

Piotr Podlipniak
Department of Musicology
Adam Mickiewicz University in Poznań, Poland
ORCID: 0000-0002-4326-559X
Emails: podlip@amu.edu.pl; podlip@poczta.onet.pl
Psychology of Music 2023, Vol. 51(4) 1397–1413 © The Author(s) 2022
<https://doi.org/10.1177/03057356221129319>

Free rider recognition – a missing link in the Baldwinian model of music evolution

Human musical behavior is driven by species-specific predispositions and culture-specific factors. As a result music from different cultures is very diverse (Mehr et al., 2019; Merriam, 1964). However, despite this diversity there are widespread features of music, known as musical universals (Brown & Jordania, 2013; Mehr et al., 2019; Nettl, 2000; Savage et al., 2015; Trehub, 2015), which suggest that music, along with crying, laughter, and speech, belongs to well recognizable human-specific auditory signaling systems. The coexistence of species-specific and culture-specific elements in music additionally suggests that the evolution of human musicality had to coincide with the evolution of cognitive plasticity that led to the development of a cultural environment (Podlipniak, 2021). Such circumstances, if stable long enough, would have created a good opportunity for gene-culture co-evolution (Lumsden & Wilson, 1982). In line with this assumption, the origin of human musicality has recently been hypothesized as the result of this co-evolutionary process (Killin, 2016, 2017, 2018; Patel, 2018, 2021; Podlipniak, 2016, 2017, 2021, 2015; Savage et al., 2021a; Shilton, 2022; Tomlinson, 2015; van der Schyff & Schiavio, 2017). Taking into account the important role of inventiveness in human musical behavior, some of these co-evolutionary scenarios of music origin have included the ‘Baldwin effect’ (Podlipniak, 2016, 2017, 2021, 2015; Savage et al., 2021a) i.e. a type of gene-culture co-evolution in which an initially invented behavioral trait is transformed by means of natural selection into an instinctive behavior (Baldwin, 1896a, 1896b). Savage et al. (2021a) have additionally elaborated on the Baldwinian scenario by postulating an ‘iterated Baldwin effect’, as an evolutionary mechanism that led to the emergence of the coevolving system we know as music. All these Baldwinian explanations of music origin have so far indicated social bonding as an adaptive value of music. In fact, there are a lot of premises that support the crucial role of music in establishing and sustaining social bonds (Dunbar et al., 2012; Pearce et al., 2015, 2017; Tarr et al., 2014). However, the unanswered question concerning the role of social bonding in the Baldwinian evolution of music is why would the cultural variation of music be necessary for social bonding? The ‘social bonding’ hypothesis does not seem to explain this issue alone due to the following reasons.

The first step in every Baldwinian scenario is a social invention (Baldwin, 1896a). In this process, learning must be costly enough in terms of energy and time in order to allow natural selection to favor instinctively behaving individuals (Dor & Jablonka, 2000). The crucial role of cultural information in this process means that the adaptive function of this behavior should be initially achieved more effectively by cultural change than by fixed instinctive features (Godfrey-Smith, 2007). In order to use cultural information in the domain of vocal signaling ‘vocal production learning’ is necessary (Merker, 2021). As far as musicality is concerned, the key sound features to be learned are pitch and rhythm (Bannan, 2009). Therefore, the actual reason for the emergence of the Baldwinian evolution of music is related to the selective pressures that were responsible for the appearance of vocal production learning of pitch and rhythm. It is difficult, however, to envisage how social bonding

could have contributed to the evolution of vocal production learning in the domain of pitch and rhythm among hominins, taking into account the postponed adaptive effect of social cooperation. The strengthening of social bonds by means of a learned behavior, even in the case of music which is relatively fast at creating social bonds (Pearce et al., 2015), is a delayed benefit in comparison to the instant profit of vocal warning obtained by innate signaling for example (Seyfarth et al., 1980). From this perspective, social bonding is a beneficial consequence of signaling (e.g. signaling commitment), rather than the main reason for the appearance of a particular vocalization. In line with this argument, Mehr et al. (2021) have suggested that music evolved as a tool of coalition and parental attention signaling. This idea does not explain, however, why music is a signaling system composed of such a large amount of culture-specific traits. The learning of culture-specific musical traits is costly in terms of time and energy. Why would natural selection have supported such a costly signaling system, instead of preferring innate culturally unchanged vocalizations? After all, both signaling of coalition strength and parental attention can be achieved by innate vocalizations as seen in chimpanzee pant-hooting (Fedurek et al., 2013) and mammalian distress calls (Root-Gutteridge et al., 2021).

Another unsolved quandary with the claim that social bonding is the reason for the appearance of the Baldwinian evolution of music, is related to the problem of the reliability of the signal. It is known that costly signals of commitment can be an important factor in influencing the acceptance of group membership (Ohtsubo & Watanabe, 2009; Power, 2017; Yamaguchi et al., 2015). Signaling commitment and trust can also lead to the strengthening of social bonds. However, as signaling can also be used as a tool of deception (Searcy & Nowicki, 2005) there is a risk that signaling commitment by an individual can be an egoistic strategy designed to cheat the community. As a result, every group that uses signals of commitment must face the challenge of the recognition of their credibility. After all, only a credible signal of commitment can be a good enough source for social bonding. From this perspective, a newly invented code, as in the case of the initial Baldwinian invention, being susceptible to deception, seems to be a poor tool for the creation of social bonds without any mechanism that ensures its credibility. Therefore, while social bonding can be a good reason for sustaining the Baldwinian feedback loop (Savage et al., 2021a) and strengthening the selective pressure during the iteration of the Baldwin effect, it is hard to account for this alone as the trigger for the Baldwinian evolution of music. The aim of this article is to indicate that the Baldwinian evolutionary scenario of music origin should be completed by adding the initial selective forces that led to the next stages of music evolution. In order to find this music origin missing link, the concept of 'free rider' recognition as an adaptive value of the first socially invented musical ritual is proposed. In other words, by indicating 'free rider' recognition as a possible additional adaptive function of music, the proposed view is an extension of the social bonding hypothesis .

The puzzle of the origin of vocal production learning among hominins

Vocal production learning is the ability to reproduce perceived sounds by voice (Janik & Knörnschild, 2021; Janik & Slater, 2000; Merker, 2012). This rare ability consists of adjusting the structure of produced sounds to the acoustic parameters of heard sounds by the means of vocal control. Apart from *Homo sapiens*, this ability has also been noticed in other mammalian taxa such as bats, cetaceans, pinnipeds, and elephants (Janik & Knörnschild, 2021) as well as in three groups of birds i.e. songbirds, parrots, and hummingbirds (Päckert, 2018). Although some convergence between the acoustic parameters of produced and perceived sounds has also been observed in the vocalizations

of other mammals, including primates (Janik & Knörnschild, 2021), it is claimed that *Homo sapiens* is the only primate endowed with vocal production learning (Fitch & Jarvis, 2013; Janik & Slater, 1997; Jarvis, 2019; Petkov & Jarvis, 2012). As imitation is a necessary condition for the development of vocal culture, vocal production learning must have been crucial for the Baldwinian evolution of music. In other words, without vocal production learning, no social invention of even the simplest song would have been possible. Merker, in his commentary to Savage and colleagues' proposal, has claimed, however, that vocal production learning could not have evolved by the Baldwinian mechanism (Merker, 2021). In response, Savage et al. (Savage et al., 2021b) have agreed, suggesting that vocal production learning evolved biologically. While the appearance of vocal production learning may have probably been the result of solely biological forces (Merker, 2021), this does not necessarily mean, however, that the origin of music has no connection with the Baldwinian model of evolution. Firstly, vocal production learning is not a 'binary trait' (Arriaga & Jarvis, 2013; Janik & Knörnschild, 2021; Martins & Boeckx, 2020; Petkov & Jarvis, 2012; Vernes et al., 2021). Indeed, it is characterized by many dimensions such as 'accuracy of the copy,' or 'type of vocal modifications' (Vernes et al., 2021). This means that once vocal learning evolved, the changes of and within its dimensions could have been induced in a Baldwinian way. Therefore, if vocal production learning among our predecessors had not evolved originally as a musical ability, then there would have still been the possibility that musicality would have appeared as a result of the Baldwin effect.

Secondly, although human vocal production learning is treated as one of the most elaborate modes of learning among mammals (Janik & Knörnschild, 2021), people seem to be especially talented in the imitation of particular sound features rather than in a precise duplication of all acoustic traits that characterize every perceived sound (Lemaitre et al., 2016). Not surprisingly, humans are most efficient in the imitation of the features crucial for the recognition of speech and singing units. Moreover, the vocal learning of the distinctive features of the mother tongue (Warlaumont, 2020) and a culture-specific music system such as pitch intervals and rhythms (Benetti & Costa-Giomi, 2019), is spontaneous and happens from infancy. The tight connection between human vocal learning and the learning of speech and singing seems to be facilitated by infants' special attention directed towards speech (Vouloumanos et al., 2010) and singing (Costa-Giomi, 2014; Costa-Giomi & Ilari, 2014). Therefore, the vocal learning that we observe today among humans seems to be especially tuned into speech and music, that is functionally involved in the expression of language-specific propositional meanings and music-specific emotional sensations.

Thirdly, the fact that vocal production learning is absent in our closest relatives – chimpanzees, suggests the relatively faster evolution of volitional vocalizations among hominins. There are, however, at least four abilities observed among some other non-human primates that, if also present in hominins, can be interpreted as the preadaptations for hominin vocal production learning: (1) the ability to adjust vocal production to social context (Seyfarth & Cheney, 2018), (2) the ability to modify certain spectral features of vocalizations (Kalan et al., 2015; Watson et al., 2015), (3) the ability to associate a particular type of vocal signal with referential meaning (Slocombe & Zuberbühler, 2005, 2006), and (4) the use of pitch and rhythm in affective prosody (Zimmermann et al., 2013). While the repertoire of primates calls is relatively constrained and rigid to acoustic modifications (Hammerschmidt & Fischer, 2008), both a primates' decision to vocalize and their choice of a particular call often depend on the social context (Seyfarth & Cheney, 2018; Slocombe & Zuberbühler, 2007). The same dependence most probably characterized the last common ancestor of humans and chimpanzees. This means that even before hominins were able to vocally control their

calls, the use of their vocalizations were susceptible to cultural change. Such an ability could have been a good starting point for the creation of vocal habits related to signaling social intentions. In addition, the ability to combine the acoustic features of vocalizations with culturally flexible meaning, which we observe nowadays among chimpanzees (Watson et al., 2015), could have opened an enormous space for coding information, restricted only by hominins' memory and perceptive resolution. Such a tendency to exploit sounds as a medium of cultural information, if only adaptive, must have been fertile ground for the evolution of vocal culture. This means that the beginning of the evolution of hominin vocal culture preceded the appearance of vocal production learning. Nevertheless, as vocal production learning is definitely a crucial ability which facilitates and accelerates cultural evolution in the domain of vocal culture, it seems reasonable to hypothesize that vocal production learning was a milestone in this process. The main questions however, are which acoustic features were the first sound objects to imitate, and what adaptive value was responsible for the appearance of vocal production learning and for the signal variability among hominins in the domain of music.

Adaptive factors in favor of musical signal variability

Judging by the interspecies comparison of signaling Griebel and Oller (2008) have indicated different functions such as intra- and intersexual competition, social cohesion including parent-offspring bonding, and deception as the potential reasons for the evolution of signal variability. Parent-infant bonding as a reason for the evolution of musical signal variability does not seem very promising. Parental singing to infants i.e. 'infant directed singing', is usually an exaggerated and simplified version of adult singing in terms of pitch contour and tonal complexity respectively (Trainor et al., 1997; Trehub et al., 1993). Lullabies are also less rhythmically complex compared to other songs (Mehr et al., 2019) and infants prefer tonal simplicity (Trainor, 1996; Trehub et al., 1993; Unyk et al., 1992) which suggest the lack of a tendency to complicate musical structure in parent-infant musical communication. Although simplicity of infant directed singing does not exclude variability the openness for complexity observed in adult-mode singing increases the scope of variability. Therefore, even if parent-infant bonding had been an initial source of musical variability (Leongómez et al., 2021), the subsequent evolutionary trajectory for musical signal variability would not have been related to this function, redirecting it towards social bonding among adults. Also, sexual selection as a source of musical signal variability does not seem to be a very convincing explanation. While sexual competition may result in the appearance of culturally variable signals, as in the case of bird and whale songs (Catchpole, 2000; Garland & McGregor, 2020; Noad et al., 2000), there are some characteristics of music that suggest that social factors rather than sexual selection played a crucial role in the origin of musical signal variability. The main clue supporting this claim is the fact that singing, in contrast to speech, tends to be simultaneous (Bannan, 2020). Both singing in unison and in polyphony imposes a coordination between singers which is costly. Even antiphonal singing necessitates the matching of harmonic series between calls separated in time from responses (Wagner & Hoeschele, 2022), which imposes coordination between singers too. The result of this coordination blends all individual displays into a more or less homogeneous signal, which makes it an ineffective strategy for individual fitness advertisement that is indispensable for sexual competition. In fact, although all communal singing can be interpreted as a signal of social cohesion, the value of this signal is measured by the level of similarity. In contrast, sexual display is oriented to show an individual advantage over other individuals. In this game there can be only one winner. Therefore, although one cannot entirely exclude any role of sexual selection in the evolution of musicality

(Darwin, 1871; Miller, 2000; Ravnani, 2018) the predominant communal character of music and its connection with social life (Blacking, 1973; Merriam, 1964; Savage et al., 2015; Turino, 2008) indicates that cooperation, not competition, must have been a more important force related to the evolution of musical signal variability. The social origin of hominin collective and antiphonal singing is additionally supported by the fact that bird (Tobias et al., 2016) and mammalian (King & McGregor, 2016; Tyack, 2008) duets and choruses are also associated with establishing stable social bonds and territoriality.

Free-riding as an inevitable component of social life

If social factors had been related to the appearance of musical signaling and its cultural variability, what would have been the actual adaptive advantages linked to the social life of hominins that would have created the pressure for this process? The obvious benefits of living in social groups include: more effective detection of, deterioration of, and defense against predators (Dunbar, 1996), more successful hunting of large prey (MacNulty et al., 2014; Scheel & Packer, 1991), increased probability of food localization (Bickerton, 2010; Bugnyar, 2013), etc. However, in order to achieve these advantages, gregarious animals have to create and sustain social bonds. This task is associated with many challenges such as inter-individual conflicts resulting from competition within a group, uneven contribution of individual efforts to the group, and the recognition of group members. All these benefits and challenges are the consequence of two antithetical forces that govern life in social groups – the ‘centripetal force’ that sustains cooperation and the ‘centrifugal force’ that promotes selfish behavior (Dunbar, 1996, p. 19; Nowak, 2006). On the one hand, in order to sustain a social group, the individual benefits of group members obtained from cooperation must exceed the profits achieved individually. On the other hand, as reproduction necessitates inter-individual competition it is impossible to eliminate every selfish behavior from a social group. An obvious egoistic strategy is to reap the social benefits without contributing one’s own efforts – ‘free riding’ (Axelrod, 1984). One possible way to achieve this aim is by using deception (Searcy & Nowicki, 2005). If some hominin vocalizations had been used as credible signals of commitment to the group, the simplest way to obtain free rider benefits would have been to mimic these credible signals. In other words, free riders would have received greater benefits than others (Grafen, 1990). In line with this reasoning, a lack of countermeasures against free-riding (e.g. in the case of a certain ‘musical’ individual being endowed with a mutation that prevented him or her from musically induced prosocial behavior) has been posed as one of the arguments used to undermine the ‘social bonding hypothesis’ (Mehr et al., 2021 but see Harrison & Seale, 2021; Wood, 2021). This means that hominins had to face yet another challenge – the recognition of deception. Importantly, in the case where deception is a part of communication, a crucial condition for signal flexibility is learnability (Griebel & Oller, 2008). From this perspective, changing or adding a new learned variant of vocalization can act as a protection against the fake signals of commitment. Learning this new variant necessitates devotion of time and energy which can test the veracity of a hominin’s intentions.

Which sounds were vocally learned first by hominins

As living primates use spectral shape (Watson et al., 2015) and F_0 (Kalan et al., 2015) as the sound signatures of objects it seems reasonable to hypothesize that hominins also used them for these same purposes. So called ‘affective prosody’ (Brown, 2017) which is observed in many living mammalian species, including all primates (Scheumann et al., 2014; Zimmermann et al., 2013), is also

based on the modulation of these acoustic parameters, providing credibility to the presence of this vocalization among hominins. Therefore, it seems to be reasonable that hominins used these acoustic features to code information at least in two important ways: (i) to communicate about external objects; for instance danger, the location of food sources, types of food, and (ii) to communicate subjective attitudes such as aggression, distress, and appeasement (Podlipniak, 2022). While some of these vocalizations were well established instinctive fixed signals, other vocalizations became subject of volitional control. From this perspective, the evolution of hominin vocal production learning is in actual fact the taking of volitional control over two types of vocalizations designed to inform about the concepts of objects and internal emotional states.

If among the vocal repertoires of ancient hominins there were vocally learned calls that were sound symbols of mental concepts referring to perceived objects such as a food source, predators, or prey, the vocal imitation of acoustic features would have started from the sound traits previously used by hominins for this same function. The fact that we observe a tendency to arbitrarily use particular sounds as food symbols among chimpanzees (Kalan et al., 2015; Watson et al., 2015), and the instinctive character of affective prosody (Filippi, 2016; Filippi et al., 2017; Scheumann et al., 2014; Zimmermann et al., 2013), suggests that hominin vocal culture also started from this type of signaling. It seems possible that the competition between hominin groups for food resources created a pressure for the use of group-specific signals that informed not only about the location of fruits but also about other individuals. In order to restrict the intelligibility of the signals, the cultural modifications of these calls would have been the best solution. As a result, the acoustic features of existing vocalizations were not only mimicked but also modified. The increasing number of concepts to communicate was probably the main pressure for the evolution of the vocal control of the aforementioned acoustic traits. This suggests that the appearance of vocal production learning among hominins was related to sound symbols of mental concepts rather than the sound expressions of preconceptual emotional sensations which characterize music.

The beginnings of hominin vocal culture and proto-music

Once the plasticity of vocalization was accessible, both the cultural evolution and gene-culture co-evolutionary mechanisms could have operated. The learnable group-specific signals could have been prone to functional flexibility (Griebel & Oller, 2008), giving the possibility to be used as a signal informing about belonging to a particular group. Taking into account that social bonding is nowadays facilitated by music in which rhythm plays a crucial role (McNeill, 1995), this type of signaling would have been most probably achieved at the beginning by means of sound synchronization. Alternatively, it has been proposed that well synchronized 'musical' signals could have been used as the signals of group consolidation (Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al., 2021) which could have functioned as an acoustic aposematism (Jordania, 2011). However, even if hominin synchronized signaling had evolved originally as an acoustic deterrent oriented against other groups of the same species (Hagen & Bryant, 2003; Hagen & Hammerstein, 2009) or against predators (Jordania, 2011), none of these functions could explain the cultural flexibility of the synchronized signals. While well synchronized signals can be an obvious indicator of consolidation, the cultural variations of such signals seems like an unreasonable expenditure of energy and would make the signal more ambiguous. In other words, the synchronization by itself, not the variations of the synchronized patterns, is enough to send a deterring message. Therefore, the cultural flexibility

of the synchronized rhythms must have evolved because of other reasons, most probably related to social bonding.

Although *Homo sapiens* is the only living primate endowed with the ability to synchronize with periodic sounds in different tempi (Honing, 2019; Patel, 2008), some studies show that the synchronization of movement with musical beat in a restricted periodicity i.e. to 600ms can be strenuously learned by chimpanzees (Hattori et al., 2013, 2015). This means that hominins were probably able to learn rituals based on movements synchronized with sounds. Interestingly, it has been proposed that the appearance of the brain mechanism that enables auditory-motor synchronization in humans was possible thanks to the evolution of vocal production learning (Patel, 2006, 2008, 2021; Patel & Iversen, 2014 but see Brown, 2022; Cook et al., 2013). Alternatively, it has been suggested that the evolution of human auditory-motor synchronization has its deeper evolutionary roots in perceptive abilities that evolved in primates much before the appearance of vocal production learning (Honing et al., 2012, 2018; Merchant & Honing, 2014). Independent of which of these hypotheses is true they do not exclude the view that the gradual evolution of vocal production learning was directed into broadening the volitional control of vocalization timing. Specifically, Patel (2021) has suggested that vocal learning was a preadaptation for the sporadic perception of and synchronization to beat. According to Patel, the advanced form of beat perception and synchronization to it, that we observe among *Homo sapiens*, is a result of gene-culture coevolution (Patel, 2021).

Another premise suggests that apart from rhythm, pitch could also have been a feature used at this time to signal group belonging. This is the crucial role of pitch in affective prosody. As social bonding is emotional (Shultz & Dunbar, 2010), which means that social relations are based on subjective internal states, the exaptation of some elements presented in affective prosody, being designed to communicate subjective attitudes, seems the most parsimonious explanation. One of these elements is pitch which can be used as an emotional signal. An effective way to signal belonging to a particular group is in the alignment of emotional states (Bharucha et al., 2011; Feldman, 2017; Shilton, 2022). As pitch is an important ingredient of affective signaling, emotional alignment can be achieved also by the synchronization of pitches between them. As vocalizations produced by the vocal cords are complex harmonic sounds, the synchronization of pitches does not necessarily mean unison singing. Instead, it can be synchronization between F_0 and other harmonics leading to polyphony (Bannan, 2012). In fact, even singing the same melody by men and women is usually a specious unison which is actually singing in octave (Bannan et al., 2023), which means synchronization between the fundamental frequency and the first harmonic. Thus, the synchronization of pitches can be viewed as a kind of spectral synchronization (Wagner & Hoeschele, 2022). In order to synchronize vocalized pitches between a group of singing individuals, every singer must predict the changes of pitch in time that will occur in their co-singers' singing. Regardless of how the synchronization of affective calls seem to be effective in signaling group belonging, the continuous changes of pitch in affective prosody are hard to predict in comparison to the relatively stable use of pitch in singing (Zatorre & Baum, 2012). This can explain the transition from the former to the latter. The use of a stable pitch in response to the recognized pitch of co-singers' vocalizations necessitates, however, the volitional control of F_0 . This is the moment when the vocal production learning of pitch would have evolved, giving the foundations for the ability of monotonous singing. This ability would have become a milestone in the evolution of singing (Bannan, 2012) and would have opened a space for the cultural variability of pitch sequences.

The role of deception in Baldwinian feedback

If the use of culturally flexible sequences of discrete pitches and rhythms became the signals of group belonging, then deception strategies could have entered the game. As the invention of a complex sound sequence forces the learners to spend time together, the use of communal singing can be a measure of commitment by checking the effort devoted to learn a group-specific vocal signal. Those individuals who devoted less time to learning the group-specific tunes (spending this time for their egoistic aims) could have been detected by the rest of the group by means of the recognition of their low standard performance. One of the hypothetical scenarios of such 'free rider' detection could have occurred when a member of the group, instead of participating in ritual group singing, took the opportunity to steal food gathered and stored by other group members. While the rest of the group strenuously learnt a new variant of ritual song by the means of many repetitions, the free rider would not have been practiced enough to learn this new variant of song. As a result, during the next communal singing the free riders' poor performance would have attracted the attention of the rest of the group risking the possibility of ostracism. In this scenario, communal singing serves as an activity which allows all members of a group to control the behavior of others. Alternatively, one can imagine that a free rider could have tried to cheat by inventing and promoting a new vocalization rather than devoting time to learn the existing song. In this scenario, however, the free rider (not endowed with musicality that characterizes modern humans) had to devote an equal amount of time and energy to invent this new song (and to persuade the rest of the group that this new song is better than the existing one) as in the case of learning the song proposed by the group. Because of this, the former scenario seems more probable. In a such long term conditions, due to inter-individual in-group competition, natural selection would have preferred those who learned faster, in other words those who avoided effort. Under these circumstances, the appearance of deception triggered an arms race between deception and the recognition of deception (Griebel & Oller, 2008). On the one hand, the effective recognition of cheaters could have led to strengthening social bonds between individuals who had recognized themselves as non-cheaters. On the other hand, the presence of 'fast learners' who abused the group – free riders – created the pressure for plasticity that enabled the inventiveness of vocalizations. This arms race is based on the canalization of pitch and rhythm learning as well as on the plasticity which is necessary for creating new vocalizations. This means that free rider recognition can be an important factor in the Baldwinian evolution of music.

The functional specificity of the Baldwinian evolution of music

What is the primordial reason for the Baldwinian evolution of music? Is this the signaling of commitment and trust, the creation and strengthening of social bonds, or the recognition of 'free riders'? In some sense, all these functions can be viewed as different sides of the same coin. There are theoretical models indicating that costly signals can co-evolve with costly cooperative traits (Salahshour, 2019). Taking this model into account, the signaling of commitment by means of vocalized sequences, being costly due to the vocal production learning of discrete pitches and rhythms, could have facilitated cooperation between hominins. Cooperation, in turn, could have induced the tendency to complicate the 'musical' sound signals. This process could have facilitated social bonds as a consequence of the fact that social bonding and cooperation need trust (Roberts, 2020). Additionally, in order to eliminate the inevitable instances of deception the recognition of 'free riders' would have had to evolve. As the arms race between deception and cooperation would

have resulted in the interplay between the canalization and plasticity of musical learning respectively, the recognition of 'free riders' would have been included into the set of forces influencing the Baldwinian evolution of music. In other words, apart from the role in social consolidation, the culture specific variations of music structure can function as a hallmark of group identity. Only learned music structure allows an individual to successfully synchronize with the other members of a group. Those individuals that are unable to synchronize with the group reveal their lack of integration and are endangered with ostracism (Podlipniak, 2017). Therefore, the adaptive value of 'musical plasticity' is not social bonding itself but the recognition of 'self-other' in terms of the assessment of trustworthiness. From this point of view, in the process of music evolution, all these three functions i.e. the signaling of commitment and trust, the creation and strengthening of social bonds, and the recognition of 'free riders' have been interdependent, leading to the appearance of functional feedback loop.

Conclusion

The presented idea concentrates on 'free rider' recognition as a function that has been so far neglected in the Baldwinian scenarios of music evolution (but see Podlipniak, 2017). However, this idea does not diminish the role of social bonding and signaling commitment, but indicates them as equally important factors in the Baldwinian evolution of music. The proposed extension of the Baldwinian model of music evolution focuses only on the ultimate explanation (Fitch, 2015; Tinbergen, 1963), leaving questions about behavioral and neurobiological mechanisms and their development in ontogeny for further research. Of course, the claims about the adaptive functions of hominins' musical behavior are difficult to test because all hominins, except for *Homo sapiens*, are extinct. Therefore, we cannot conduct any experiments on our ancestral species that were not endowed with contemporary human musicality, or observe the behavioral changes that had been occurring in our ancestral lineage. The scope of data that can shed light on the possible role of 'free rider' recognition in the process of shaping our musicality is therefore restricted to that data which can be obtained from interspecies comparative studies and from research on modern humans. However, neither living primates nor modern humans cannot be treated as the reliable models of hominins since both their brains and behavioral repertoires differ from those of hominins as a result of phylogenetic distance. Nevertheless, both some of our and our close animal relatives' traits can be interpreted as the remnants and pre-adaptations of hominins' abilities respectively. Therefore, a useful method to detect possible pre-adaptations for the ability to recognize 'free riders' by the means of music would be by looking for the use of vocalizations as hallmarks of group identity among chimpanzees (cf. e.g. Crockford et al., 2004). Similarly, the suggested idea of 'free rider' recognition can allow us to predict behavioral facts observed among modern humans that imply a possible role of this function in the evolution of human musicality. One such implication could be the level of social cohesion obtained by the means of communal singing. For example, a possible way to trace the remnants of a 'free rider' recognition strategy is to compare the level of social cohesion between the 'devoted' singers of a spontaneously created choir and the singers who avoid singing or who sing out of tune. Another way is to measure the behavioral, physiological, and neural correlates of ostracism (Hudac, 2019; McGuire & Raleigh, 1986; Morese et al., 2019) among the aforementioned 'poor' singers before and after singing. Both the higher level of social cohesion among 'devoted' singers in comparison to 'poor' singers and ostracism towards 'poor' singers, if observed, cannot be explained solely by social bonding, credible signaling, and mate selection theories. Another source of premises that could suggest that 'free rider' recognition could have been an important factor in

shaping our musicality is the research on the convergent evolution of vocalizations. As a convergent evolution of similar traits is usually the result of similar selective pressures (Losos, 2017), the use of culture specific vocalizations as the tools for 'free rider' recognition by animals phylogenetically distant from us (such as birds) could support the presented view.

As human musicality is a set of abilities (Fitch, 2015; Honing, 2018) rather than one uniform trait, its origin has probably been a complex process that has been influenced by many selective pressures. Therefore, the hypotheses of music origin must take into account multifaceted evolutionary paths that have led to the appearance of different abilities. In fact, many contemporary scientific efforts and studies have shown that among our abilities used in music production and perception only some can be treated as music-specific. Nonetheless, looking for their origin is not only an abstract, theoretical task but can also contribute to answering many questions such as: what is the scope of the possible use of music in the solution of social conflicts resulting from suspicion of free riding, and what is an optimal strategy for music education? In the former case, learning and singing together new songs by feuding parties should reduce the conflict, while in the latter, the greater care for individual speed of learning in choirs and musical ensembles would improve the development of teamwork skills. Another important conclusion related to the Baldwinian scenario of music origin is the fact that the so far proposed different adaptive functions of music are not mutually exclusive (Harrison & Seale, 2021). Instead, they could have influenced the appearance of different musical features. The last but not least postulate is that research on the functions of music should take into account pragmatics. After all, the interpretation of any signal can depend on the context (Seyfarth & Cheney, 2018). Since this way of attributing meaning has been observed among chimpanzees (Kalan et al., 2015) one should take this into account in the evolutionary scenarios of music origin. For example, the interpretation of the same well synchronized singing could have been experienced by hominins as formidable, as in the case when one was listening to foreigners and as encouraging when the listener belonged to the singers' group. This means that what acted as a deterrent from one perspective, could have functioned at the same time as a social glue from another.

Acknowledgments

I would like to thank the reviewers for their useful suggestions and inspiring questions. I would also like to thank Peter Kośmider-Jones for his language consultation.

Funding

This research was funded in whole by National Science Centre, Poland [Grant number 2021/41/B/HS1/00541].

References

- Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*(1), 96–116.
<https://doi.org/https://doi.org/10.1016/j.bandl.2012.10.002>
- Axelrod, R. M. (1984). *The evolution of cooperation*. Basic Books.
- Baldwin, J. M. (1896a). A New Factor in Evolution. *The American Naturalist*, *30*(354), 441–451.
<https://doi.org/10.1086/276408>

- Baldwin, J. M. (1896b). A New Factor in Evolution (Continued). *The American Naturalist*, 30(355), 536–553. <https://doi.org/10.1086/276428>
- Bannan, N. (2009). Language out of Music: The Four Dimensions of Vocal Learning. *The Australian Journal of Anthropology*, 19(3), 272–293. <https://doi.org/10.1111/j.1835-9310.2008.tb00354.x>
- Bannan, N. (2012). Harmony and its Role in Human Evolution. In N. Bannan (Ed.), *Music, Language, and Human Evolution* (pp. 288–340). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199227341.003.0012>
- Bannan, N. (2020). An evolutionary perspective on the human capacity for singing. In F. A. Russo, B. Ilari, & A. J. Cohen (Eds.), *The Routledge Companion to Interdisciplinary Studies in Singing, Volume I: Development* (pp. 39–51). Routledge. <https://doi.org/10.4324/9781315163734-3>
- Bannan, N., Bamford, J., & Dunbar, R. I. M. (2023). The evolution of gender dimorphism in the human voice: the role of octave equivalence. *Current Anthropology*, <https://psyarxiv.com/f4j6b/>.
- Benetti, L., & Costa-Giomi, E. (2019). Infant Vocal Imitation of Music. *Journal of Research in Music Education*, 67(4), 381–398. <https://doi.org/10.1177/0022429419890328>
- Bharucha, J. J., Curtis, M., & Paroo, K. (2011). Musical communication as alignment of brain states. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 139–155). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199553426.003.0016>
- Bickerton, D. (2010). *Adam's tongue: how humans made language, how language made humans*. Hill and Wang.
- Blacking, J. (1973). *How musical is man?* University of Washington Press.
- 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. (2022). Group dancing as the evolutionary origin of rhythmic entrainment in humans. *New Ideas in Psychology*, 64, 100902. <https://doi.org/https://doi.org/10.1016/j.newideapsych.2021.100902>
- Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, 41(2), 229–248. <https://doi.org/10.1177/0305735611425896>
- Bugnyar, T. (2013). Social cognition in ravens. *Comparative Cognition & Behavior Reviews*, 8, 1–12. <https://doi.org/10.3819/ccbr.2013.80001>
- Catchpole, C. K. (2000). Sexual Selection and the Evolution of Song and Brain Structure in Acrocephalus Warblers. *Advances in the Study of Behavior*, 29, 45–97. [https://doi.org/https://doi.org/10.1016/S0065-3454\(08\)60103-5](https://doi.org/https://doi.org/10.1016/S0065-3454(08)60103-5)
- 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>
- Costa-Giomi, E. (2014). Mode of Presentation Affects Infants' Preferential Attention to Singing and Speech. *Music Perception: An Interdisciplinary Journal*, 32(2), 160–169. <https://doi.org/10.1525/mp.2014.32.2.160>
- Costa-Giomi, E., & Ilari, B. (2014). Infants' Preferential Attention to Sung and Spoken Stimuli. *Journal*

of *Research in Music Education*, 62(2), 188–194. <https://doi.org/10.1177/0022429414530564>

- 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>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex* (1st ed.). John Murray.
- Dor, D., & Jablonka, E. (2000). From Cultural Selection to Genetic Selection: A Framework for the Evolution of Language. *Selection*, 1(1), 33–56. <https://doi.org/10.1556/Select.1.2000.1-3.5>
- Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.
- Dunbar, R. I. M., 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/epjournal-2536>
- 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/https://doi.org/10.1016/j.anbehav.2013.05.010>
- Feldman, R. (2017). The Neurobiology of Human Attachments. *Trends in Cognitive Sciences*, 21(2), 80–99. <https://doi.org/https://doi.org/10.1016/j.tics.2016.11.007>
- Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. *Frontiers in Psychology*, 7(SEP). <https://doi.org/10.3389/fpsyg.2016.01393>
- Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele, M., Ocklenburg, S., de Boer, B., Sturdy, C. B., Newen, A., & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, 284(1859), 20170990. <https://doi.org/10.1098/rspb.2017.0990>
- Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664), 20140091. <https://doi.org/10.1098/rstb.2014.0091>
- Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In M. A. Arbib (Ed.), *Language, Music and the Brain* (pp. 499–539). The MIT Press.
- Garland, E. C., & McGregor, P. K. (2020). Cultural Transmission, Evolution, and Revolution in Vocal Displays: Insights From Bird and Whale Song. *Frontiers in Psychology*, 11, 544929. <https://doi.org/10.3389/fpsyg.2020.544929>
- Godfrey-Smith, P. (2007). Between Baldwin Scepticism and Baldwin Boosterism. In B. H. Weber & D. J. Depew (Eds.), *Evolution and Learning: The Baldwin Effect Reconsidered* (pp. 53–67). The MIT Press.
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144(4), 517–546. [https://doi.org/https://doi.org/10.1016/S0022-5193\(05\)80088-8](https://doi.org/https://doi.org/10.1016/S0022-5193(05)80088-8)
- Griebel, U., & Oller, D. K. (2008). Evolutionary Forces Favoring Communicative Flexibility. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (pp. 9–40). The MIT Press.

<https://doi.org/10.7551/mitpress/7650.003.0006>

- Hagen, E. H., & Bryant, G. A. (2003). Music and Dance As a Coalition Signaling System. *Human Nature*, *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*, *13*(2 Suppl), 291–320. <https://doi.org/10.1177/1029864909013002131>
- Hammerschmidt, K., & Fischer, J. (2008). Constraints in Primate Vocal Production. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (pp. 93–119). The MIT Press. <https://doi.org/10.7551/mitpress/9780262151214.003.0005>
- Harrison, P. M. C., & Seale, M. (2021). Against unitary theories of music evolution. *Behavioral and Brain Sciences*, *44*, e76. <https://doi.org/DOI:10.1017/S0140525X20001314>
- Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, *3*(1), 1566. <https://doi.org/10.1038/srep01566>
- Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2015). Distractor Effect of Auditory Rhythms on Self-Paced Tapping in Chimpanzees and Humans. *PloS One*, *10*(7), e0130682–e0130682. <https://doi.org/10.1371/journal.pone.0130682>
- Honing, H. (2018). On the biological basis of musicality. *Annals of the New York Academy of Sciences*, *1423*(1), 51–56. <https://doi.org/https://doi.org/10.1111/nyas.13638>
- Honing, H. (2019). *The evolving animal orchestra: in search of what makes us musical* (S. Macdonald (trans.)). MIT Press.
- 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>
- Honing, H., Merchant, H., Háden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PloS One*, *7*(12), e51369–e51369. <https://doi.org/10.1371/journal.pone.0051369>
- Hudac, C. M. (2019). Social priming modulates the neural response to ostracism: a new exploratory approach. *Social Neuroscience*, *14*(3), 313–327. <https://doi.org/10.1080/17470919.2018.1463926>
- Janik, V. M., & Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200244. <https://doi.org/10.1098/rstb.2020.0244>
- Janik, V. M., & Slater, P. J. B. (1997). Vocal Learning in Mammals. *Advances in the Study of Behavior*, *26*(C), 59–99. [https://doi.org/10.1016/S0065-3454\(08\)60377-0](https://doi.org/10.1016/S0065-3454(08)60377-0)
- Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. <https://doi.org/10.1006/anbe.2000.1410>
- Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50 LP – 54. <https://doi.org/10.1126/science.aax0287>

- Jordania, J. (2011). *Why Do People Sing? Music in Human Evolution*. Logos.
- Kalan, A. K., Mundry, R., & Boesch, C. (2015). Wild chimpanzees modify food call structure with respect to tree size for a particular fruit species. *Animal Behaviour*, *101*, 1–9. <https://doi.org/10.1016/j.anbehav.2014.12.011>
- Killin, A. (2016). Rethinking music's status as adaptation versus technology: a niche construction perspective. *Ethnomusicology Forum*, *25*, 1–24. <https://doi.org/10.1080/17411912.2016.1159141>
- Killin, A. (2017). Plio-Pleistocene Foundations of Hominin Musicality: Coevolution of Cognition, Sociality, and Music. *Biological Theory*, *12*(4), 222–235. <https://doi.org/10.1007/s13752-017-0274-6>
- Killin, A. (2018). The origins of music: Evidence, theory, and prospects. *Music & Science*, *1*, 2059204317751971. <https://doi.org/10.1177/2059204317751971>
- Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., & Susini, P. (2016). Vocal Imitations of Non-Vocal Sounds. *PLOS ONE*, *11*(12), e0168167. <https://doi.org/10.1371/journal.pone.0168167>
- Leongómez, J. D., Havlíček, J., & Roberts, S. C. (2021). Musicality in human vocal communication: an evolutionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 20200391. <https://doi.org/10.1098/rstb.2020.0391>
- Losos, J. B. (2017). *Improbable destinies: fate, chance, and the future of evolution*. Riverhead Books.
- Lumsden, C. J., & Wilson, E. O. (1982). Précis of Genes, Mind, and Culture. *The Behavioral and Brain Sciences*, *5*, 1–37. <https://doi.org/10.1142/5786>
- MacNulty, D. R., Tallian, A., Stahler, D. R., & Smith, D. W. (2014). Influence of group size on the success of wolves hunting bison. *PloS One*, *9*(11), e112884–e112884. <https://doi.org/10.1371/journal.pone.0112884>
- Martins, P. T., & Boeckx, C. (2020). Vocal learning: Beyond the continuum. *PLOS Biology*, *18*(3), e3000672. <https://doi.org/10.1371/journal.pbio.3000672>
- McGuire, M. T., & Raleigh, M. J. (1986). Behavioral and physiological correlates of ostracism. *Ethology and Sociobiology*, *7*(3), 187–200. [https://doi.org/https://doi.org/10.1016/0162-3095\(86\)90047-6](https://doi.org/https://doi.org/10.1016/0162-3095(86)90047-6)
- McNeill, W. H. (1995). *Keeping together in time : dance and drill in human history*. Harvard University Press.
- Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*, e60. <https://doi.org/DOI:10.1017/S0140525X20000345>
- 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*, *366*(970), eaax0868. <https://doi.org/10.1126/science.aax0868>
- 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>

Merker, B. (2012). The Vocal Learning Constellation. In N. Bannan (Ed.), *Music, Language, and Human Evolution* (pp. 215–260). Oxford University Press.

<https://doi.org/10.1093/acprof:osobl/9780199227341.003.0009>

Merker, B. (2021). Music, bonding, and human evolution: A critique. *Behavioral and Brain Sciences*, 44, e83. <https://doi.org/DOI: 10.1017/S0140525X20001429>

Merriam, A. P. (1964). *The anthropology of music*. Northwestern University Press.

Miller, G. F. (2000). Evolution of Human Music Through Sexual Selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 329–360). The MIT Press.

<https://doi.org/10.1177/004057368303900411>

Morese, R., Lamm, C., Bosco, F. M., Valentini, M. C., & Silani, G. (2019). Social support modulates the neural correlates underlying social exclusion. *Social Cognitive and Affective Neuroscience*, 14(6), 633–643. <https://doi.org/10.1093/scan/nsz033>

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In Nils L Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463–472). The MIT Press.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M.-N., & Jenner, K. C. S. (2000). Cultural revolution in whale songs. *Nature*, 408(6812), 537. <https://doi.org/10.1038/35046199>

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563.

<https://doi.org/10.1126/science.1133755>

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, 30(2), 114–123.

<https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2008.09.004>

Päckert, M. (2018). Song: The Learned Language of Three Major Bird Clades. In D. T. Tietze (Ed.), *Bird Species: How They Arise, Modify and Vanish* (pp. 75–94). Springer International Publishing.

https://doi.org/10.1007/978-3-319-91689-7_5

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

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D. (2018). Music as a Transformative Technology of the Mind: An Update. In H. Honing (Ed.), *The Origins of Musicality* (pp. 113–126). The MIT Press.

<https://doi.org/10.7551/mitpress/10636.003.0009>

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

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8, 57. <https://doi.org/10.3389/fnsys.2014.00057>

Pearce, E., Launay, J., & Dunbar, R. I. M. (2015). The ice-breaker effect : singing mediates fast social bonding. *Royal Society Open Science*, 2(10), 1–9. <https://doi.org/10.1098/rsos.150221>

- Pearce, E., Launay, J., MacCarron, P., & Dunbar, R. I. M. (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>
- Petkov, C., & Jarvis, E. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, 4, 12. <https://www.frontiersin.org/article/10.3389/fnevo.2012.00012>
- Podlipniak, P. (2016). The evolutionary origin of pitch centre recognition. *Psychology of Music*, 44(3), 527–543. <https://doi.org/10.1177/0305735615577249>
- Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. *Frontiers in Neuroscience*, 11(Oct), 542. <https://doi.org/10.3389/fnins.2017.00542>
- Podlipniak, P. (2021). The Role of Canalization and Plasticity in the Evolution of Musical Creativity. *Frontiers in Neuroscience*, 15, 267. <https://doi.org/10.3389/fnins.2021.607887>
- Podlipniak, P. (2022). Pitch syntax as part of an ancient protolanguage. *Lingua*, 271, 103238a. <https://doi.org/10.1016/J.LINGUA.2021.103238>
- Podlipniak, P. (2015). The origin of music and the Baldwin effect. In J. Ginsborg, A. Lamont, & S. Bramley (Eds.), *Proceedings of Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 671–677). Royal Northern College of Music.
- Power, E. A. (2017). Discerning devotion: Testing the signaling theory of religion. *Evolution and Human Behavior*, 38(1), 82–91. <https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2016.07.003>
- Ravignani, A. (2018). Darwin, Sexual Selection, and the Origins of Music. *Trends in Ecology and Evolution*, 33(10), 716–719. <https://doi.org/10.1016/j.tree.2018.07.006>
- Roberts, G. (2020). Honest signaling of cooperative intentions. *Behavioral Ecology*, 31(4), 922–932. <https://doi.org/10.1093/beheco/araa035>
- Root-Gutteridge, H., Ratcliffe, V. F., Neumann, J., Timarchi, L., Yeung, C., Korzeniowska, A. T., Mathevon, N., & Reby, D. (2021). Effect of pitch range on dogs' response to conspecific vs. heterospecific distress cries. *Scientific Reports*, 11(1), 19723. <https://doi.org/10.1038/s41598-021-98967-w>
- Salahshour, M. (2019). Evolution of costly signaling and partial cooperation. *Scientific Reports*, 9(1), 8792. <https://doi.org/10.1038/s41598-019-45272-2>
- 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. (2021a). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, 44, e59. <https://doi.org/DOL: 10.1017/S0140525X20000333>
- Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021b). Toward inclusive theories of the evolution of musicality. *Behavioral and Brain Sciences*, 44, e121. <https://doi.org/DOL: 10.1017/S0140525X21000042>
- Scheel, D., & Packer, C. (1991). Group hunting behaviour of lions: a search for cooperation. *Animal*

- Behaviour*, 41(4), 697–709. [https://doi.org/https://doi.org/10.1016/S0003-3472\(05\)80907-8](https://doi.org/https://doi.org/10.1016/S0003-3472(05)80907-8)
- Scheumann, M., Hasting, A. S., Kotz, S. A., & Zimmermann, E. (2014). The Voice of Emotion across Species: How Do Human Listeners Recognize Animals' Affective States? *PLoS ONE*, 9(3), Article ID e91192. <https://doi.org/10.1371/journal.pone.0091192>
- Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press.
- Seyfarth, R. M., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current Opinion in Behavioral Sciences*, 21, 56–61. <https://doi.org/https://doi.org/10.1016/j.cobeha.2018.02.005>
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science*, 210(4471), 801–803. <https://doi.org/10.2307/1684570>
- Shilton, D. (2022). Sweet Participation: The Evolution of Music as an Interactive Technology. *Music & Science*, 5, 20592043221084710. <https://doi.org/10.1177/20592043221084710>
- Shultz, S., & Dunbar, R. (2010). Bondedness and sociality. *Behaviour*, 147(7), 775–803. <https://doi.org/https://doi.org/10.1163/000579510X501151>
- Slocombe, K. E., & Zuberbühler, K. (2005). Functionally Referential Communication in a Chimpanzee. *Current Biology*, 15(19), 1779–1784. <https://doi.org/10.1016/j.cub.2005.08.068>
- Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, 72(5), 989–999. <https://doi.org/https://doi.org/10.1016/j.anbehav.2006.01.030>
- Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, 104(43), 17228–17233. <https://doi.org/10.1073/pnas.0706741104>
- Tarr, B., Launay, J., & Dunbar, R. I. M. (2014). Music and social bonding: “Self-other” merging and neurohormonal mechanisms. *Frontiers in Psychology*, 5, 1096. <https://doi.org/10.3389/fpsyg.2014.01096>
- Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift Für Tierpsychologie*, 20(4), 410–433. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>
- Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., & Nakagawa, S. (2016). Territoriality, Social Bonds, and the Evolution of Communal Signaling in Birds. *Frontiers in Ecology and Evolution*, 4, 00074. <https://www.frontiersin.org/article/10.3389/fevo.2016.00074>
- Tomlinson, G. (2015). *A million years of music: the emergence of human modernity*. The MIT Press.
- Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development*, 19(1), 83–92. [https://doi.org/10.1016/S0163-6383\(96\)90046-6](https://doi.org/10.1016/S0163-6383(96)90046-6)
- Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, 20(3), 383–396. [https://doi.org/10.1016/S0163-6383\(97\)90009-6](https://doi.org/10.1016/S0163-6383(97)90009-6)
- Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proceedings of the National*

- Academy of Sciences*, 112(29), 8809 LP – 8810. <https://doi.org/10.1073/pnas.1510724112>
- Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior and Development*, 16(3), 285–295. [https://doi.org/10.1016/0163-6383\(93\)80036-8](https://doi.org/10.1016/0163-6383(93)80036-8)
- Turino, T. (2008). *Music as social life: the politics of participation*. University of Chicago Press.
- Unyk, A. M., Trehub, S. E., Trainor, L. J., & Schellenberg, E. G. (1992). Lullabies and Simplicity: A Cross-Cultural Perspective. *Psychology of Music*, 20(1), 15–28. <https://doi.org/10.1177/0305735692201002>
- van der Schyff, D., & Schiavio, A. (2017). Evolutionary Musicology Meets Embodied Cognition: Biocultural Coevolution and the Enactive Origins of Human Musicality. *Frontiers in Neuroscience*, 11, 519. <https://www.frontiersin.org/article/10.3389/fnins.2017.00519>
- 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 B: Biological Sciences*, 376(1836), 20200236. <https://doi.org/10.1098/rstb.2020.0236>
- Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The Tuning of Human Neonates' Preference for Speech. *Child Development*, 81(2), 517–527. <http://www.jstor.org/stable/40598998>
- Wagner, B., & Hoeschele, M. (2022). The links between pitch, timbre, musicality, and social bonding from cross-species research. *Comparative Cognition & Behavior Reviews*, 17, 13–32. <https://doi.org/10.3819/CCBR.2022.170002>
- Warlaumont, A. S. (2020). Infant Vocal Learning and Speech Production. In C. S. Tamis-LeMonda & J. J. Lockman (Eds.), *The Cambridge Handbook of Infant Development: Brain, Behavior, and Cultural Context* (pp. 602–631). Cambridge University Press. [https://doi.org/DOI: 10.1017/9781108351959.022](https://doi.org/DOI:10.1017/9781108351959.022)
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, 25(4), 495–499. <https://doi.org/10.1016/j.cub.2014.12.032>
- Wood, C. (2021). Musical bonds are orthogonal to symbolic language and norms. *Behavioral and Brain Sciences*, 44, e119. [https://doi.org/DOI: 10.1017/S0140525X20001272](https://doi.org/DOI:10.1017/S0140525X20001272)
- Yamaguchi, M., Smith, A., & Ohtsubo, Y. (2015). Commitment signals in friendship and romantic relationships. *Evolution and Human Behavior*, 36(6), 467–474. <https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2015.05.002>
- Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology*, 10(7), Article ID e1001372. <https://doi.org/10.1371/journal.pbio.1001372>
- Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: A comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of Emotional Communication: from Sounds in Nonhuman Mammals to Speech and Music in Man*. (pp. 116–132). Oxford University Press.