (2000). *Gibbon songs and human music from an evolutionary perspective.* In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 103–123). The MIT Press.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in cognitive sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Ghazanfar A. A. (2009). The multisensory roles for auditory cortex in primate vocal communication. *Hearing research*, *258*(1-2), 113–120. https://doi.org/10.1016/j.heares.2009.04.003

Gilmore, S. A., & Russo, F. A. (2021). Neural and Behavioral Evidence for Vibrotactile Beat Perception and Bimodal Enhancement. *Journal of cognitive neuroscience*, *33(4)*, 635–650. https://doi.org/10.1162/jocn_a_01673

Gordon, R. L., Ravignani, A., Hyland Bruno, J., Robinson, C. M., Scartozzi, A., Embalabala, R., Niarchou, M., 23andMe Research Team, Cox, N. J., & Creanza, N. (2021). Linking the genomic signatures of human beat synchronization and learned song in birds. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *376*(1835), 20200329. https://doi.org/10.1098/rstb.2020.0329

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of cognitive neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Grammer, K., & Eibl-Eibesfeldt, I. (1990). *The ritualization of laughter*. In W. Koch (Ed.), *Naturlichkeit der sprache und der kultur: Acta colloquii* (pp. 192-214). Bochum: Brockmeyer.

Granot, R., Frankel, Y., Gritsenko, V., Lerer, E., Gritsenko, I., Bachner-Melman, R., S. Israel, & Ebstein, R. (2007). Provisional evidence that the arginine vasopressin 1a receptor gene is associated with musical memory. *Evolution and Human Behavior, 28(5)*, 313-318. https://doi.org/10.1016/j.evolhumbehav.2007.05.003.

Grape, C., Sandgren, M., Hansson, L. O., Ericson, M., & Theorell, T. (2003). Does singing promote well-being?: An empirical study of professional and amateur singers during a singing lesson. *Integrative physiological and behavioral science : the official journal of the Pavlovian Society*, *38*(1), 65–74. https://doi.org/10.1007/BF02734261

Greenfield, M. (2006). Mechanisms and evolution of chorusing interactions in acoustic insects and anurans. *The Journal of the Acoustical Society of America, 119*, 3222. https://doi.org/10.1121/1.4785923

Greenfield, M. D., & Schul, J. (2008). Mechanisms and evolution of synchronous chorusing: emergent properties and adaptive functions in Neoconocephalus katydids (Orthoptera: Tettigoniidae). *Journal of comparative psychology (Washington, D.C. : 1983)*, *122*(3), 289–297. https://doi.org/10.1037/0735-7036.122.3.289

Greenfield, M.D., Marin-Cudraz, T., & Party, V. (2017). Evolution of synchronies in insect choruses. *Biological Journal of The Linnean Society, 122(3)*, 487-504. https://doi.org/10.1093/biolinnean/blx096

Greenfield, M.D., Aihara, I., Amichay, G., Anichini, M., & Nityananda, V. (2021). Rhythm interaction in animal groups: selective attention in communication networks. *Philosophical Transactions of the Royal Society B, 376(1835)*, 20200338. https://doi.org/10.1098/rstb.2020.0338

Griffiths, N. K., & Reay, J. L. (2018). The relative importance of aural and visual information in the evaluation of Western canon music performance by musicians and nonmusicians. *Music Perception, 35*(3), 364–375. https://doi.org/10.1525/mp.2018.35.3.364

Grillner, S., & Robertson, B. (2016). The Basal Ganglia Over 500 Million Years. *Current biology : CB*, *26*(20), R1088–R1100. https://doi.org/10.1016/j.cub.2016.06.041

Gruber, T. (2014). Wild-born orangutans (Pongo abelii) engage in triadic interactions during play. *International Journal of Primatology, 35*(2), 411–424. https://doi.org/10.1007/s10764-013-9745-1

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human nature (Hawthorne, N.Y.)*, *14*(1), 21–51. https://doi.org/10.1007/s12110-003-1015-z

Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae, Special Issue,* 291–320. https://doi.org/10.1177/1029864909013002131

Hall, K., Oram, M. W., Campbell, M. W., Eppley, T. M., Byrne, R. W., & de Waal, F. B. (2017). Chimpanzee uses manipulative gaze cues to conceal and reveal information to foraging competitor. *American journal of primatology*, *79*(3), 1–11. https://doi.org/10.1002/ajp.22622

Handel, S., Todd, S. K., & Zoidis, A. M. (2009). Rhythmic structure in humpback whale (Megaptera novaeangliae) songs: preliminary implications for song production and perception. *The Journal of the Acoustical Society of America*, *125*(6), EL225–EL230. https://doi.org/10.1121/1.3124712

Handel, S., Todd, S. K., & Zoidis, A. M. (2012). Hierarchical and rhythmic organization in the songs of humpback whales (Megaptera novaeangliae). *Bioacoustics, 21*(2), 141–156. https://doi.org/10.1080/09524622.2012.668324

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: implications for musical structure learning. *Cognitive psychology*, *50*(4), 354–377. https://doi.org/10.1016/j.cogpsych.2004.09.00

Hannon, E. E., Vanden Bosch der Nederlanden, C. M., & Tichko, P. (2012). Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Annals of the New York Academy of Sciences*, *1252*, 92–99. https://doi.org/10.1111/j.1749-6632.2012.06466.x

Hannon, E. E., Schachner, A., & Nave-Blodgett, J. E. (2017). Babies know bad dancing when they see it: Older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays. *Journal of experimental child psychology*, *159*, 159–174. https://doi.org/10.1016/j.jecp.2017.01.006

Hannon, E. E., Crittenden, A. N., Snyder, J. S., & Nave, K. M. (2021). An evolutionary theory of music needs to care about developmental timing. *The Behavioral and brain sciences*, *44*, e74. https://doi.org/10.1017/S0140525X20001168

Hartbauer, M., & Römer, H. (2016). Rhythm Generation and Rhythm Perception in Insects: The Evolution of Synchronous Choruses. *Frontiers in neuroscience*, *10*, 223.

https://doi.org/10.3389/fnins.2016.00223

Hartbauer, M., Haitzinger, L., Kainz, M., & Römer, H. (2014). Competition and cooperation in a synchronous bushcricket chorus. *Royal Society open science*, *1(2)*, 140167. https://doi.org/10.1098/rsos.140167

Hasegawa, A., Okanoya, K., Hasegawa, T., & Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific reports*, *1*, 120. https://doi.org/10.1038/srep00120

Hattori, Y., & Tomonaga, M. (2020). Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proceedings of the National Academy of Sciences of the United States of America*, *117*(2), 936–942. https://doi.org/10.1073/pnas.1910318116

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports, 3,* 1566. https://doi.org/10.1038/srep01566

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: testing the iambic/trochaic law. *Perception & psychophysics*, *69*(1), 113–122. https://doi.org/10.3758/bf03194458

Heimbauer, L. A., Beran, M. J., & Owren, M. J. (2021). A chimpanzee recognizes varied acoustical versions of sine-wave and noise-vocoded speech. *Animal cognition*, *24*(4), 843–854. https://doi.org/10.1007/s10071-021-01478-4

Heinsohn, R., Zdenek, C., Cunningham, R., Endler, J., & Langmore, N. (2017). Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. *Science Advances, 3(6)*, e1602399. https://doi.org/10.1126/sciadv.1602399

Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., Scheffler, K., Wetzel, S., Seifritz, E., & Cattapan-Ludewig, K. (2010). Musical training induces functional plasticity in human hippocampus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(4), 1377–1384. https://doi.org/10.1523/JNEUROSCI.4513-09.2010

Herzing, D., & Dolphin, W. (2015). Synchronous and Rhythmic Vocalizations and Correlated Underwater Behavior of Free-ranging Atlantic Spotted Dolphins (Stenella frontalis) and Bottlenose Dolphins (Tursiops truncatus) in the Bahamas. *Animal Behavior and Cognition, 2(1)*, 14-29. Https://doi.org/10.12966/abc.02.02.2015

Hoeschele, M., & Bowling, D. L. (2016). Sex Differences in Rhythmic Preferences in the Budgerigar (*Melopsittacus undulatus*): A Comparative Study with Humans. *Frontiers in psychology*, *7*, 1543. https://doi.org/10.3389/fpsyg.2016.01543

Hoeschele, M., & Fitch, W. T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Animal cognition*, *19*(3), 643–654. https://doi.org/10.1007/s10071-016-0968-3

Honing H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*, 85–91. https://doi.org/10.1111/j.1749-6632.2011.06402.x

Honing, H. (2019). *The Evolving Animal Orchestra. In search of what makes us musical.* MIT Press, Cambridge (MA). https://doi.org/10.7551/mitpress/11839.001.0001

Honing, H., & Ploeger, A. (2012). Cognition and the evolution of music: pitfalls and prospects. *Topics in cognitive science*, *4*(4), 513–524. https://doi.org/10.1111/j.1756-8765.2012.01210.x

Honing, H., Cate, C.T., Peretz, I., & Trehub, S. (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B:*

Biological Sciences, 370(1664), 20140088. https://doi.org/10.1098/rstb.2014.0088

Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018). Rhesus Monkeys (*Macaca mulatta*) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. *Frontiers in neuroscience*, *12*, 475. https://doi.org/10.3389/fnins.2018.00475

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27*(6), 949–961. https://doi.org/10.1521/soco.2009.27.6.949

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10383–10388. https://doi.org/10.1073/pnas.1402039111

Hribar, A. (2011). Comparing Great Apes' and Children's Ability to Reason by Analogy. PhD Thesis, Heidelberg University.

Hulse, S. H., Humpal, J., & Cynx, J. (1984). Processing of rhythmic sound structures by birds. *Annals of the New York Academy of Sciences*, *423*, 407–419. https://doi.org/10.1111/j.1749-6632.1984.tb23449.x

Huron D. (2001). Is music an evolutionary adaptation?. *Annals of the New York Academy of Sciences*, *930*, 43–61. https://doi.org/10.1111/j.1749-6632.2001.tb05724.x

Huron D. (2012). Two challenges in cognitive musicology. *Topics in cognitive science*, *4*(4), 678–684. https://doi.org/10.1111/j.1756-8765.2012.01224.x

Iversen, J. R., & Balasubramaniam, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences, 8,* 175–180. https://doi.org/10.1016/j.cobeha.2016.02.027

Iversen, J., Patel, A., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124 4*, 2263-71. https://doi.org/10.1121/1.2973189

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, *124*(4), 2263–2271. https://doi.org/10.1121/1.2973189

Jacoby, N., & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current biology : CB*, *27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Janata P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral cortex (New York, N.Y. : 1991)*, *19*(11), 2579–2594. https://doi.org/10.1093/cercor/bhp008

Jarvis, E. D., & Mello, C. V. (2000). Molecular mapping of brain areas involved in parrot vocal communication. *The Journal of comparative neurology*, *419*(1), 1–31. https://doi.org/10.1002/(sici)1096-9861(20000327)419:1

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and brain sciences*, *31*(5), 559–621. https://doi.org/10.1017/S0140525X08005293

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception, 23*(1), 1–27. https://doi.org/10.1525/mp.2005.23.1.1

Kello, C. T., Bella, S. D., Médé, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society, Interface*, *14*(135), 20170231. https://doi.org/10.1098/rsif.2017.0231

Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of experimental child psychology*, *102*(3), 299–314. https://doi.org/10.1016/j.jecp.2008.07.005

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior, 31*(5), 354–364. https://doi.org/10.1016/j.evolhumbehav.2010.04.004

Kisilevsky, B. S., & Muir, D. W. (1991). Human fetal and subsequent newborn responses to sound and vibration. *Infant Behavior & Development, 14*(1), 1–26. https://doi.org/10.1016/0163-6383(91)90051-S

Kisilevsky, B. S., Muir, D. W., & Low, J. A. (1992). Maturation of human fetal responses to vibroacoustic stimulation. *Child development*, *63*(6), 1497–1508. https://doi.org/10.2307/1131571

Knörnschild, M., Blüml, S., Steidl, P., Eckenweber, M., & Nagy, M. (2017). Bat songs as acoustic beacons - male territorial songs attract dispersing females. *Scientific reports*, *7*(1), 13918. https://doi.org/10.1038/s41598-017-14434-5

Koelsch S. (2009). A neuroscientific perspective on music therapy. *Annals of the New York Academy of Sciences*, *1169*, 374–384. https://doi.org/10.1111/j.1749-6632.2009.04592.x

Koelsch S. (2010). Towards a neural basis of music-evoked emotions. *Trends in cognitive sciences*, *14*(3), 131–137. https://doi.org/10.1016/j.tics.2010.01.002

Koelsch, S., Offermanns, K., & Franzke, P. (2010). Music in the treatment of affective disorders: An exploratory investigation of a new method for music-therapeutic research. *Music*

Perception, 27(4), 307–316. https://doi.org/10.1525/mp.2010.27.4.307

Koelsch S. (2014). Brain correlates of music-evoked emotions. *Nature reviews. Neuroscience*, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Kogan, N. (1997). Reflections on aesthetics and evolution. *Critical Review, 11(2)*, 193-210. https://doi.org/10.1080/08913819708443453

Kokal, I., Engel, A., Kirschner, S., & Keysers, C. (2011). Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment--if the rhythm comes easily. *PloS one*, *6*(11), e27272. https://doi.org/10.1371/journal.pone.0027272

Koshimori, Y., Strafella, A. P., Valli, M., Sharma, V., Cho, S. S., Houle, S., & Thaut, M. H. (2019). Motor Synchronization to Rhythmic Auditory Stimulation (RAS) Attenuates Dopaminergic Responses in Ventral Striatum in Young Healthy Adults: [^{11}C]-(+)-PHNO PET Study. *Frontiers in neuroscience*, *13*, 106. https://doi.org/10.3389/fnins.2019.00106

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences, 22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science (New York, N.Y.)*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Lakens, D. (2010). Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology, 46(5)*, 701-708. https://doi.org/10.1016/j.jesp.2010.03.015

Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition, 29*(1), 1–14. https://doi.org/10.1521/soco.2011.29.1.1

Lameira A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience and biobehavioral reviews*, *83*, 429–439. https://doi.org/10.1016/j.neubiorev.2017.09.021

Lameira, A. R., Eerola, T., & Ravignani, A. (2019). Coupled whole-body rhythmic entrainment between two chimpanzees. *Scientific reports*, *9*(1), 18914. https://doi.org/10.1038/s41598-019-55360-y

Landis-Shack, N., Heinz, A. J., & Bonn-Miller, M. O. (2017). Music Therapy for Posttraumatic Stress in Adults: A Theoretical Review. *Psychomusicology*, *27*(4), 334–342. https://doi.org/10.1037/pmu0000192

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119

Large, E. W., & Gray, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a bonobo (Pan paniscus). *Journal of comparative psychology (Washington, D.C. : 1983)*, *129*(4), 317–328. https://doi.org/10.1037/com0000011

Larsson M. (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Animal cognition*, *17*(1), 1–14. https://doi.org/10.1007/s10071-013-0678-z

Larsson, M., & Abbott, B. W. (2018). Is the Capacity for Vocal Learning in Vertebrates Rooted in Fish Schooling Behavior?. *Evolutionary biology*, *45*(4), 359–373. https://doi.org/10.1007/s11692-018-9457-8

Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal Steps in the Development of Rhythmic Behavior in Humans. *Music & Science, 2*. https://doi.org/10.1177/2059204319892617

Launay, J., Tarr, B., & Dunbar, R. (2016). Synchrony as an Adaptive Mechanism for Large-Scale Human Social Bonding. *Ethology, 122(10)*, 779-789. https://doi.org/10.1111/eth.12528

Leanderson, R., Sundberg, J., & Euler, C.V. (1987). Breathing muscle activity and subglottal pressure dynamics in singing and speech. *Journal of Voice, 1(3)*, 258-261. https://doi.org/10.1016/S0892-1997(87)80009-7

Lecanuet, J. P., Graniere-Deferre, C., Jacquet, A. Y., & DeCasper, A. J. (2000). Fetal discrimination of low-pitched musical notes. *Developmental psychobiology*, *36*(1), 29–39. https://doi.org/10.1002/(SICI)1098-2302(200001)36:1<29::AID-DEV4>3.0.CO;2-J

Lehmann, J., Korstjens, A. H., & Dunbar, R. I. M. (2007). Group size, grooming and social cohesion in primates. *Animal Behaviour, 74*(6), 1617–1629. https://doi.org/10.1016/j.anbehav.2006.10.025

Leman, M., Moelants, D., Varewyck, M., Styns, F., van Noorden, L., & Martens, J. P. (2013). Activating and relaxing music entrains the speed of beat synchronized walking. *PloS one*, *8*(7), e67932. https://doi.org/10.1371/journal.pone.0067932

Leubner, D., & Hinterberger, T. (2017). Reviewing the Effectiveness of Music Interventions in Treating Depression. *Frontiers in psychology*, *8*, 1109. https://doi.org/10.3389/fpsyg.2017.01109

Loersch, C., & Arbuckle, N. L. (2013). Unraveling the mystery of music: music as an evolved group process. *Journal of personality and social psychology*, *105*(5), 777–798. https://doi.org/10.1037/a0033691

Lopresti-Goodman, S. M., Kameka, M., & Dube, A. (2012). Stereotypical behaviors in chimpanzees rescued from the african bushmeat and pet trade. *Behavioral sciences (Basel, Switzerland)*, *3*(1), 1–20. https://doi.org/10.3390/bs3010001

Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), 14563–14568. https://doi.org/10.1073/pnas.1515087112

Ma, W., Fiveash, A., & Thompson, W. (2019). Spontaneous emergence of language-like and music-like vocalizations from an artificial protolanguage. *Semiotica, 2019(229)*, 1 - 23. https://doi.org/10.1515/sem-2018-0139

MacDougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of applied physiology (Bethesda, Md. : 1985)*, *99*(3), 1164–1173. https://doi.org/10.1152/japplphysiol.00138.2005

MacLean, E., & Hare, B. (2013). Spontaneous triadic engagement in bonobos (Pan paniscus) and chimpanzees (Pan troglodytes). *Journal of comparative psychology (Washington, D.C. : 1983)*, *127*(3), 245–255. https://doi.org/10.1037/a0030935

Madison G. (2014). Sensori-motor synchronisation variability decreases as the number of metrical levels in the stimulus signal increases. *Acta psychologica*, *147*, 10–16. https://doi.org/10.1016/j.actpsy.2013.10.002Madison, G., Holmquist, J., & Vestin, M. (2018). Musical improvisation skill in a prospective partner is associated with mate value and preferences, consistent with sexual selection and parental investment theory: Implications for the origin of music. *Evolution and Human Behavior, 39*(1), 120–129. https://doi.org/10.1016/j.evolhumbehav.2017.10.005

Malinowski B (1922*). Argonauts of the Western Pacific*. Routledge and Kegan Paul, London.

Malinowski, B. (1929). *The sexual life of savages in northwestern Melanesia.* [La vie sexuelle des sauvages du Nord-Ouest de la Mélanesie. (Trans. by S. Jankelevitch.)]. Liveright.

Marin, M. M., Schober, R., Gingras, B., & Leder, H. (2017). Misattribution of musical arousal increases sexual attraction towards opposite-sex faces in females. *PloS one*, *12*(9), e0183531. https://doi.org/10.1371/journal.pone.0183531

Mathevon, N., Casey, C., Reichmuth, C., & Charrier, I. (2017). Northern Elephant Seals Memorize the Rhythm and Timbre of Their Rivals' Voices. *Current biology : CB*, *27*(15), 2352–2356.e2. https://doi.org/10.1016/j.cub.2017.06.035

Matsusaka T. (2004). When does play panting occur during social play in wild chimpanzees?. *Primates; journal of primatology*, *45*(4), 221–229. https://doi.org/10.1007/s10329-004-0090-z

McAuley, J. D. (2010). *Tempo and rhythm*. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 165–199). Springer Science + Business Media. https://doi.org/10.1007/978-1-4419-6114-3_6

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception, 23*(1), 29–59. https://doi.org/10.1525/mp.2005.23.1.29

McPherson, T., Berger, D., Alagapan, S., & Fröhlich, F. (2019). Active and Passive Rhythmic Music Therapy Interventions Differentially Modulate Sympathetic Autonomic Nervous System Activity. *Journal of music therapy*, *56*(3), 240–264. https://doi.org/10.1093/jmt/thz007

Mead M (1976). *Growing up in New Guinea*. Morrow, New York (original publication 1930).

Medrano, L., Salinas, M., Salas, I., Guevara, P., Aguayo, A., Jacobsen, J.K., & Baker, C. (1994). Sex identification of humpback whales, Megaptera novaeangliae, on the wintering grounds of the Mexican Pacific Ocean. *Canadian Journal of Zoology, 72*, 1771-

1774. https://doi.org/10.1139/z94-239

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior, 38*(5), 674–684. https://doi.org/10.1016/j.evolhumbehav.2016.12.005

Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and Function in Human Song. *Current biology : CB*, *28*(3), 356–368.e5. https://doi.org/10.1016/j.cub.2017.12.042

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science (New York, N.Y.)*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *The Behavioral and brain sciences*, *44*, e60. https://doi.org/10.1017/S0140525X20000345

Menne-Lothmann, C., Jacobs, N., Derom, C., Thiery, E., van Os, J., & Wichers, M. (2012). Genetic and environmental causes of individual differences in daily life positive affect and reward experience and its overlap with stress-sensitivity. *Behavior genetics*, *42*(5), 778–786. https://doi.org/10.1007/s10519-012-9553-y

Mercado E., III (2018). The Sonar Model for Humpback Whale Song Revised. *Frontiers in psychology*, *9*, 1156. https://doi.org/10.3389/fpsyg.2018.01156

Mercado, E., 3rd, & Handel, S. (2012). Understanding the structure of humpback whale songs (L). *The Journal of the Acoustical Society of America*, *132*(5), 2947–2950.

https://doi.org/10.1121/1.4757643

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual review of neuroscience*, *36*, 313–336. https://doi.org/10.1146/annurev-neuro-062012-170349

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience*, *7*, 274. https://doi.org/10.3389/fnins.2013.00274

Merchant, H., Pérez, O., Bartolo, R., Méndez, J. C., Mendoza, G., Gámez, J., Yc, K., & Prado, L. (2015). Sensorimotor neural dynamics during isochronous tapping in the medial premotor cortex of the macaque. *The European journal of neuroscience*, *41*(5), 586–602. https://doi.org/10.1111/ejn.12811

Merker, B. (2000). *Synchronous chorusing and human origins*. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 315–327). Cambridge, Mass: The MIT Press.

Merker, B. (2002). Music: The missing Humboldt system. *Musicae Scientiae*, 6, 3–21. https://doi.org/10.1177/102986490200600101

Merker, B., Madison, G. S., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45(1), 4–17. https://doi.org/10.1016/j.cortex.2008.06.011.

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *370*(1664), 20140095. https://doi.org/10.1098/rstb.2014.0095

Miani A. (2016). Sexual arousal and rhythmic synchronization: A possible effect of vasopressin. *Medical hypotheses*, *93*, 122–125.

https://doi.org/10.1016/j.mehy.2016.05.030

Miller, G. 2000. *Evolution of human music through sexual selection*. In Wallin, N.L., Merker, B. & Brown, S. (Eds). *The Origins of Music* (pp. 329-360). MIT Press. Cambridge, MA.

Miller, N. S., Kwak, Y., Bohnen, N. I., Müller, M. L., Dayalu, P., & Seidler, R. D. (2013). The pattern of striatal dopaminergic denervation explains sensorimotor synchronization accuracy in Parkinson's disease. *Behavioural brain research*, *257*, 100–110. https://doi.org/10.1016/j.bbr.2013.09.032

Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour, 135*(8-9), 1041-1064. https://doi.org/10.1163/156853998792913483

Mithen, S. 2005. *The singing Neanderthals. The origins of music, language, mind and body*. Widenfeld & Nicholson, London.

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, *152*, 150–159. https://doi.org/10.1016/j.cognition.2016.03.023

Moore, B. L., Connor, R. C., Allen, S. J., Krützen, M., & King, S. L. (2020). Acoustic coordination by allied male dolphins in a cooperative context. *Proceedings. Biological sciences*, *287*(1924), 20192944. https://doi.org/10.1098/rspb.2019.2944

Morris, G. K., Kerr, G. E., and Fullard, J. H. (1978). Phonotactic preferences of female meadow katydids (Orthoptera: *Tettigoniidae*: *Conocephalus nigropleurum*). *Can. J. Zool.* 56, 1479–1487. https://doi.org/10.1139/z78-205.

Moser, C.J., Lee-Rubin, H., Bainbridge, C., Atwood, S., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C., Neff, M.B., Martin, A., Cirelli, L.K., Trehub, S., Song, J., Kim, M., Schachner, A., Vardy, T., Atkinson, Q., Antfolk, J., Madhivanan, P., Siddaiah, A., Placek, C., Salali, G.D., Keestra, S., Singh, M., Collins, S.A., Patton, J., Scaff, C., Stieglitz, J., Moya, C., Sagar, R.R., Wood, B.M., Krasnow, M.M., & Mehr, S.A. (2020). Acoustic regularities in infant-directed vocalizations across cultures. *bioRxiv*. https://doi.org/10.1101/2020.04.09.032995

Mosing, M. A., Pedersen, N. L., Madison, G., & Ullén, F. (2014). Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PloS one*, *9*(11), e113874. https://doi.org/10.1371/journal.pone.0113874

Mosing, M., Verweij, K., Madison, G., Pedersen, N.L., Zietsch, B., & Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior, 36(5)*, 359-366. https://doi.org/10.1016/j.evolhumbehav.2015.02.004

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., Schneider, A., & Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *NeuroImage*, *117*, 267–283. https://doi.org/10.1016/j.neuroimage.2015.05.020

Namu, Yang Erche, Mathieu, Christine. 2004. *Leaving Mother Lake: A Girlhood at the Edge of the World*, Back Bay Books.

Nelson, P. C., & Young, E. D. (2010). Neural correlates of context-dependent perceptual enhancement in the inferior colliculus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(19), 6577–6587. https://doi.org/10.1523/JNEUROSCI.0277-10.2010

Nettl, B. (2015). *The study of ethnomusicology: Thirty-three discussions*, Urbana, IL: University of Illinois Press.

Nettl, B. (2000) *An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture*. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of music* (pp. 463–472). Cambridge, Mass: The MIT Press.

Niarchou, M., Gustavson, D.E., Sathirapongsasuti, J., Anglada-Tort, M., Eising, E., Bell, E., McArthur, E., Straub, P., McAuley, J., Capra, J.A., Ullén, F., Creanza, N., Mosing, M., Hinds, D., Davis, L., Jacoby, N., & Gordon, R. (2021). Unravelling the genetic architecture of musical rhythm: a large-scale genome-wide association study of beat synchronization. *bioRxiv*. https://doi.org/10.1101/836197

Nieder, A., & Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *375*(1789), 20190054. https://doi.org/10.1098/rstb.2019.0054

Nilsson U. (2009). Soothing music can increase oxytocin levels during bed rest after open-heart surgery: a randomised control trial. *Journal of clinical nursing*, *18*(15), 2153–2161. https://doi.org/10.1111/j.1365-2702.2008.02718.x

Noda, T., Amemiya, T., Shiramatsu, T. I., & Takahashi, H. (2017). Stimulus Phase Locking of Cortical Oscillations for Rhythmic Tone Sequences in Rats. *Frontiers in neural circuits*, *11*, 2. https://doi.org/10.3389/fncir.2017.00002

Nozaradan, S., Schwartze, M., Obermeier, C., & Kotz, S. A. (2017). Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex; a journal devoted to the study of the nervous system and behavior*, *95*, 156–168.

https://doi.org/10.1016/j.cortex.2017.08.015

O'Bryan, L.R., Munson, B., & Wilson, M. (2018). Chimpanzee food-associated calls display low referential potential in a wild population. *Journal of the Acoustical Society of America, 143*, 1788-1788. https://doi.org/10.1121/1.5035852

Obozova, T., Smirnova, A., Zorina, Z., & Wasserman, E. (2015). Analogical reasoning in amazons. *Animal cognition*, *18*(6), 1363–1371. https://doi.org/10.1007/s10071-015-0882-0

Oesch, N., & Dunbar, R. I. M. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics, 43*(Part B), 95–106. https://doi.org/10.1016/j.jneuroling.2016.09.008

Oikkonen, J., Huang, Y., Onkamo, P., Ukkola-Vuoti, L., Raijas, P., Karma, K., Vieland, V. J., & Järvelä, I. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular psychiatry*, *20*(2), 275–282. https://doi.org/10.1038/mp.2014.8

Oikkonen, J., Onkamo, P., Järvelä, I., & Kanduri, C. (2016). Convergent evidence for the molecular basis of musical traits. *Scientific reports*, *6*, 39707. https://doi.org/10.1038/srep39707

Ooishi, Y., Mukai, H., Watanabe, K., Kawato, S., & Kashino, M. (2017). Increase in salivary oxytocin and decrease in salivary cortisol after listening to relaxing slow-tempo and exciting fast-tempo music. *PloS one*, *12*(12), e0189075. https://doi.org/10.1371/journal.pone.0189075

Ozdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*(2), 628–635.

https://doi.org/10.1016/j.neuroimage.2006.07.013

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behavioural processes*, *60*(2), 133–155. https://doi.org/10.1016/s0376-6357(02)00080-3

Parncutt, R., & Chuckrow, R. (2017). Chuckrow’s theory of the prenatal origin of music. *Musicae Scientiae, 23*, 403 - 425. https://doi.org/10.1177/1029864917738130

Parr, L. A., & Waller, B. M. (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Social cognitive and affective neuroscience*, *1*(3), 221–228. https://doi.org/10.1093/scan/nsl031

Party, Virginie, Brunel-Pons, Odette, Greenfield, Michael D. 2014. Priority of precedence: receiver psychology, female preference for leading calls and sexual selection in insect choruses, *Animal Behaviour*, *87*, 175-185. https://doi.org/10.1016/j.anbehav.2013.10.029

Patel A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A.D.(2006). Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Perception, 24(1)*, 99-104. https://doi.org/10.1525/mp.2006.24.1.99

Patel A. D. (2021). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *376*(1835), 20200326. https://doi.org/10.1098/rstb.2020.0326

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current biology : CB*, *19*(10), 827–830. https://doi.org/10.1016/j.cub.2009.03.038

Pearce, E., Launay, J., MacCarron, P., & Dunbar, R. (2017). Tuning in to others: Exploring relational and collective bonding in singing and non-singing groups over time. *Psychology of Music, 45(4)*, 496 - 512. https://doi.org/10.1177/0305735616667543

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature neuroscience*, *6*(7), 688–691. https://doi.org/10.1038/nn1083

Persson, T., Sauciuc, G. A., & Madsen, E. A. (2018). Spontaneous cross-species imitation in interactions between chimpanzees and zoo visitors. *Primates; journal of primatology*, *59*(1), 19–29. https://doi.org/10.1007/s10329-017-0624-9

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in evolutionary neuroscience*, *4*, 12. https://doi.org/10.3389/fnevo.2012.00012

Pilozzi, A., Carro, C., & Huang, X. (2020). Roles of β-Endorphin in Stress, Behavior, Neuroinflammation, and Brain Energy Metabolism. *International journal of molecular sciences*, *22*(1), 338. https://doi.org/10.3390/ijms22010338

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *376*(1835), 20200334. https://doi.org/10.1098/rstb.2020.0334

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in psychology*, *5*, 1048. https://doi.org/10.3389/fpsyg.2014.01048

Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H., & Järvelä, I. (2008). Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *Journal of medical genetics*, *45*(7), 451–456.

https://doi.org/10.1136/jmg.2007.056366

Rabinowitch, T.-C., & Meltzoff, A. N. (2017a). Synchronized movement experience enhances peer cooperation in preschool children. *Journal of Experimental Child Psychology, 160,* 21–32. https://doi.org/10.1016/j.jecp.2017.03.001.

Rabinowitch, T. C., & Meltzoff, A. N. (2017). Joint Rhythmic Movement Increases 4-Year-Old Children's Prosocial Sharing and Fairness Toward Peers. *Frontiers in psychology*, *8*, 1050. https://doi.org/10.3389/fpsyg.2017.01050

Rajendran, V. G., Harper, N. S., & Schnupp, J. (2020). Auditory cortical representation of music favours the perceived beat. *Royal Society open science*, *7*(3), 191194. https://doi.org/10.1098/rsos.191194

Ravignani, A., Sonnweber, R.S, Stobbe, N., Fitch, WT. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biology Letters 9(6),* 20130852. https://doi.org/10.1098/rsbl.2013.0852

Ravignani, A., & Madison, G. (2017). The Paradox of Isochrony in the Evolution of Human Rhythm. *Frontiers in psychology*, *8*, 1820. https://doi.org/10.3389/fpsyg.2017.01820

Ravignani, A., & Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across sensory modalities. *Cognition*, *161*, 74–79. https://doi.org/10.1016/j.cognition.2017.01.005

Ravignani, A., Bowling, D. L., & Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in psychology*, *5*, 1118. https://doi.org/10.3389/fpsyg.2014.01118

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour, 1*, 0007. https://doi.org/10.1038/s41562-016-0007

Ravignani, A., & Norton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution, 2(1)*, 4-19. https://doi.org/10.1093/jole/lzx002

Ravignani, A., Chiandetti, C., & Kotz, S. A. (2019). *Rhythm and music in animal signals*. In J. C. Choe (Ed.), *Encyclopedia of animal behavior* (pp. 615–622). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.90729-5

Ravignani, A., Kello, C. T., de Reus, K., Kotz, S. A., Dalla Bella, S., Méndez-Aróstegui, M., Rapado-Tamarit, B., Rubio-Garcia, A., & de Boer, B. (2019). Ontogeny of vocal rhythms in harbor seal pups: an exploratory study. *Current zoology*, *65*(1), 107–120. https://doi.org/10.1093/cz/zoy055

Reber, S. A., Šlipogor, V., Oh, J., Ravignani, A., Hoeschele, M., Bugnyar, T., & Fitch, W. T. (2019). Common marmosets are sensitive to simple dependencies at variable distances in an artificial grammar. *Evolution and human behavior : official journal of the Human Behavior and Evolution Society*, *40*(2), 214–221. https://doi.org/10.1016/j.evolhumbehav.2018.11.006

Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PloS one*, *8*(8), e71182. https://doi.org/10.1371/journal.pone.0071182

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic bulletin & review*, *20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Richman, B. (1978). The synchronization of voices by gelada monkeys. *Primates, 19*, 569-581. https://doi.org/10.1007/BF02373317

Richter, J., & Ostovar, R. (2016). "It Don't Mean a Thing if It Ain't Got that Swing"- an Alternative Concept for Understanding the Evolution of Dance and Music in Human Beings. *Frontiers in human neuroscience*, *10*, 485. https://doi.org/10.3389/fnhum.2016.00485

Riedl, R., & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 63–91. https://doi.org/10.1037/a0026318

Roederer, J. G. (1984). The search for a survival value of music. *Music Perception, 1*(3), 350–356. https://doi.org/10.2307/40285265

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current biology : CB*, *30*(18), 3544–3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Rogers T. L. (2017). Calling underwater is a costly signal: size-related differences in the call rates of Antarctic leopard seals. *Current zoology*, *63*(4), 433–443. https://doi.org/10.1093/cz/zox028

Rogers, C. N., Ross, A. P., Sahu, S. P., Siegel, E. R., Dooyema, J. M., Cree, M. A., Stopa, E. G., Young, L. J., Rilling, J. K., Albers, H. E., & Preuss, T. M. (2018). Oxytocin- and arginine vasopressin-containing fibers in the cortex of humans, chimpanzees, and rhesus macaques. *American journal of primatology*, *80*(10), e22875. https://doi.org/10.1002/ajp.22875

Rosenblum L. D. (2008). Speech Perception as a Multimodal Phenomenon. *Current directions in psychological science*, *17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rouse, A. A., Cook, P. F., Large, E. W., & Reichmuth, C. (2016). Beat Keeping in a Sea Lion As

Coupled Oscillation: Implications for Comparative Understanding of Human Rhythm. *Frontiers in neuroscience*, *10*, 257. https://doi.org/10.3389/fnins.2016.00257

Rouse, A. A., Patel, A. D., & Kao, M. H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(29), e2026130118. https://doi.org/10.1073/pnas.2026130118

Rousseau, J.-J. (1781). *Essay on the Origin of Languages. English translation by J. H. Moran and A. Gode (1986)*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226923284.001.0001

Ryan, M., Fox, J., Wilczynski, W. *et al.* Sexual selection for sensory exploitation in the frog *Physalaemus pustulosus*. *Nature* 343(6253), 66–67 (1990). https://doi.org/10.1038/343066a0

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music, 39(3)*, 307 - 327. https://doi.org/10.1177/0305735610374894

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music, 35*(1), 88–109. https://doi.org/10.1177/0305735607068889

Sachs C (1962). *The wellsprings of music*. M Nijhoff, The Hague. https://doi.org/10.1007/978-94-015-1059-2

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, *19*(2), 86–91. https://doi.org/10.1016/j.tics.2014.12.001

Samuni, L., Preis, A., Mundry, R., Deschner, T., Crockford, C., & Wittig, R. M. (2017).

Oxytocin reactivity during intergroup conflict in wild chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 268–273. https://doi.org/10.1073/pnas.1616812114

Sauciuc, G., Klang, K., Zlakowska, J., & Persson, T. (2019a). Social antecedents of rhythmic behaviours in chimpanzees (Pan troglodytes). *Folia Primatologica, 91*, 290-291.

Sauciuc GA, Zlakowska J, Klang KJ, Bååth R, Persson T. (2019b). Social antecedents and consequences of rhythmic behaviours in chimpanzees (*Pan troglodytes*), 15th *SweCog* Conference, University of Umeå, Sweden, p. 8.

Sauciuc, G. A., Zlakowska, J., Persson, T., Lenninger, S., & Alenkaer Madsen, E. (2020). Imitation recognition and its prosocial effects in 6-month old infants. *PloS one*, *15*(5), e0232717. https://doi.org/10.1371/journal.pone.0232717

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *The Behavioral and brain sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current biology : CB*, *19*(10), 831–836. https://doi.org/10.1016/j.cub.2009.03.061

Schlinger, B., & Brenowitz, E.A. (2002). Neural and Hormonal Control of Birdsong. In Pfaff, D.W., Arnold, A.P., Fahrbach, S.E., Etgen, A.M., Rubin, R.T. (Eds.) *Hormones, brain* *and behavior* (pp. 799-839). Academic Press. https://doi.org/10.1016/B978-012532104-4/50035-4

Schneider, J.N., & Mercado, E. (2019). Characterizing the rhythm and tempo of sound production by singing whales. *Bioacoustics, 28(3)*, 239 - 256. https://doi.org/10.1080/09524622.2018.1428827

Schruth, D.M., Templeton, C.N., & Holman, D. (2019). A definition of song from human music universals observed in primate calls. *bioRxiv*. Https://doi.org/https://doi.org/10.1101/649459

Seeger, A. (1987). *Why Suyá sing: a musical anthropology of an amazonian people*. Cambridge: Cambridge University Press.

Seesjärvi, E., Särkämö, T., Vuoksimaa, E., Tervaniemi, M., Peretz, I., & Kaprio, J. (2016). The Nature and Nurture of Melody: A Twin Study of Musical Pitch and Rhythm Perception. *Behavior genetics*, *46*(4), 506–515. https://doi.org/10.1007/s10519-015-9774-y

Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A., & Brosch, M. (2013). Rhythm sensitivity in macaque monkeys. *Frontiers in systems neuroscience*, *7*, 49. https://doi.org/10.3389/fnsys.2013.00049

Serrano, A., & Terhune, J. (2001). Within-call repetition may be an anti-masking strategy in underwater calls of harp seals (*Pagophilus groenlandicus*). *Canadian Journal of Zoology, 79*, 1410-1413. https://doi.org/10.1139/z01-107

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., & Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *The Lancet. Neurology*, *16*(8), 648–660. https://doi.org/10.1016/S1474-4422(17)30168-0

Slabbekoorn, H., ten Cate, C. (1999) Collared dove responses to playback: Slaves to the rhythm. *Ethology*, 105, 377-391. https://doi.org/10.1046/j.1439-0310.1999.00420.x

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current biology : CB*, *25*(2), 256–260. https://doi.org/10.1016/j.cub.2014.11.063

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-015-0840-x

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine.* 56, 396–408.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Cambridge, Blackwell.

Spierings, M., Hubert, J., & Ten Cate, C. (2017). Selective auditory grouping by zebra finches: testing the iambic-trochaic law. *Animal cognition*, *20*(4), 665–675. https://doi.org/10.1007/s10071-017-1089-3

Stanwood, G. (2019). *Dopamine and Stress*. In Fink, G. (ed.), *Stress: Physiology, Biochemistry, and Pathology*. Pp. 105-114. Academic Press. https://doi.org/10.1016/B978-0-12-813146-6.00009-6

Stokes, P., Shotbolt, P., Mehta, M., Turkheimer, E., Benecke, A., Copeland, C., Turkheimer, F., Lingford-Hughes, A., & Howes, O. (2013). Nature or Nurture? Determining the Heritability of Human Striatal Dopamine Function: an [18F]-DOPA PET Study. *Neuropsychopharmacology, 38*, 485-491. https://doi.org/10.1038/npp.2012.207

Su Y. H. (2016). Sensorimotor Synchronization with Different Metrical Levels of Point-Light Dance Movements. *Frontiers in human neuroscience*, *10*, 186. https://doi.org/10.3389/fnhum.2016.00186

Sundberg, J. (2018). *The Singing voice*. In S. Frühholz & P. Berslin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 117-142). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.6

Takeya, R., Kameda, M., Patel, A. D., & Tanaka, M. (2017). Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Scientific reports*, *7*(1), 6127. https://doi.org/10.1038/s41598-017-06417-3

Takeya, R., Patel, A. D., & Tanaka, M. (2018). Temporal Generalization of Synchronized Saccades Beyond the Trained Range in Monkeys. *Frontiers in psychology*, *9*, 2172. https://doi.org/10.3389/fpsyg.2018.02172

Tanner, J. E., & Byrne, R. W. (2010). Triadic and collaborative play by gorillas in social games with objects. *Animal cognition*, *13*(4), 591–607. https://doi.org/10.1007/s10071-009-0308-y

Tarr, B., Launay, J., Cohen, E., & Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology letters*, *11*(10), 20150767. https://doi.org/10.1098/rsbl.2015.0767

Tarr, B., Launay, J., & Dunbar, R. I. (2016). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and human behavior : official journal of the Human Behavior and Evolution Society*, *37*(5), 343–349. https://doi.org/10.1016/j.evolhumbehav.2016.02.004

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011 ten Cate, C., & Spierings, M. (2019). Rules, rhythm and grouping: Auditory pattern perception by birds. *Animal Behaviour, 151,* 249–257. https://doi.org/10.1016/j.anbehav.2018.11.010

Ten Cate, C., Spierings, M., Hubert, J., & Honing, H. (2016). Can Birds Perceive Rhythmic Patterns? A Review and Experiments on a Songbird and a Parrot Species. *Frontiers in psychology*, *7*, 730. https://doi.org/10.3389/fpsyg.2016.00730

Terleph, T. A., Malaivijitnond, S., & Reichard, U. H. (2017). Male white-handed gibbons flexibly time duet contributions. *Behavioral Ecology and Sociobiology, 72,* Article 16. https://doi.org/10.1007/s00265-017-2432-z

Thaut, M. H., McIntosh, G. C., & Hoemberg, V. (2015). Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. *Frontiers in psychology*, *5*, 1185. https://doi.org/10.3389/fpsyg.2014.01185

Thayer, R. E., Newman, J. R., & McClain, T. M. (1994). Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension. *Journal of personality and social psychology*, *67*(5), 910–925. https://doi.org/10.1037//0022-3514.67.5.910

Tifferet, S., Gaziel, O., and Baram, Y. (2012). Guitar increases male facebook attractiveness. Preliminary support for the sexual selection theory of music. *Letters on Evolutionary Behavioral Science, 3(1)*, 4–6. https://doi.org/10.5178/lebs.2012.18

Toiviainen, P., Burunat, I., Brattico, E., Vuust, P., & Alluri, V. (2020). The chronnectome of musical beat. *NeuroImage*, *216*, 116191. https://doi.org/10.1016/j.neuroimage.2019.116191

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current*

Anthropology, 53(6), 673–692. https://doi.org/10.1086/668207

Toro JM (2016) Something old, something new: combining mechanisms during language acquisition. *Current Directions in Psychological Science*, 25(2),130–134. https://doi.org/10.1177/0963721416629645

Toro, J. M., & Nespor, M. (2015). Experience-dependent emergence of a grouping bias. *Biology letters*, *11*(9), 20150374. https://doi.org/10.1098/rsbl.2015.0374

Toro, J. M., & Hoeschele, M. (2017). Generalizing prosodic patterns by a non-vocal learning mammal. *Animal Cognition, 20*(2), 179–185. https://doi.org/10.1007/s10071-016-1036-8

Trainor, L. (2018). *The origins of music: auditory scene analysis, evolution, and culture in musical creation.* In Honing H. (ed.), *The origins of musicality*, (pp. 81-112). MIT, Cambridge MA.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior & Development, 16*(3), 285–295. https://doi.org/10.1016/0163-6383(93)80036-8

Trehub S. E. (2021). Challenging infant-directed singing as a credible signal of maternal attention. *The Behavioral and brain sciences*, *44*, e117. https://doi.org/10.1017/S0140525X20001442

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55–64. https://doi.org/10.1016/j.neuroimage.2014.09.009

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55–64.

https://doi.org/10.1016/j.neuroimage.2014.09.009Tsay C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14580–14585. https://doi.org/10.1073/pnas.1221454110

Tunçgenç, B., Cohen, E., & Fawcett, C. (2015). Rock with me: the role of movement synchrony in infants' social and nonsocial choices. *Child development*, *86*(3), 976–984. https://doi.org/10.1111/cdev.12354

Tunçgenç, B., & Cohen, E. (2018). Interpersonal movement synchrony facilitates pro-social behavior in children's peer-play. *Developmental science*, *21*(1), 10.1111/desc.12505. https://doi.org/10.1111/desc.12505

Tyack, P. (1981). Interactions between singing Hawaiian humpback whales and conspecifics nearby. *Behavioral Ecology and Sociobiology*, *8*(2), 105-116. https://doi.org/10.1007/BF00300822

Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., & Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PloS one*, *4*(5), e5534. https://doi.org/10.1371/journal.pone.0005534

Ullén, F., Mosing, M., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Personality and Individual Differences, 63*, 87-93. https://doi.org/10.1016/j.paid.2014.01.057

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology, 46*(4), 693–695. https://doi.org/10.1016/j.jesp.2010.03.004

Valentova, J. V., Tureček, P., Varella, M., Šebesta, P., Mendes, F., Pereira, K. J., Kubicová, L., Stolařová, P., & Havlíček, J. (2019). Vocal Parameters of Speech and Singing Covary and Are Related to Vocal Attractiveness, Body Measures, and Sociosexuality: A Cross-Cultural Study. *Frontiers in psychology*, *10*, 2029. https://doi.org/10.3389/fpsyg.2019.02029

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., Ten Cate, C., & Janik, V. M. (2021). The multi-dimensional nature of vocal learning. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *376*(1836), 20200236. https://doi.org/10.1098/rstb.2020.0236

Vuust, P., & Kringelbach, M. (2010). The Pleasure of Making Sense of Music. *Interdisciplinary Science Reviews, 35*, 166 - 182. https://doi.org/10.1179/030801810X12723585301192

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music--brain responses to rhythmic incongruity. *Cortex; a journal devoted to the study of the nervous system and behavior*, *45*(1), 80–92. https://doi.org/10.1016/j.cortex.2008.05.014

Watanabe, J. M., & Smuts, B. B. (1999). Explaining religion without explaining it away: Trust, truth, and the evolution of cooperation in Roy A. Rappaport's "The Obvious Aspects of Ritual". *American Anthropologist, 101*(1), 98–112. https://doi.org/10.1525/aa.1999.101.1.98

Webb, S.J.N., Schapiro, S.J., Hopkins, W. D. (2019). Polymorphisms in Vasopressin V1a Receptor Gene (AVPR1A) are Associated with Play, Affiliative, and Exploratory Behavior in Chimpanzees (Pan troglodytes). The American Society of Primatologists Conference.

Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic bulletin & review*, *23*(6), 1647–1659. https://doi.org/10.3758/s13423-016-1013-x

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological science*, *20*(1), 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2468–2471. https://doi.org/10.1073/pnas.0809035106

Yu, L., & Tomonaga, M. (2018). Effect of visual cues in addition to moderate auditory cues on temporal coordination: A comparative study in humans and chimpanzees. *bioRxiv*. https://doi.org/10.1101/290379

Yuhi, T., Kyuta, H., Mori, H. A., Murakami, C., Furuhara, K., Okuno, M., Takahashi, M., Fuji, D., & Higashida, H. (2017). Salivary Oxytocin Concentration Changes during a Group Drumming Intervention for Maltreated School Children. *Brain sciences*, *7*(11), 152. https://doi.org/10.3390/brainsci7110152

Zarate J. M. (2013). The neural control of singing. *Frontiers in human neuroscience*, *7*, 237. https://doi.org/10.3389/fnhum.2013.00237

Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *Journal of neurophysiology*, *102*(6), 3191–3202. https://doi.org/10.1152/jn.00066.2009

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: singing a different tune. *PLoS biology*, *10*(7), e1001372. https://doi.org/10.1371/journal.pbio.1001372

Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*(Suppl 2), 10430–10437. https://doi.org/10.1073/pnas.1301228110

Zatorre R. J. (2015). Musical pleasure and reward: mechanisms and dysfunction. *Annals of the New York Academy of Sciences*, *1337*, 202–211. https://doi.org/10.1111/nyas.12677

Zwamborn, E. M. J., & Whitehead, H. (2017). Repeated call sequences and behavioural context in long-finned pilot whales off Cape Breton, Nova Scotia, Canada. *Bioacoustics, 26*(2), 169–183. https://doi.org/10.1080/09524622.2016.1233457

Table 1 - Functional theories

(adaptive functions)

Proposal	**Supporting findings / arguments**	**Contradicting findings / arguments**	**Relevant reading**
Sexual selection**:** Protomusical behaviors have evolved as mating displays, and function as honest fitness signals and/ or aesthetic displays. Rhythmic structure supports mating by facilitating species recognition through signal conservation, or by increasing broadcast range through a beacon effect.	- Universality of the courtship function of musical behaviors - Male bias for musical behaviors in humans - Preference for sexual partners with musical ability - The courtship function of ‘song’ in other species (insects, birds, whales, seals) - Male bias for ‘song’ in other species - ‘Song’ is regulated by sex hormones in frogs, birds, seals	- Musical behaviors are found in a broad range of contexts in humans - Both sexes may use music for courtship. - In many societies, the most skilled performers are those beyond reproductive age - Musical skills do not predict mating success - Musical skills do not correlate with fitness - In other species, ‘song’ is often used for territorial defense	Darwin, 1871; Frazer & Mercado, 2000; Greenfield, 2006; Greenfield & Schul, 2008; Greenfield et al., 2017; Hartbauer & Römer, 2016; Madison et al., 2018; Medrano et al., 1996; Merker, 2000; Merker et al., 2009; Mercado, 2018; Miller, 2000; Mosing et al., 2015; Namu & Mathieu, 2004; Rogers, 2017; Savage et al., 2015; Schlinger & Brenowitz, 2017
Vocal emotional communication**:** Speech and song are evolutionarily related: a) musical communication preceded speech; b) musicality is a by-product of speech-related adaptations; c) a rudimentary communication system diverged into a referential communication system (speech) and an emotional one (music)	- Similarities between speech and song: overlapping neural and structural mechanisms, similar mechanisms for emotional modulation / induction - Evidence that a single communication system can split into an emotional vs. referential system, with the emotional system being more music-like	- Musicality is more than song - Diverging neuro-cognitive and physiological mechanisms between song and speech	Brown, 2000a,b, 2017; Darwin, 1871; Fitch, 2006; Ma et al., 2019; Mithen, 2005; Musso et al., 2015; Ozdemir et al., 2006; Peretz & Coltheart 2003; Rousseau, 1761; Spencer, 1857; Sundberg, 2018; Zarate, 2013; Zatorre & Baum (2012)
Social function**:** Musicality has emerged under selection pressures related to human sociality.	- Musical behaviors reach, involve, and connect many people simultaneously - Rhythmic synchronization leads to feelings of liking, rapport, trust, affiliation, entitativity, cooperation and generosity - The prosocial effects of rhythmic synchronization emerge early in ontogeny thus suggesting ancestral evolutionary roots - High quality musical performances (that are synchronized) suggest higher coalition quality - Rhythmic synchronization is suggestive of strength and cohesion	- Musical behaviors have adaptive functions beyond those exerted at group level	Brown, 2000a; Bryant 2013; Cirelli et al., 2014, 2017; Dunbar, 2004, 2012; Fessler & Holbrook, 2016; Freeman, 2001; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Kisrchener & Tomasello, 2009; Kogan, 1997; Lakens & Stel, 2011; Mehr et al., 2020; Merker et al., 2009; Reddish et al., 2013, 2014; Roederer, 1984; Savage et al., 2021; Tarr et al., 2016; Tungenc et al., 2015; Valdesolo et al., 2010; Wiltermuth & Heath, 2009
Generic social functions: Musicality supports group functions regarding within/between group cooperation or inter-group competition.			
Social bonding: Musicality has evolved to replace grooming as hominid group-size increased.			
Social cooperation: Musical behaviors promote affiliation and altruism, thereby increasing the effectiveness of collective actions.			
Coalition signaling: Proto-musical behaviors advertised group fitness and group quality (cohesion, coordination ability, commitment) in the formation of inter-group alliances.			

Mother-infant interaction is the primordial context of musicality.	- Infant-directed song is universal across human cultures - Infant-directed song is most prevalent in mother-infant interaction - Compared to other primates, humans have higher demands for parental investment since a) human infants cannot cling onto their parents' body; b) humans have shorter interbirth intervals and longer postnatal development	- Musical behaviors have adaptive functions beyond those potentially exerted in the context of mother-infant interaction - Musicality is not limited to song - The formal features of infant-directed song diverge from those of musical behaviors in other functional contexts - Features of childcare in foraging cultures, whereby extended breast feeding increases interbirth interval, and infants almost constantly carried	Dissanayake, 2000, 2009a; Mehr & Krasnow, 2017; Mehr et al., 2021; Moser et al., 2020; Trehub et al., 2021
Mater-infant bonding Affiliative signals acquired music-like feature in the context of mother-infant communication through ritualization.			
Credible signaling of parental attention Musicality has emerged as infant-directed song, which is based on ancestral hominid contact calls that have evolved to encode cues of attention to offspring.			
Coping strategy: Musical behaviors have evolved under selective pressures related to a) socio-ecological stressors; b) enhanced episodic cognition, leading to constant apprehension about the future, and thus to stress and anxiety.	- Musical behaviors have therapeutic effects, and are used for stress alleviation in everyday life and clinical settings - Musical behaviors are universally linked to rituals related to events of vital significance for the group or for the individual, where outcome uncertainty is likely to have caused stress and anxiety - Musical stimuli engage neural networks engaged in simulation of future events - Rhythmic entrainment in chimpanzees may be linked to stressful situations	- In other species, behaviors that are considered precursors or analogues of human musical behaviors are present in functional contexts that may not entail a need for coping strategies - Musical behaviors are also found in contexts that are not likely to generate stress and anxiety	Arom, 2000; Brown, 2000a; Cross, 2003; de Witte et al., 2020; Dissanayake, 2009b; Koelsch, 2009, 2014; Lameira et al., 2019; McPherson et al., 2019; Mithen, 2005; Nettle, 1983, 2015; Saarikallio, 2011; Thaut et al., 2015; Thayer et al., 1994

Table 2: Neural adaptations in the evolution of human rhythm cognition

Proposal	**Supporting findings / arguments**	**Contradicting findings / arguments**	**Relevant reading**
Vocal learning and rhythmic synchronization hypothesis: audiomotor entrainment is a by-product of neural adaptations that support vocal mimicry / flexible vocal learning	- Initial evidence of audiomotor entrainment in vocal learning species exclusively - Some similarities between speech and music with respect to the temporal structure of units	- Evidence of audiomotor entrainment in non-vocal learning species	Brown, 2007; Cook et al., 2013; Gámez et al., 2018; Hattori et al., 2013; Large & Gray, 2015; Merchant & Honing, 2014; Noda et al., 2017; Patel, 2006; Patel et al., 2009; Patel, 2021; Rouse et al, 2021; Schachner et al., 2009; Takeya et al., 2017
The gradual audiomotor evolution hypothesis: complex and tightly coupled audiomotor neural circuitry enables relative (beat-based) timing, which in turn enables audiomotor	- Humans exhibit flexible entrainment, other apes exhibit spontaneous entrainment, but less flexible	- Audiomotor entrainment is also present in non-primate species	Cook et al., 2013; Gámez et al., 2018; Gan et al., 2015; Hattori et al., 2013; Iversen et al., 2015; Large & Gray,

entrainment. This circuitry exhibits a gradual evolution in the primate lineage	and complex compared to humans, while monkeys can acquire entrainment with extensive training - More robust and complex audiomotor couplings in the human vs. monkey brain	- Humans also exhibit entrainment that involves other reception channels than the auditory one	2015; Merchant and Honing, 2014; Noda et al., 2017; Takeya et al., 2017; ten Cate & Spierings, 2019
Neural resonance supported by dynamic attending and predictive coding: entrainment is widespread across taxa and results from oscillatory interactions between neurons in the motor and sensory areas. Rhythmic regularity generates predictive coding (i.e. predictions about event recurrence) and supports dynamic attending, i.e. the optimal allocation of attentional resources to stimuli expected to be most salient or informative	- Rhythmic entrainment is present / can be learned by a variety of species - Oscillatory interactions between motor and sensory areas are elicited by rhythmic stimuli in human and nonhuman brains	- Incompatible with evidence that organisms that exhibit the relevant neural auditory and motor brain areas fail to exhibit cognitive abilities that support entrainment in humans	Bartolo et al., 2014; Cook et al., 2013; Large & Gray, 2015; Large & Jones, 1999; Noda et al., 2017; Rouse et al., 2016; ten Cate & Spierings, 2019; Vuust et al., 2009; Wilson & Cook, 2016
Acoustic advantages hypothesis: entrainment is based on audiomotor and genetic adaptations derived from locomotor-reparatory couplings that in fish enable synchronization guided by incidental sounds of locomotion, in order to reduce interferences caused by such sounds. In the tetrapod descendants of ancestral fishes, these couplings further evolved into couplings between forepaw motor processing and vocal communication	- Rhythmic entrainment is present / can be learned by a variety of species	- Incompatible with evidence that organisms that exhibit the relevant neural structures fail to exhibit cognitive abilities that support entrainment in humans	Larsson, 2013; Larsson & Abbott, 2018; ten Cate & Spierings, 2019
Bipedal experience in utero**:** maternal walking provides sensory-motor experience to human fetuses, thereby stimulating the prenatal development of sensory-motor neural couplings, and inducing perceptual biases in the fetus	- In humans, the optimal tempo of rhythmic synchronization overlaps with the optimal tempo of walking and popular dance music	- Nonhuman animals that are not exposed to rhythms derived from bipedal walking during fetal development can spontaneously entrain to human music - The optimal tempo of entrainment in chimpanzees is comparable to that of humans, although chimpanzees do not experience rhythms derived from bipedal walking during fetal development	Hattori et al., 2013; Larsson et al., 2019; Leman et al., 2013; MacDougall & Moore, 2005; Parncutt & Chuckrow, 2017; Repp & Su, 2013

Table 3: Cognitive mechanisms in the evolution of human rhythm cognition

Proposal	Supporting findings / arguments	Contradicting findings / arguments	Relevant reading

Perceptual biases rooted in attentional, learning and / or memory processes that have evolved for the purposes of (a) auditory scene analysis; (b) vocal / non-vocal communication; (c) courtship and mating; (d) group function; (e) parent-offspring bonding; (e) coping.	- The tendency to assign the basic beat to the lowest-pitched instrument reflects inner ear adaptations for the temporal processing of sound, which prioritize low-pitch - The universality of certain rhythmic features across musical cultures, including isochrony, the preference for certain types of relationships and contrasts - (for findings and arguments linked to the functions (b) – (e), see Table 1)		Jacoby & McDermott, 2017; Ravignani et al., 2016; Savage et al., 2015; Trainor, 2018 and references cited in Table 1, concerning the adaptive pressures that could have given rise to such perceptual biases
Perceptual grouping, i.e. the tendency to group sequential stimuli into units based on proximity, similarity, etc., which in human music may support meter perception	- The 'iambic-trochaic law' in humans, whereby a sequence of alternating sounds is chunked into groups of two sounds, and perceived as a: (a) iambic pattern (accent on the second sound) for sequences of short-long alternating sounds; (b) trochaic pattern (accent on the first sound) for sequences of high-low / loud-soft alternating sounds. - Duple meter based on a strong-weak alternation is universal in human music - Animals can be trained to discriminate trochaic and iambic sequences, which are generalized to ambiguous stimuli		Abboub et al., 2016; Fitch, 2016; Hay & Diehl 2007; Hoeschele & Fitch, 2016; Iversen et al. 2008; Mora et al., 2013; Patel, 2006; Savage et al., 2015; Spierings et al., 2017; Toro & Hoeschele, 2017;Toro & Nespor, 2015;
Relational processing, i.e. the ability to map relations between relations, supports relative timing and thus isochrony perception. More complex forms of relational processing, such as hierarchical grouping and recursion enable meter perception and production.	- Isochrony is a universal feature of musical rhythm - Isochrony is also present in other species - Species that entrain spontaneously also exhibit relational reasoning in other domains - Metric structure based on hierarchical grouping and recursion is widespread in human music	- Currently debated if isochrony detection is based on relative timing in the tested species	Celma-Miralles & Toro, 2020; Fitch, 2013; Hoeschele & Bowling, 2016; Hribar et al., 2011; Obozova et al., 2015; Ravignani & Sonnweber, 2017; Savage et al., 2015; Sonnweber et al., 2015; ten Cate & Spierings, 2019

Table 4: Socio-cognitive and motivational mechanisms in the evolution of human rhythm cognition

Proposal	**Supporting findings / arguments**	**Contradicting findings / arguments**	**Relevant reading**

Intrinsic reward and rhythmic synchronization hypothesis: vocal learning species show an intrinsic motivation to entrain to auditory rhythms	- Parrots, which are flexible vocal learners, appear capable of spontaneous audiomotor entrainment	- Chimpanzees, which are not considered flexible vocal learners, exhibit spontaneous – albeit rudimentary - audiomotor entrainment	Hattori et al., 2013; Large & Gray, 2015; Schachner et al., 2009; Takeya et al., 2017; ten Cate & Spierings, 2019
Acoustically mediated emotional contagion: 'contagious' vocal signals (which trigger similar signals and corresponding emotional states in conspecifics) are a homologous precursor of human musicality.	- The pant hoot and the play pant of chimpanzees are 'contagious' vocalizations, entailing that this is an ancestral hominid trait - The pant hoot and the play pant of chimpanzees exhibit motif repetition	- This form of emotional contagion is likely present in other species as well; more research is needed to establish if this is a homologous or analogous trait in human evolution	Davila-Ross et al., 2011; Fritz & Koelsch, 2013; Matsusaka, 2004; Savage et al. 2015
The contagious heterophony hypothesis: unsynchronized group calls are an analogous precursor of music and speech. In ancestral humans, these calls have evolved from reflexive vocal contagion to intentionally synchronized chorusing	- Vocal chorusing in several species, both closely and distantly related to humans	- Potential intentionality in the coordinated calls of chimpanzees and dolphins	Brown, 2007; Crockford et al., 2004; Fedurek et al., 2013; Mitani & Gros-Luis, 1998; Moore et al., 2020
Propensity for gathering to display: the motivation to gather and display in large groups is a primitive, homologous trait of musicality	- Chimpanzees exhibit communal displays that combine loud calls, drumming, and exaggerated bodily movements		Geissman, 2000; Merker et al., 2015
Shared intentionality: rhythmic synchronization relies on a human-specific motivation to share and co-represent mental states with others, which enables joint action coordination towards common goals, and has emerged about 400 000 years ago	- Synchronization-induced prosociality is only present when synchronization is motivated by shared intentionality - Shared intentionality facilitates accurate rhythmic synchronization in toddlers	- The roots of musicality may be traced even earlier, given evidence of shared intentionality in other ape species	Gruber, 2013; Kirschner & Tomasello, 2009; MacLean & Hare, 2013; Persson et al., 2018; Reddish et al., 2013; Tanner & Byrne, 2010

Table 5: Neurohormonal mechanisms in the evolution of human rhythm cognition

Proposal	**Supporting findings / arguments**	**Contradicting findings / arguments**	**Relevant reading**
Endorphins contribute to the positive effects of rhythmic movement and rhythmic synchronization, in particular sense of reward, social bonding, coping with stress and anxiety	- Active engagement in group musical behaviors (communal singing, chanting, drumming, dancing) trigger endorphin release - Music exerts influence on neural assemblies that contribute to emotion regulation through endorphin release - Anxiolytic behaviors in other species are rhythmic		Chamove, 1989; Cohen et al., 2009; Dunbar et al., 2012; Koelsch, 2014; Lameira et al., 2019; Launay et al., 2016; Tarr et al. 2015, 2016
Dopamine is involved in mediating the processing and production of musical rhythm, and the feeling of pleasure elicited by engaging with music	- Several mechanisms involved in rhythm cognition are dopamine-dependent, such as expectations about rhythmic structure and their violations,		Ferreri et al., 2019; Koelsch, 2014; Koshimori et al., 2019; Merchant et al., 2013; Miller et al., 2013; Salimpoor et al.,

	temporal processing in the millisecond range, the reward feeling elicited by music - Dopamine-rich areas are stimulated by music listening and are involved in rhythmic motor control		2015; Vuust & Kringelback, 2010; Zatorre & Salimpoor (2013); Zatorre, 2015
Oxytocin Oxytocin is an underlying mechanism of the social bonding function of musicality	- Music-making, and passive music listening triggers an increase of oxytocin - Oxytocin administration improves rhythmic synchronization		Freeman, 2000; Gebauer et al., 2014; Grape et al., 2002; Nilsson, 2009; Ooishi et al., 2017; Yuhi et al., 2017
Sex steroid hormones.	- Birdsong is highly dependent on the release of sex steroid hormones - Potential indirect influence of sex steroids on rhythm cognition and behavior in humans due to influences exerted on oxytocin		Baribeau & Agnastou (2015; Schlinger & Brenowitz (2002)

Table 6: The genetic bases of human rhythm cognition

Aspect investigated	**Findings**		**References**
Heritability	- Twin studies	- Moderate heritability levels (50%) for similarity judgments of rhythmic structure - Lack of heritability for the ability to detect off-beat sounds based on twin studies	Seesjärvi et al., 2015; Ullén et al., 2014
	- Genome-wide association assays	- Modest heritability (13-16%) for rhythmic synchronization - Modest heritability (21%) for rhythm discrimination	Niarchou et al., 2019; Oikkonen et al., 2014; Pulli et al., 2008
Genes and genetic correlations	- Musical skill, in general, and rhythmic synchronization, in particular, are highly polygenic traits that share molecular mechanisms with traits implicated in emotion, cognition, neurodevelopment and health: lung function, processing speed, strong hand grip, evening chronotype - Polymorphisms in the genes for 1a receptor of arginine-vasopressin predict creative dance, rhythm perception, rhythm memory - The genetic signature of human rhythmic synchronization is enriched for genes expressed after song in area X of the avian brain, which is analogous to the mammalian basal ganglia;		Bachner et al., 2005; Gordon et al., 2021; Granot et al., 2007; Niarchou et al., 2019; Oikkonen et al., 2016; Ukkola et al., 2009