

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
Musicae Scientiae, Online-first, 1–22 © The Author(s) 2024
<https://doi.org/10.1177/10298649241255694>

The evolution of musicality and cross-domain co-evolutionary interactions

Abstract

The evolution of human *musicality* has often been linked to the evolution of the faculty of language since the development of musical and linguistic abilities seems to share a common phase in their ontogenesis. Apart from that, both singing and speaking are, on the one hand, universal forms of human vocal expression, and on the other hand, consist of culturally specific elements. Such a probable co-occurrence of the predisposition to speak and sing, with the cultural variability of both of these forms of communication, has prompted researchers to indicate gene-culture coevolution as the probable mechanism responsible for the emergence of human *musicality* and the faculty of language. However, in most evolutionary scenarios proposed so far, the evolutionary paths of music and language followed independently after divergence from a common precursor. This article, based on observations of contemporary interactions between language and music, presents a different view in which musical and language-like forms of proto-communication interacted leading to the repurposing of some of their neural mechanisms. In this process, the Baldwinian interplay between plasticity and canalization has been proposed as the most probable evolutionary mechanism that shaped our *musicality*. The premises that support the presence of cross-domain coevolutionary interactions in the contemporary communicative niche of *Homo sapiens* are indicated.

Music and natural language are inseparable parts of contemporary human culture, which often interact in various types of lyrical expression such as songs and melo-recitation. Although the term *music* has been used in the Western tradition to refer to many different forms of expression involving sound, songs and instrumental performances have certain functional and structural similarities that can be observed in all, often disparate, human cultures (Mehr et al., 2018; 2019; Savage et al., 2015). These similarities are probably not merely coincidental but could have emerged from a set of abilities that are shared by *Homo sapiens*, referred to as *musicality* (Fitch, 2015; Honing, 2018), and therefore form a natural basis for music. The crucial consequence of the influence of human *musicality* on music is that music is a form of expression involving sound divided into discrete units of pitches and rhythms. From this perspective, music and natural language, the analogous discrete units of which are phonemes, are human-specific forms of communication through sound exemplifying the Humboldt system (Merker, 2002). Despite these similarities both music and language, as observed in human cultures around the world, are enormously diverse communicative systems. This diversity would not be possible without cultural inventiveness which, as a part of gene-culture coevolution, is more and more often treated as an important element in the evolution of cognitive abilities (Jablonka & Lamb, 2005; Laland, 2017; Whitehead et al., 2019), including *musicality* (Patel, 2023; Tomlinson, 2015). Even if current gene-culture coevolutionary explanations of the evolution of music (Killin, 2017; Podlipniak, 2017; Savage et al., 2021a; Shilton, 2022; Tomlinson, 2015) allow for different, non-mutually exclusive, adaptive functions of music, they still point to functionally independent evolutionary pathways that could have led to the appearance of human *musicality*. Admittedly, Savage et al. (2021) have indicated that elements of *musicality* may originally have arisen because of other adaptations that were not music-specific. However, they have emphasized that only later in the process of the gene-culture coevolution of human *musicality* were these formerly non-music specific abilities exapted (i.e., they were pre-existing traits that were repurposed to fulfil a new function; see Gould & Vrba, 1982) and modified due to the adaptive function of music, which in their view was to promote social bonding. According to Savage et al. (2021) a clear sequence of events occurred, starting with adaptive abilities that arose because of adaptive values unrelated to *musicality*, followed by the process of gene-culture coevolution. This in turn led to the development of human *musicality*, thanks to the adaptive value of social bonding. While this sequence allows for the exaptation of previously non-specific musical abilities, the whole process of gene-culture coevolution of human *musicality* took place independently from the evolution of other non-musical communicative abilities. From this perspective, the gene-culture coevolution of human *musicality* was unidirectional and restricted to one domain, in that the coevolutionary path and all the selective pressures that shaped our capacity for music were related solely to the adaptive value(s) of music.

In this article another view is proposed, in which the Baldwinian coevolution of a language-like propositional domain and a music-like emotional domain of communication interacted in such a way that the abilities for one were exapted by the other and vice versa. However, in this process, in contrast to classical exaptation by natural selection in genetic evolution (Gould & Vrba, 1982), the main cause of selection leading to exaptation was the cultural invention of using some elements of one domain in order to achieve functions specific to the other domain. Specifically, certain features of spoken language such as intonation and syntax were initially adopted socially from music-like emotional vocalizations. Similarly, some types of concept originating in associations formerly reserved solely for language-like propositional communication, such as timbral symbolism and iconicity (Imai & Kita, 2014), were taken over by music. One method used to infer the evolutionary interplay of music-like and language-like communication is based on Bateson's model of double description (Bateson, 1979; Hui et al., 2008), which is a type of logical abduction.

The origin of *musicality* and Baldwinian loops

The debate about the origin of human *musicality* has been dominated by the search for the adaptive value of music because, for many scholars, musical behavior seems to be biologically useless (Pinker, 1997; Wilson, 2013). Nonetheless, since Darwin, many different adaptive functions have been proposed as the possible reason for the evolution of *musicality*, such as strengthening social bonds (Dunbar, 2012; Roederer, 1984; Savage et al., 2021a; Storr, 1992), mate attraction (Darwin, 1871; Miller, 2000; Ravignani, 2018), enhancing mother-infant affiliative interactions (Dissanayake, 2001), parent-infant communication and bonding (Leongómez et al., 2021), eliciting attention in parent-infant competition (Mehr & Krasnow, 2017), signaling social strength (Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al., 2021), free-rider recognition (Podlipniak, 2022b), deterring predators (Jordania, 2011), and vocal grooming (Dunbar, 1996). However, since *musicality* is a set of distinct abilities (Fitch, 2015; Honing, 2018), they might have evolved independently because they had different adaptive values. They need not be mutually exclusive, for the purposes of explaining the evolution of *musicality* (Harrison & Seale, 2021; Juslin, 2021; Savage et al., 2021b), so long as they can be included in an evolutionary scenario that explains the phylogeny of *musicality*. Since hominins can also respond to environmental challenges by means of phenotypic adaptations (i.e., by means of acquired features) such as habits and traditions (Avital & Jablonka, 2000; S. E. Fisher & Ridley, 2013; Jablonka & Lamb, 2005), the possibility that such adaptations also played an important role in the evolution of *musicality* cannot be ruled out. This possibility leads to the assumption that, rather than being solely a result of genetic evolution, human *musicality* could have been a product of gene-culture coevolution (Killin, 2016, 2017, 2018; Patel, 2018, 2021, 2023; Podlipniak, 2016, 2017, 2021; Savage et al., 2021a; Tomlinson, 2015), a process in which cultural and genetic evolutions interact, leading to the appearance and inheritance of new traits (Lumsden & Wilson, 1982; Richerson et al., 2010).

A special type of gene-culture coevolution is Baldwinian evolution (Baldwin, 1896a, 1896b). In this process, a culturally invented behavioral trait, which is adaptive and in which the learning is time-consuming and effortful, comes under genetic control. Although this process may look like Lamarckian inheritance of acquired traits, the actual reason for the inheritance of such a behavioral trait is the accidental appearance of a mutation enabling a predisposition to learn this behavior faster and with less effort (Hall, 2001), which is then favored by natural selection (although for an alternative explanation, see Hughes, 2012). Importantly, in Baldwinian evolution, genetic inheritance follows cultural inheritance (West-Eberhard, 2005), in which the crucial role is played by learning (Jablonka & Lamb, 2005). Learning, like inventing new behaviors, necessitates a specific kind of developmental (or phenotypic) plasticity (Pigliucci, 2001; West-Eberhard, 2003) that enables ontogenetic modification of behavior in response to the environment (including the cultural environment). This kind of plasticity is called behavioral plasticity (Dor & Jablonka, 2010; Mery & Burns, 2010) and is grounded in neural plasticity (Dor & Jablonka, 2014). However, neural plasticity can differ depending on the specializations that have evolved. Patel (2023) refers to two types of neural plasticity proposed by Greenough et al. (1987): *experience-expectant* and *experience-dependent* plasticity. Both are mechanisms that enable the acquisition of culture-specific traits such as language-specific vocabulary and writing. However, while the phonological system specific to a particular language is acquired as a result of experience-expectant plasticity, the learning of writing necessitates experience-dependent plasticity. As a result, speech—as a default form of language—is a human universal, but literacy is not observed in all human cultures (Pinker, 1994). This means that experience-expectant plasticity is restricted by a canalized neural system, whereas experience-dependent plasticity opens more space for learning.

These two types of plasticity seem to be the solutions for different environmental challenges. Greenough et al. (1987) have proposed that experience-expectant plasticity evolved as a way of dealing with environmental challenges that are ubiquitous and stable. In contrast, experience-dependent plasticity developed, according to them, as a tool for storing information that is related to experiences that are unique to a particular individual, such as the location of a source of food or shelter. In the case of hominin culture, however, experience-dependent plasticity permits the individual not only to remember information unique to them, but also to learn cultural innovations that are widespread throughout the whole group, such as writing. In a stable cultural environment, experience-expectant plasticity that enables fast learning during sensitive periods has an advantage over experience-dependent plasticity, which necessitates more time-consuming and effortful learning. From the Baldwinian point of view, experience-expectant plasticity related to a particular trait is therefore a result of canalization, but it allows cultural evolution of this trait, such as a particular language (Dor & Jablonka, 2010, 2014) or culture-specific music (Jan, 2018, 2022; Savage, 2019). By contrast, experience-dependent plasticity is a source of behavioral innovations that go beyond the canalized behavior and enable open-ended cultural evolution.

As music has both universal (S. Brown & Jordania, 2013; Mehr et al., 2019; Savage et al., 2015) and idiosyncratic features, the origin of musical behavior could be considered in terms of Baldwinian evolution (Podlipniak, 2017, 2021; Savage et al., 2021a). After all, the Baldwinian model predicts that the behavior is partly predisposed and partly culture-dependent (Jablonka & Lamb, 2005). Moreover, since music is a multifaceted form of communication in which the different elements depend on different specialized abilities, then Baldwinian transformation could have happened many times. In line with this assumption, Savage et al. (2021) have proposed that the evolution of *musicality* should be considered in terms of the “iterated Baldwin effect” (p. 3), that is, a process in which culturally invented musical behavioral innovation creates a niche enabling the selection of genetically controlled elements of *musicality*, allowing subsequent innovation and so on, resulting in “a virtuous spiral” (Savage et al., 2021, p. 3). However, because the processing of music and language by modern human brains involves the same neural structures, at least to some extent (Steinbeis & Koelsch, 2008a), it would be useful to take into account the broadening of the musical niche into a communicative niche comprising both proto-musical and proto-lingual forms of communication. In this scenario, the cycle of plasticity and canalization that characterizes the iterated Baldwinian process went beyond the musical niche, drawing inspiration for innovations from both music-like and language-like behaviors. The proposed extension of the musical niche to a communicative niche assumes that the existing proto-musical cognitive tools could have been used by hominins to fulfil new communicative functions specific to a proto-language. Similarly, the cognitive tools specific to a proto-language could have been applied to music-like communication.

Developmental plasticity is required for exaptation of this nature (Hughes, 2012). The exaptation of cognitive tools involves implementing an existing neural submodule, or neural circuitry, in a functionally new module, or circuit. Such repurposing, to use Schlaudt’s term (2022), or *neuronal recycling* (Dehaene, 2005), demands experience-dependent plasticity and must be first achieved in the domain of culture. In some sense, a cultural niche tinkers with and creates a new device from pre-existing cognitive tools, rather like natural selection, which can be seen as a tinkerer using everything at their disposal to produce a useful tool (Jacob, 1977). The cultural repurposing of an existing neural circuit is therefore the first attempt to cope with the environmental challenge. If this attempt is successful, the next step is canalization by means of expensive cultural inheritance based on learning this repurposing. Since the cost of learning becomes a burden, in a population where a particular individual was accidentally endowed with the predisposition to learn a new behavior faster and with less effort, natural selection starts to favor this individual and then its progeny. Only then does

natural selection canalize a new behavioral trait, in the long term, by genetic inheritance. However, a novel behavior changes the niche, which creates new challenges leading to a new circle of plasticity, exaptation, and canalization. This is the iterated Baldwin effect described by Savage and his colleagues (2021a). In the case of vocal communication among early hominins, even as distant from *Homo sapiens* as *Ardipithecus ramidus* (Clark & Henneberg, 2017), this niche probably involved the use of different communicative tools, because culture usually tests different variants of behavior to achieve a particular goal (see, for example, the different methods of opening milk bottles used by certain species of birds such as *Parus major*, *Parus coeruleus*, and *Parus ater*; J. Fisher & Hinde, 1949; Hawkins, 1950; Jablonka & Lamb, 2005).

The two forms of human vocal communication

Singing and speaking, or more broadly, music and natural language, have often been understood as functionally different forms of communication and could have had their beginnings in hominin vocalizations. However, while individual vocalizations could have been the predecessors of speech, group vocalizations could have been more closely related to music (Jan, 2022), as music seems to be predominantly effective for producing different sounds simultaneously. In this case, vocalizations should have possessed certain features that could have acted as the anchor points for synchronization. In fact, singing, like instrumental music, uses rhythm and pitch to achieve this goal. Although interindividual synchronization does not necessarily involve the coordination of pitches in time, the use of pitch as an additional anchor point increases the complexity and thus the combinatoriality of a system. Auditory-motor synchronization (Zatorre et al., 2007) enables the coordination of matching sounds in time for the purposes of collective singing or playing music. When singers or instrumental musicians align pitches according to culturally specific rules, they can do this because they know how to imitate the fundamental frequency of harmonic sound (F_0) (Bannan, 2008, 2012) producing a range of textures: monophony, heterophony, homophony, and polyphony. In contrast, speech is mainly used in a responsorial way that involves sequential interactions (S. Brown, 2022b). This difference between music and speech can thus be described as the “choric/dialogic distinction” (Haiduk & Fitch, 2022, p. 1) respectively. However, it should not be treated as absolute in the sense that it precludes possible exceptions to the rules of collectivity and turn-taking. Probably because of the collective nature of music, songs differ from speech, mainly in that in singing the pitches are stable and discrete, while in speaking they are continuous (Zatorre & Baum, 2012). Conversely, people do sing alone such as in the shower (in the Western sense), or to keep themselves company (Falk, 2004), whilst lead vocalists are known to carry a solo melody 'alone' in ensemble contexts. Sometimes the pitch contour of a melodic phrase is not intervallic but plastic, enabling the use of melisma and glissandi. Pitch sliding can be choric over long passages with performers on individual parts, such as in the case of isophony (Nikolsky, 2018; Gill, 2023). Nevertheless, the distinction proposed by Haiduk and Fitch (2022) reflects general tendencies in the ways that sounds are organized in singing and speaking. Importantly, these two different types of vocal communication require specific characteristics of vocal learning (Merker, 2012). While speaking necessitates the imitation of distinctive spectral features of sounds, singing requires the volitional control of F_0 (Bannan, 2008, 2012).

As far as the content of communication is concerned, Shilton (2022) has indicated that, while natural language is focused on external objects, music acts as a tool of cooperative interactions by means of temporal and tonal alignment, as described above. In other words, natural language is oriented to extrinsic meaning, whereas music is connected to intrinsic meaning. These two types of meaning are relevant to the distinction between implicit and explicit knowledge (Schilhab & Gerlach, 2008). In both cases, communication leads to the synchronization of brain states (Abrams et al., 2013; Jiang et

al., 2012; Pérez et al., 2017). However, the communication of intrinsic meaning is achieved by directly eliciting motor and emotional states. In contrast, concepts have to be inferred from patterns of sound for the vocal communication of extrinsic meaning. This is not to say that intrinsic meaning does not influence extrinsic meaning. Being evolutionarily older, intrinsic meaning can scaffold conceptual meaning, as in the case of the influence of tactile sensations on conceptual knowledge (Ackerman et al., 2010) or the emotions induced by timbre (Wallmark et al., 2018; 2019). After all, emotions also serve as a mechanism for assessing the external world, that is, by efficiently rating the ecological relevance of sound sources (Ma & Thompson, 2015). However, the use of internally and externally oriented communication systems fulfil the different functions of motivating reactions to stimuli and creating a conscious model of reality, respectively. Since these two functions of vocal communication seem to be detached from each other among chimpanzees (Watson et al., 2015), one can assume that hominins used two types of vocal communication systems before language and music evolved, namely externally and internally oriented protolanguages (Podlipniak, 2022a). Today music and speech are typically accompanied by gestures, which has led to claims that both these communicative systems are integrated with gestural expressions (Kelly et al., 2010; Nussbaum, 2007). However, while affective gestures are present during both singing and speaking, semantic gestures (deictic, iconic, and symbolic) seem to dominate in language, which suggests that music differs from natural language also in the domain of gesturing. Although the division between music as an intrinsic communicative system and language as an extrinsic communicative system reflects the fundamental characteristics of these two types of human expression, this difference is not absolute. It should be emphasized that both natural language and music convey intrinsic and extrinsic meanings respectively. Every natural form of speech consists of suprasegmental (prosody) and segmental features (consonants and vowels) that transmit meaning in different ways. These two components of speech, nonverbal and verbal, are often described as affective vocalization and articulate speech, and are assumed to be based on two different brain pathways (Ackermann et al., 2014). Speech prosody, apart from its many possible contributions to propositional meaning, transmits affective meaning in a similar way to music, that is by eliciting emotions directly (Frühholz et al., 2014). In other ways, music is often reported as a source of referential meaning (S. Brown, 2022b; Cross, 2009; Cross & Woodruff, 2010; Jan, 2022; Koelsch, 2013; Koelsch et al., 2004; Patel, 2008; Tomlinson, 2023), and although musical semantics are usually thought to be much more ambiguous than the semantics of prose (Cross, 2005), neuroimaging studies have shown overlaps between the parts of the brain responsible for processing meaning in music and in language (Koelsch, 2005, 2011; Koelsch et al., 2004; Painter & Koelsch, 2011; Steinbeis & Koelsch, 2008b).

Common-precursor and multi-source models of music and language evolution

The observed overlaps between musical and speech communication have inspired many scholars to look for the common origin of language and music (Bannan, 2008; S. Brown, 2000, 2017, 2022b; Darwin, 1871; Fitch, 2013; Jan, 2022; Livingstone, 1973; Rousseau, 1998; Spencer, 1890). Many of these proposals have additionally assumed that the *musicality* and the faculty of language came into existence as the result of biological evolution (Bannan, 2008; S. Brown, 2000, 2017, 2022b; Darwin, 1871; Fitch, 2013; Jan, 2022). The dominant view among these hypotheses is a linear model in which language and music evolved from a common vocal precursor. The main premise for this explanation is based on the similarities between speech prosody and music (London, 2012; Palmer & Hutchins, 2006; Patel et al., 2006; Patel & Daniele, 2003), which include pitch contour, rhythm, stress, loudness, tempo, and pauses. The use of these features in order to express emotions, so-called affective prosody (or expressive dynamics), is characterized by many intercultural, and even interspecies similarities (Filippi, 2016, 2020; Filippi et al., 2017; Merker, 2003; Zimmermann et al., 2013). Speech and music also seem to possess a common first phase in their ontogenetic development (Brandt et

al., 2012; McMullen & Saffran, 2004). Having a common developmental origin and shared prosodic features can be explained by the descent from an evolutionary common precursor. After all, homologies are evidence of shared ancestry. The common-precursor models assume that after a *musilanguage* phase (S. Brown, 2000) music and language started to evolve separately. In other words, after the split from a common precursor the evolution of *musicality* and language faculty took disparate and independent, uni-domain evolutionary paths. The exception to this standard view is Brown's theory (2022b) that a protomusic co-opted a rhythmic system that evolved independently from music as a part of dance (S. Brown, 2022a). He has not explained, however, what mechanism led to this co-optation.

An alternative explanation suggests that, rather than having one common precursor, music and language have multiple sources. According to this multi-source explanation, hominins created a communicative niche consisting partly of instinctive affective prosody combined with affective gestures, and partly of culturally invented signals such as iconic and symbolic gestures and sounds. As communication about internal states and external objects have different functions, the two forms of communication, music-like (internally oriented) and language-like (externally oriented), could initially have evolved separately (Podlipniak, 2022a). However, as the social niche started to become more and more complex, thereby creating new challenges, existing forms of communication appeared insufficient. Thus instead of expanding the existing music-like and language-like communicative tools by introducing new features, hominins could have begun to use elements of one form of communication to enhance the communication capabilities of the other. It is well known that speech prosody can influence social interactions by conveying clues to the speaker's internal state, such as politeness, impoliteness, dominance, or submissiveness (P. Brown & Levinson, 1987; Culpeper, 2011; Culpeper et al., 2003; Ponsot et al., 2018); these clues can affect the way that lexical and grammatical content is interpreted by the listener. Interpretation is also likely to be influenced by the speaker's affective gestures, which can function as pragmatic gestures (Lopez-Ozieblo, 2020). It is therefore plausible that hominins could have used music-like tools for communicating internal states to enhance their use of language-like tools for communicating external social relations. Natural selection could have favored such repurposing because it is more economical than creating new structures. The examples of interactions between music and language that can be observed in contemporary cultures, and interpreted as repurposing, suggest that a similar process could have happened in our species' ancestral-hominin cultures.

Cross-domain interactions between modern communicative phenomena

The observed differences between the phonological systems of contemporary languages and music seem to be the results of experience-expectant plasticity, as people acquire these systems via implicit learning during childhood (McMullen & Saffran, 2004). The appearance of some forms of communication in certain populations must however have demanded experience-dependent plasticity. Good examples of such communication systems are whistled (Meyer, 2008, 2015) and drum languages (Akinbo, 2021; Arewa & Adekola, 1980; Seifart et al., 2018). In both cases, music-specific elements such as pitch and rhythm are used to code speech-specific features to convey propositional meaning. The users of whistled and drum languages can transmit propositional meaning by emulating the tonal and rhythmic patterns of spoken language using sound sequencing (Akinbo, 2021; Seifart et al., 2018) and also, in the case of whistled languages, the spectral characteristics of harmonic sounds (F_0 and formants) (Meyer, 2015). The invention of both of these communication systems was probably a way of overcoming the limitations of spoken language related to the short range of the propagation of speech sounds. Deserving of special attention here, however, is that behavioral plasticity consists in this instance of using the resources of an existing system to perform a function

specific to another system. This re-use or repurposing probably results in the reorganization of certain neural circuits. It has been discovered, for instance, that native users of a whistled language in the mountains of Northeast Turkey exhibit a decrease in left-hemisphere and increase in right-hemisphere activity, such that symmetric hemispheric processing can be observed when they are listening to and understanding the whistled language as opposed to speech (Güntürkün et al., 2015).

People who are native speakers of tonal and non-tonal languages also exhibit lateralization differences (Wang et al., 2004). Left-hemispheric lateralization of lexical tone processing has been found in native tonal-language speakers (Chien et al., 2020; Gu et al., 2013). There are also structural differences between the brains of native- and non-native tonal-language speakers such that the former have a greater density of gray and white matter near the right anterior temporal lobe and the left insula (Crinion et al., 2009). Except when used as part of prosody, relative changes in pitch become phonological features in tonal languages and are used as cues influencing the meaning of words (Maddieson, 2005) and/or their grammatical relationships. Although a lexical tone is not an interval between two musical pitches, musical training enhances the recognition of lexical tones (Patel & Iversen, 2007; Wong et al., 2007); this suggests a possible interaction between music and speech in the cultural development of these communicative systems. The differences between the ways in which tonal- and non-tonal languages are processed shows that the choice of a particular culture, in this case whether or not to use pitch as part of a language phonological system, can induce the reorganization of neural circuitry via experience-expectant plasticity. It could also be that the use of tonal speech in a cultural environment was responsible for producing particular genotypes, because correlations have been found between the population frequency of a variant of the development-related gene *ASPM* and the distribution of tonal languages in that population (Dediu, 2021; Dediu & Ladd, 2007), and between the same gene and the ability to perceive lexical tones (Wong et al., 2012, 2020). Thus Baldwinian evolution could contribute to language diversity. Another well-documented reorganization of brain circuitry in response to cultural demands concerns differences between musicians and non-musicians (Leipold et al., 2021). In fact, the changes in musicians' brain networks are a widely used example of experience-dependent plasticity (Leipold et al., 2021; Münte et al., 2002). An important behavioral difference between musicians and non-musicians is that musicians are trained to organize their perception of musical structure by using concepts, often reflected in music notation, such as particular pitch intervals, rhythm values, and tonal functions, while non-musicians are unaware of these constraints when they listen to music. Since conceptual thinking is not a domain of musical communication this behavioral difference can be interpreted as a result of a conceptualization of music that has been imposed culturally (Zbikowski, 2002).

Culture can have a negative impact on the various forms of vocalization described above by suppressing the development of cognitive abilities to which human beings are predisposed. One example related to vocalization and pitch perception is octave equivalence, or the ability to perceive that two pitches an octave apart are similar (Hoeschele et al., 2012). This ability is widespread among human populations, and it has been claimed that it is a universal feature of music perception (S. Brown & Jordania, 2013; Harwood, 1976). In one study, however, researchers played sequences of sounds to indigenous Tsimane people from the Amazonian rainforest and North Americans. Unlike the latter, the Tsimane participants ignored octave similarity when they reproduced the sequences. This may indicate that octave equivalence depends on the experience of culture-specific music. Or perhaps the Tsimane participants ignored it because their own music "appears to lack group performance and harmony" (Jacoby et al., 2019, p. 3230; see also McDermott et al., 2016). It has recently been suggested that octave equivalence originates in social bonding produced by chorusing (Bannan et al., 2022), so the non-communal use of singing can be interpreted as a culturally driven

change of music function leading to the suppression of a predisposition to develop octave equivalence. In this case, the trajectory of cultural evolution changes the cultural communicative niche in such a way that natural selection no longer prefers individuals who experience pitches an octave apart as perceptually similar. A comparable effect whereby the ability to recognize tone in speech is suppressed can be induced by an environment in which language is non-tonal. A similar process may have led to the suppression of perfect pitch at the expense of developing linguistic abilities, as Steven Mithen (2006) suggests. Nonetheless, the existence of cultural repurposing or suppression of cognitive abilities related to communication observed today suggests that the same effects could have influenced the evolution of music and language in the past too.

Cultural invention, canalization, and exaptation as drivers in the evolution of *musicality* and language faculty

An intriguing feature of the functional organization of the brain is activity in the structures responsible for the processing of language syntax during the processing of music (Koelsch, 2005; Li et al., 2023; Steinbeis & Koelsch, 2008a). In order to explain overlaps between these structures, Patel (2003) has proposed the shared syntactic integration resource hypothesis, also known as the resource-sharing framework (Patel, 2011). In these models the shared neural structures (resources) operate on domain-specific knowledge localized in different areas of the brain. The widely accepted view is that musical syntax is a result of employing syntactic abilities that evolved originally as a part of the language faculty (Lerdahl & Jackendoff, 1983; Patel, 2008; Pinker, 1997). These shared neural resources are therefore seen as elements of a cognitive mechanism underlying language. In an alternative model, in which specialized structures are used by a variety of neural networks (Peretz et al., 2015), shared neural resources can be viewed as an integrated part of a music-specific network (although for yet another explanation, see Asano et al., 2022). The incorporation of these resources into a music-specific network can be explained by culturally driven repurposing, as described above, and also probably by canalization through natural selection. From the psychological point of view, the main difference between the functions of the language-specific and music-specific syntactic systems is that patterns of sound are mapped on to a hierarchy of concepts in the language-specific system, and on to pre-conceptual hierarchical experiences of stability in the music-specific system. This difference could also be seen in terms of the use of hierarchical control over abstract rules in the case of language and patterns of stability and instability in motor networks linked to emotions (Asano et al., 2021) in the case of music. Considering that pre-conceptual experiences of stability were probably part of a form of vocal communication that is evolutionarily older than language, it is more likely that shared neural resources originated in and were incorporated from a music-like communication system into a language-like communication system than vice versa (Podlipniak, 2023). In this view, pitch hierarchy evolved before language grammar (Podlipniak, 2016), probably with increased control of the larynx, subglottal system and supralaryngeal tract in *Homo ergaster*, affording later-evolving species (e.g., *Homo erectus*, *Homo heidelbergensis*, and the Neanderthals), the ability, like *H. sapiens*, to control a melody consciously (Deacon, 2000; Morley, 2013, 2014; Wurz, 2009). It may be, however, that the pre-conceptual experiences of stability and instability that became the basis for a syntactic hierarchy in a music-like communication system evolved earlier, as sensations accompanying the motor expression of rhythm patterns by gestures and vocalizations. In that case, considering that vocal abilities began to increase in *Ardipithecus ramidus* (Clark & Henneberg, 2017), it is even possible that the first sound hierarchies appeared before the evolution of the *Homo* genus. Regardless of which proximal function of syntax evolved first (i.e., expressing the hierarchy of concepts in language or pre-conceptual experiences of stability in music), the exaptation of cognitive machinery from one of these communication systems to the other is a convincing explanation for the evolution of resource-sharing networks captured in Patel's model. Importantly, the main results of

this cross-domain coevolutionary interaction are the canalized syntactic properties of modern language and music.

Besides octave equivalence, relative pitch is a cognitive ability that could have been repurposed by hominins to fulfil a new function. This ability allows us to recognize transposed melodies as being examples of the same prototype independent of differences of pitch between the original and transposed melodies. Patel (2023) has speculated that relative pitch could have evolved as a speech-specific specialization and was only later employed in the context of music. However, relative pitch could also have evolved because of its adaptive value to music-like emotional vocalizations consisting of a primitive pitch hierarchy based on pre-conceptual sensations of stability. It could then have been used in language-like propositional vocalizations as a result of social invention. For example, a pitch contour contributing to a music-like emotional vocalizations could have become, independent of its absolute pitch, not only a precursor of speech prosody but could have also been used by early hominins in language-like propositional vocalizations in order to indicate attitudes to events or other hominins. In a broad sense, these attitudes – referred to in linguistics as grammatical *mood* (Gil, 2021) (e.g., interrogative or imperative, referring to questions and statements respectively) – resemble internal preconceptual states (e.g., questions representing uncertainty and statements representing certainty) thus producing the original content of music-like vocalizations. This was probably not accidental. Nowadays, mood is typically conveyed by intonation in the majority of languages (Jun, 2005; Warren & Calhoun, 2021), suggesting that intonation is a tool that has been canalized in the course of evolution. In other words, the use of pitch in spoken language to indicate cues or questions (Chien et al., 2020; Gussenhoven, 2016; Gussenhoven & Chen, 2000; Jun, 2005; Ma et al., 2011) could have been adopted by hominins from music-like vocalizations and then canalized via the Baldwinian process.

Overlapping brain structures do not only explain the processing of musical and linguistic syntax and prosody but also musical and linguistic semantics (Koelsch, 2005, 2011; Painter & Koelsch, 2011; Steinbeis & Koelsch, 2008b). The extent to which music has meaning is an intriguing question that is much debated. It is generally agreed that music lacks propositional semantics (Lerdahl, 2013) but many examples of specific pieces are interpreted as conveying an intersubjectively consistent extra-musical meaning (Koelsch, 2013; Patel, 2008). The role of emotions in creating concepts may hold a clue to the emergence of the propositional meaning of music. The neural processing of emotional and semantic information, respectively, converges in the lateral zone of the inferior frontal gyrus pars orbitalis (Belyk et al., 2017), suggesting that emotions were implicated in the evolution of propositional semantics. Importantly, this structure is also a part of the affective prosody (Belyk & Brown, 2014) and music production networks (Bianco et al., 2022), indicating a potential link between the expression of emotions and semantic signals. In line with this view, Filippi (2020) proposes that vocal emotional expressions facilitated the evolution of language semantics because vocal patterns became associated with meaning. Alternatively, this process could be explained in terms of a language-like communication system repurposing elements of music-like communication. Since the emotional pre-conceptual sensations that accompany the experience of music can be divided into relatively distinct sets of emotions such as fear, pleasure, anger, joy, and sadness, it can be speculated that these sensations informed the development of these concepts in language.

A music-like communication system could have repurposed elements of language-like communication too. Spoken language consists of sequences of vowels and consonants that we recognize based on the spectral characteristics of the sounds. It is therefore the changing spectral characteristics that make speech so effective in conveying propositional meaning. From this perspective, the experience of listening to spoken words is similar to the experience of listening to a series of fluctuating timbres.

Although it is traditionally assumed that the relationship between the sounds of words and their meaning is arbitrary (Saussure, 1959), several studies have shown that certain psychoacoustic features of the sounds of words can have a universal, non-arbitrary relationship to their meaning (D'Anselmo et al., 2019; Dingemanse et al., 2015; Erben Johansson et al., 2020; Monaghan et al., 2014; Preziosi & Coane, 2017). This phenomenon is called *sound symbolism*; Imai and Kita (2014) claim that it is an important part of natural language that can shed light on the evolution of meaning. Non-arbitrary sound-meaning mappings seem to be deeply engrained in human cognition (Erben Johansson et al., 2020) and may represent the canalized vestiges of the language-like propositional tool of communication used by hominins. Since timbre is important in music, and also conveys iconicity and symbolism, it can be speculated that music owes sound symbolism to the repurposing of non-arbitrary sound-meaning mappings from language-like propositional vocalizations. It is also probable that the associative tendencies governing sound symbolism in speech are also present in music. For example, there is evidence of cross-modal correspondences between timbre and the visual or tactile domains (Wallmark, 2019; Wallmark & Allen, 2020).

Propositional meaning can also be conveyed by gestures. This being the case, it can be asked if propositional meaning emerged in music because elements of motor behaviors were repurposed in the auditory domain or vice versa. The findings of research involving the participation of individuals who stammer provide an answer. Stammering reduces the number of gestures accompanying speech (Mayberry & Jaques, 2010) but does not affect singing (Wan et al., 2010). This may be because gestures have different roles in speech and music. Affective gestures were repurposed by language-like communication from music-like communication, but not the other way round. If the communicative niche proposed by Brown (2022a) included dance, which evolved separately from music-like vocalizations, the affective gestures repurposed by a language-like communication could have had their roots in motor behavior specific solely for dance.

I propose that all the cross-domain interactions I have described are important in the evolutionary history of both human *musicality* and the faculty of language. The evolutionary process common to both is as follows: cultural invention of a particular behavior in one domain; canalization of this behavior as a part of this domain; cultural exaptation of this behavior into another domain; and finally canalization of this behavior as a part of this new domain. If all these new culturally canalized behaviors were adaptive, they could have undergone Baldwinian evolution whereby an individual genetically predisposed to learn a particular behavior faster and with less effort would be born sooner or later. The predisposition to learn behaviors involving cross-domain interactions would set the individual on a genetically driven developmental path in which interdomain neural connections would lead to the emergence of a new functionally specialized neural network with its origins in developmental plasticity. Considering that the processing of modern language and music involves many shared structures in the brain, the interactions I have proposed are likely candidates for explaining their evolutionary origin.

Conclusions

My argument has focused on the role of cultural flexibility in the evolution of human *musicality* and the language faculty as a part of a dynamic process of interaction between these two sets of communicative abilities. The examples of potential interactions and repurposing that I have presented are not, of course, exhaustive. I could also have described cross-domain interactions in relation, for example, to the volitional control of affective prosody in speech; the perception that pitches, rhythms, and phonemes are discrete; and the combinatoriality of discrete musical and speech units. Moreover, the communicative niche of early hominins was probably not restricted to vocal modes of communication. The most useful and opportunistic strategy for hominins to exchange

information was probably to combine auditory and visual signals (Zlatev et al., 2020). The fact that both singing and speaking are typically accompanied by involuntary gestures suggests that gestural modes of communication could also have interacted with both the music-like and language-like vocalizations of early hominins. Visual signals as the source of iconicity and symbolism could have been invented primarily in the gestural domain in the form of pantomime, which is also a good candidate precursor of semantics (Zlatev et al., 2017).

My argument could be criticized for contradicting the principle of parsimony, which states that the most convincing explanation should involve the fewest assumptions or entities that fit the evidence. This principle is also applied in evolutionary biology (Sober, 1988). However, the reconstruction of phylogeny by means of the shortest evolutionary tree, according to the principle of parsimony, does not necessarily reflect the actual phylogeny (Stewart, 1993). I have presented a number of examples of repurposing in contemporary culture, and of overlaps between the neural processing of music and language. Together, they form the basis of my premise that cross-domain interactions are an important mechanism underlying the evolution of both human *musicality* and the faculty of language. Theories of the evolution of music, along Baldwinian lines, are incomplete if cross-domain interactions are not considered. More research is needed, of course, to elucidate the possible cross-domain evolutionary paths that have led to the emergence of *musicality* as we know it today. Nevertheless, the cross-domain coevolutionary interactions that, according to my proposal, drove the evolution of human *musicality* and the faculty of language bring us closer to understanding the intricate phylogenetic relationships that exist between music and natural language.

Acknowledgments

I would like to thank two anonymous reviewers for their many useful comments and suggestions. I would also like to thank Jane Ginsborg for her helpful advice as well as Peter Kośmider-Jones for his language consultation on the first draft of this manuscript.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

References

- Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience*, 37(9), 1458–1469. <https://doi.org/10.1111/ejn.12173>
- Ackerman, J. M., Nocera, C. C., & Bargh, J. A. (2010). Incidental haptic sensations influence social judgments and decisions. *Science*, 328(5986), 1712–1715. <https://doi.org/10.1126/science.1189993>
- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, 37(06), 529–546. <https://doi.org/10.1017/S0140525X13003099>
- Akinbo, S. K. (2021). The language of Gángan, a Yorùbá talking drum. *Frontiers in Communication*, 6, Article 180. <https://doi.org/10.3389/fcomm.2021.650382>

- Arewa, O., & Adekola, N. (1980). Redundancy principles of statistical communications as applied to Yoruba talking-drum. *Anthropos*, 75(1/2), 185–202. <http://www.jstor.org/stable/40460588>
- Asano, R., Boeckx, C., & Fujita, K. (2022). Moving beyond domain-specific versus domain-general options in cognitive neuroscience. *Cortex*, 154, 259–268. <https://doi.org/https://doi.org/10.1016/j.cortex.2022.05.004>
- Asano, R., Boeckx, C., & Seifert, U. (2021). Hierarchical control as a shared neurocognitive mechanism for language and music. *Cognition*, 216, Article 104847. <https://doi.org/https://doi.org/10.1016/j.cognition.2021.104847>
- Avital, E., & Jablonka, E. (2000). *Animal traditions: Behavioural inheritance in evolution*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511542251>
- 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. (2008). 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., Bamford, J., & Dunbar, R. I. M. (2022). The evolution of gender dimorphism in the human voice: the role of octave equivalence. *PsyArXiv*, <https://psyarxiv.com/f4j6b/>.
- Bateson, G. (1979). *Mind and nature: A necessary unity*. Dutton.
- Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, 9(9), 1395–1403. <https://doi.org/10.1093/scan/nst124>
- Belyk, M., Brown, S., Lim, J., & Kotz, S. A. (2017). Convergence of semantics and emotional expression within the IFG pars orbitalis. *NeuroImage*, 156, 240–248. <https://doi.org/10.1016/j.neuroimage.2017.04.020>
- Bianco, R., Novembre, G., Ringer, H., Kohler, N., Keller, P. E., Villringer, A., & Sammler, D. (2022). Lateral prefrontal cortex is a hub for music production from structural rules to movements. *Cerebral Cortex*, 32(18), 3878–3895. <https://doi.org/10.1093/cercor/bhab454>
- Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology*, 3, Article 327. <https://doi.org/10.3389/fpsyg.2012.00327>
- Brown, P., & Levinson, S. C. (1987). Politeness: Some universals in language usage. In *Politeness: Some universals in language usage*. Cambridge University Press.
- Brown, S. (2000). The ‘musilanguage’ model of musical evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). The MIT Press.
- Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, 8, Article 1894. <https://doi.org/10.3389/fpsyg.2017.01894>
- Brown, S. (2022a). Group dancing as the evolutionary origin of rhythmic entrainment in humans. *New Ideas in Psychology*, 64, Article 100902. <https://doi.org/10.1016/j.newideapsych.2021.100902>

- Brown, S. (2022b). *The unification of the arts: A framework for understanding what the arts share and why*. Oxford University Press. <https://doi.org/10.1093/oso/9780198864875.001.0001>
- Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, 41(2), 229–248. <https://doi.org/10.1177/0305735611425896>
- Chien, P.-J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*, 41(7), 1842–1858. <https://doi.org/10.1002/hbm.24916>
- Clark, G., & Henneberg, M. (2017). *Ardipithecus ramidus* and the evolution of language and singing: An early origin for hominin vocal capability. *HOMO*, 68(2), 101–121. <https://doi.org/10.1016/j.jchb.2017.03.001>
- Crinion, J. T., Green, D. W., Chung, R., Ali, N., Grogan, A., Price, G. R., Mechelli, A., & Price, C. J. (2009). Neuroanatomical markers of speaking Chinese. *Human Brain Mapping*, 30(12), 4108–4115. <https://doi.org/10.1002/hbm.20832>
- Cross, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical Communication* (pp. 27–44). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198529361.003.0002>
- Cross, I. (2009). The evolutionary nature of musical meaning. *Musicae Scientiae*, 13(2), 179–200. <https://doi.org/10.1177/1029864909013002091>
- Cross, I., & Woodruff, G. E. (2010). Music as a communicative medium. In *The Prehistory of Language* (pp. 77–98). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199545872.003.0005>
- Culpeper, J. (2011). "It's not what you said, it's how you said it!" Prosody and impoliteness". In Linguistic Politeness Research Group (Ed.), *Discursive approaches to politeness* (pp. 57–84). De Gruyter Mouton. <https://doi.org/10.1515/9783110238679.57>
- Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, 35(10), 1545–1579. [https://doi.org/10.1016/S0378-2166\(02\)00118-2](https://doi.org/10.1016/S0378-2166(02)00118-2)
- D'Anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing meaning from word sounds of unfamiliar languages: A cross-cultural sound symbolism study. *Frontiers in Psychology*, 10, Article 593. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00593>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex* (1st ed.). John Murray.
- Deacon, T. W. (2000). Evolutionary perspectives on language and brain plasticity. *Journal of Communication Disorders*, 33(4), 273–291. [https://doi.org/10.1016/S0021-9924\(00\)00025-3](https://doi.org/10.1016/S0021-9924(00)00025-3)
- Dediu, D. (2021). Tone and genes: New cross-linguistic data and methods support the weak negative effect of the "derived" allele of ASPM on tone, but not of Microcephalin. *PLOS ONE*, 16(6), Article e0253546. <https://doi.org/10.1371/journal.pone.0253546>
- Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, 104(26), 10944–10949. <https://doi.org/10.1073/pnas.0610848104>
- Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The 'neuronal recycling' hypothesis. In S. Dehaene, J.-R. Duhamel, M. D. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain: A fyxsen foundation symposium*. The MIT Press. <https://doi.org/10.7551/mitpress/3136.003.0012>

- Dingemans, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615. <https://doi.org/10.1016/J.TICS.2015.07.013>
- Dissanayake, E. (2001). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). MIT Press.
- Dor, D., & Jablonka, E. (2010). Plasticity and canalization in the evolution of linguistic communication: an evolutionary developmental approach. In R. K. Larson, V. Deprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 135–147). Cambridge University Press. <https://doi.org/10.1017/CBO9780511817755.010>
- Dor, D., & Jablonka, E. (2014). Why we need to move from gene-culture co-evolution to culturally driven co-evolution. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language* (pp. 14–30). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199665327.003.0002>
- Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.
- Dunbar, R. I. M. (2012). On the evolutionary function of song and dance. In N. Bannan (Ed.), *Music, language, and human evolution* (pp. 201–214). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199227341.003.0008>
- Erben Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). *The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features*. 24(2), 253–310. <https://doi.org/doi:10.1515/lingty-2020-2034>
- Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, 27(2004), 491–541. <https://doi.org/10.1017/S0140525X04000111>
- Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. *Frontiers in Psychology*, 7, Article 1393. <https://doi.org/10.3389/fpsyg.2016.01393>
- Filippi, P. (2020). Emotional voice intonation: A communication code at the origins of speech processing and word-meaning associations? *Journal of Nonverbal Behavior*, 44(4), 395–417. <https://doi.org/10.1007/s10919-020-00337-z>
- 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), Article 20170990. <https://doi.org/10.1098/rspb.2017.0990>
- Fisher, J., & Hinde, R. A. (1949). The opening of milk bottles by birds. *British Birds*, 42, 47–57.
- Fisher, S. E., & Ridley, M. (2013). Culture, genes, and the human revolution. *Science*, 340(6135), 929–930. <https://doi.org/10.1126/science.1236171>
- Fitch, W. T. (2013). Musical protolanguage: Darwin's theory of language evolution revisited. In J. J. Bolhuis & M. Everaert (Eds.), *Birdsong, speech, and language: exploring the evolution of mind and brain* (pp. 489–503). The MIT Press.
- Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664), Article 20140091. <https://doi.org/10.1098/rstb.2014.0091>
- Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, 123, 1–17.

<https://doi.org/10.1016/j.pneurobio.2014.09.003>

- Gil, D. (2021). Tense–aspect–mood marking, language-family size and the evolution of predication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1824), Article 20200194. <https://doi.org/10.1098/rstb.2020.0194>
- Gill, F. (2023). Finding wolves [Music recorded by Frances Flute the Bellows Mender]. On *Sonic Debitage*, Bandcamp.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation-A missing term in the science of form. *Paleobiology*, 8(1), 4–15. <http://www.jstor.org/stable/2400563>
- Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development*, 58(3), 539–559. <https://doi.org/10.2307/1130197>
- Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *NeuroImage*, 83, 637–645. <https://doi.org/10.1016/j.neuroimage.2013.02.080>
- Güntürkün, O., Güntürkün, M., & Hahn, C. (2015). Whistled Turkish alters language asymmetries. *Current Biology*, 25(16), R706–R708. <https://doi.org/10.1016/j.cub.2015.06.067>
- Gusenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8(2), 425–434. <https://doi.org/10.1111/tops.12197>
- Gusenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In D. Guan (Ed.) *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP)*, Vol. 2 (pp. 91–94). China Military Friendship Publish. <https://doi.org/10.21437/ICSLP.2000-216>
- 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), 291–320. <https://doi.org/10.1177/1029864909013002131>
- Haiduk, F., & Fitch, W. T. (2022). Understanding design features of music and language: The choric/dialogic distinction. *Frontiers in Psychology*, 13, Article 786899. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.786899>
- Hall, B. K. (2001). Organic selection: Proximate environmental effects on the evolution of morphology and behaviour. *Biology and Philosophy*, 16(2), 215–237. <https://doi.org/10.1023/A:1006773408919>
- Harrison, P. M. C., & Seale, M. (2021). Against unitary theories of music evolution. *Behavioral and Brain Sciences*, 44, Article e76. <https://doi.org/DOI: 10.1017/S0140525X20001314>
- Harwood, D. L. (1976). Universals in music: A perspective from cognitive psychology. *Ethnomusicology*, 20(3), 521–533. <https://doi.org/10.2307/851047>
- Hawkins, T. H. (1950). Opening of milk bottles by birds. *Nature*, 165(4194), 435–436. <https://doi.org/10.1038/165435a0>
- Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, & Psychophysics*, 74(8), 1742–1760. <https://doi.org/10.3758/s13414-012-0364-2>
- Honing, H. (2018). On the biological basis of musicality. *Annals of the New York Academy of Sciences*,

1423(1), 51–56. <https://doi.org/10.1111/nyas.13638>

- Hughes, A. L. (2012). Evolution of adaptive phenotypic traits without positive Darwinian selection. *Heredity*, 108(4), 347–353. <https://doi.org/10.1038/hdy.2011.97>
- Hui, J., Cashman, T., & Deacon, T. (2008). Bateson's method: Double description. What is it? How does it work? What do we learn? In J. Hoffmeyer (Ed.), *A legacy for living systems: Gregory Bateson as precursor to biosemiotics* (pp. 77–92). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6706-8_6
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130298–20130298. <https://doi.org/10.1098/rstb.2013.0298>
- Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT Press.
- Jacob, F. (1977). Evolution and tinkering. *Science*, 196(4295), 1161–1166. <https://doi.org/10.1126/science.860134>
- Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, 29(19), 3229–3243. <https://doi.org/10.1016/j.cub.2019.08.020>
- Jaques, J., & Mayberry, R. I. (2000). Gesture production during stuttered speech: insights into the nature of gesture–speech integration. In D. McNeill (Ed.), *Language and gesture* (pp. 199–214). Cambridge University Press. <https://doi.org/10.1017/CBO9780511620850.013>
- Jan, S. B. (2018). “The two brothers”: Reconciling perceptual-cognitive and statistical models of musical evolution. *Frontiers in Psychology*, 9, Article 344. <https://doi.org/10.3389/fpsyg.2018.00344>
- Jan, S. B. (2022). *Music in evolution and evolution in music*. Open Book Publishers.
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(45), 16064–16069. <https://doi.org/10.1523/JNEUROSCI.2926-12.2012>
- Jordania, J. (2011). *Why do people sing? Music in human evolution*. Logos.
- Jun, S.-A. (2005). Prosodic Typology. In S.-A. Jun & S.-A. Jun (Eds.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0016>
- Juslin, P. N. (2021). Mind the gap: The mediating role of emotion mechanisms in social bonding through musical activities. *Behavioral and Brain Sciences*, 44, Article e80. <https://doi.org/10.1017/S0140525X2000120X>
- Kelly, S. D., & Ngo Tran, Q.-A. (2023). Exploring the emotional functions of co-speech hand gesture in language and communication. *Topics in Cognitive Science*, Article 12657. <https://doi.org/10.1111/tops.12657>
- 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->

- Killin, A. (2018). The origins of music: Evidence, theory, and prospects. *Music & Science*, *1*, 2059204317751971. <https://doi.org/10.1177/2059204317751971>
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, *15*(2), 207–212. <https://doi.org/10.1016/j.conb.2005.03.005>
- Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. <https://doi.org/10.1016/j.plrev.2011.04.004>
- Koelsch, S. (2013). *Brain and music*. Wiley-Blackwell. <https://doi.org/10.1017/CBO9781107415324.004>
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307. <https://doi.org/10.1038/nn1197>
- Laland, K. N. (2017). *Darwin's unfinished symphony: how culture made the human mind*. Princeton University Press.
- Leipold, S., Klein, C., & Jäncke, L. (2021). Musical expertise shapes functional and structural brain networks independent of absolute pitch ability. *The Journal of Neuroscience*, *41*(11), 2496–2511. <https://doi.org/10.1523/JNEUROSCI.1985-20.2020>
- 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), Article 20200391. <https://doi.org/10.1098/rstb.2020.0391>
- Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, music, and the brain* (pp. 257–272). The MIT Press. <https://doi.org/10.7551/mitpress/9780262018104.003.0010>
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. The MIT Press.
- Li, D., Wang, X., Li, Y., Song, D., & Ma, W. (2023). Resource sharedness between language and music processing: An ERP study. *Journal of Neurolinguistics*, *67*, Article 101136. <https://doi.org/https://doi.org/10.1016/j.jneuroling.2023.101136>
- Livingstone, F. B. (1973). Did the Australopithecines sing? *Current Anthropology*, *14*(1/2), 25–29. <https://doi.org/10.1086/201402>
- London, J. (2012). Three things linguists need to know about rhythm and time in music. *Empirical Musicology Review*, *7*(1–2), 5–11.
- Lopez-Ozieblo, R. (2020). Proposing a revised functional classification of pragmatic gestures. *Lingua*, *247*, Article 102870. <https://doi.org/10.1016/j.lingua.2020.102870>
- 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>
- Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America*, *129*(2), 1012–1023. <https://doi.org/10.1121/1.3531840>
- Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proceedings of the National Academy of Sciences*, *112*(47), 14563–14568. <https://doi.org/10.1073/PNAS.1515087112>

- Maddieson, I. (2005). Tone. In M. Haspelmath & H.-J. Bibiko (Eds.), *The world atlas of language structures* (pp. 58–61). Oxford University Press.
- McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550. <https://doi.org/10.1038/nature18635>
- McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289–311. <https://doi.org/10.1525/mp.2004.21.3.289>
- 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., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*, Article e60. <https://doi.org/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), Article eaax0868. <https://doi.org/10.1126/science.aax0868>
- Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. *Current Biology*, *28*(3), 356–368. <https://doi.org/10.1016/j.cub.2017.12.042>
- 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. (2003). Is there a biology of music? And why does it matter? In R. Kopiez, A. C. Lehmann, I. Wolther, & C. Wolf (Eds.), *Proceedings of the 5th triennial ESCOM conference* (pp. 402–405). Hanover University of Music and Drama.
- Merker, B. (2002). Music: The missing Humboldt system. *Musicae Scientiae*, *6*, 3–21. <https://doi.org/10.1177/102986490200600101>
- Mery, F., & Burns, J. G. (2010). Behavioural plasticity: an interaction between evolution and experience. *Evolutionary Ecology*, *24*(3), 571–583. <https://doi.org/10.1007/s10682-009-9336-y>
- Meyer, J. (2008). Typology and acoustic strategies of whistled languages: Phonetic comparison and perceptual cues of whistled vowels. *Journal of the International Phonetic Association*, *38*(1), 69–94. <https://doi.org/10.1017/S0025100308003277>
- Meyer, J. (2015). *Whistled languages: a worldwide inquiry on human whistled speech*. Springer-Verlag.
- 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>
- Mithen, S. J. (2006). *The singing Neanderthals: the origins of music, language, mind, and body*. Harvard University Press.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), Article 20130299. <https://doi.org/10.1098/rstb.2013.0299>

- Morley, I. (2013). *The prehistory of music: human evolution, archaeology, and the origins of musicality*. Oxford University Press.
- Morley, I. (2014). A multi-disciplinary approach to the origins of music: perspectives from anthropology, archaeology, cognition and behaviour. *Journal of Anthropological Sciences*, 92, 147–177.
- Münste, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6), 473–478. <https://doi.org/10.1038/nrn843>
- Nikolsky, A. (2018). Commentary: The 'musilanguage' model of language evolution. *Frontiers in Psychology*, 9, Article 75.
- Nussbaum, C. O. (2007). *The musical representation: meaning, ontology, and emotion*. MIT Press.
- Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology*, 48(5), 645–655. <https://doi.org/10.1111/j.1469-8986.2010.01134.x>
- Palmer, C., & Hutchins, S. B. T.-P. of L. and M. (2006). What is musical prosody? *Psychology of Learning and Motivation*, 46, 245–278. [https://doi.org/10.1016/S0079-7421\(06\)46007-2](https://doi.org/10.1016/S0079-7421(06)46007-2)
- 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. (2008). *Music, language, and the brain*. Oxford University Press.
- Patel, A. D. (2011). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204–223). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199553426.003.0022>
- 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), Article 20200326. <https://doi.org/10.1098/rstb.2020.0326>
- Patel, A. D. (2023). Human musicality and gene-culture coevolution: Ten concepts to guide productive exploration. In E. H. Margulis, P. Loui, & D. Loughridge (Eds.), *The science-music borderlands: Reckoning with the past and imagining the future* (pp. 15–37). The MIT Press. <https://doi.org/10.7551/mitpress/14186.003.0006>
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87(1), B35–B45. [https://doi.org/10.1016/S0010-0277\(02\)00187-7](https://doi.org/10.1016/S0010-0277(02)00187-7)
- Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9), 369–372. <https://doi.org/10.1016/j.tics.2007.08.003>
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5), 3034–3047. <https://doi.org/10.1121/1.2179657>
- Peretz, I., Vuvar, D. T., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370, Article 20140090. <https://doi.org/10.1098/rstb.2014.0090>

- Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports*, 7(1), Article 4190. <https://doi.org/10.1038/s41598-017-04464-4>
- Pigliucci, M. (2001). Characters and environments. In G. P. Wagner (Ed.), *The character concept in evolutionary biology* (pp. 363–388). Academic Press. <https://doi.org/10.1016/B978-012730055-9/50028-8>
- Pinker, S. (1994). *The language instinct: How the mind creates language*. Harper Perennial.
- Pinker, S. (1997). *How the mind works*. Norton.
- 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, Article 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, Article 267. <https://doi.org/10.3389/fnins.2021.607887>
- Podlipniak, P. (2022a). Pitch syntax as part of an ancient protolanguage. *Lingua*, 271, Article 103238a. <https://doi.org/10.1016/J.LINGUA.2021.103238>
- Podlipniak, P. (2022b). Free rider recognition—A missing link in the Baldwinian model of music evolution. *Psychology of Music*, Article 03057356221129319. <https://doi.org/10.1177/03057356221129319>
- Podlipniak, P. (2023). Free rider recognition—A missing link in the Baldwinian model of music evolution. *Psychology of Music*, 51(4), 1397–1413. <https://doi.org/10.1177/03057356221129319>
- Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, 115(15), 3972–3977. <https://doi.org/10.1073/pnas.1716090115>
- Preziosi, M. A., & Coane, J. H. (2017). Remembering that big things sound big: Sound symbolism and associative memory. *Cognitive Research: Principles and Implications*, 2(1), Article 10. <https://doi.org/10.1186/s41235-016-0047-y>
- 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>
- Richerson, P. J., Boyd, R., & Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), 8985–8992. <https://doi.org/10.1073/pnas.0914631107>
- Roederer, J. G. (1984). The search for a survival value of music. *Music Perception: An Interdisciplinary Journal*, 1(3), 350–356. <https://doi.org/10.2307/40285265>
- Rousseau, J.-J. (1998). *Essay on the origin of languages and writings related to music (Collected writings of Rousseau)* (J. T. Scott Ed. & trans.). University Press of New England.
- Saussure, F. de. (1959). *Course in general linguistics* (B. Wade, P. Meisel, & H. Saussy Eds.). Columbia University Press. <https://doi.org/LK> - <https://worldcat.org/title/553329>
- Savage, P. E. (2019). Cultural evolution of music. *Palgrave Communications*, 5(1), 1–12. <https://doi.org/10.1057/s41599-019-0221-1>

- 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*, Article e59. <https://doi.org/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*, Article e121. <https://doi.org/10.1017/S0140525X21000042>
- Schilhab, T. S. S., & Gerlach, C. (2008). Connections in action – bridging implicit and explicit domains. In J. Hoffmeyer (Ed.), *A legacy for living systems: Gregory Bateson as precursor to biosemiotics* (pp. 135–144). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6706-8_9
- Schlaudt, O. (2022). Exaptation in the co-evolution of technology and mind: New perspectives from some old literature. *Philosophy & Technology*, *35*(2), Article 48. <https://doi.org/10.1007/s13347-022-00545-z>
- Seifart, F., Meyer, J., Grawunder, S., & Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science*, *5*(4), Article 170354. <https://doi.org/10.1098/rsos.170354>
- Shilton, D. (2022). Sweet participation: The evolution of music as an interactive technology. *Music & Science*, *5*, Article 20592043221084710. <https://doi.org/10.1177/20592043221084710>
- Sober, E. (1988). *Reconstructing the past: Parsimony, evolution, and inference*. MIT Press.
- Spencer, H. (1890). The origin of music. *Mind*, *15*, 449–468. <https://doi.org/10.2307/2247370>
- Steinbeis, N., & Koelsch, S. (2008a). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169–1178. <https://doi.org/10.1093/cercor/bhm149>
- Steinbeis, N., & Koelsch, S. (2008b). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE*, *3*(5), Article e2226. <https://doi.org/10.1371/journal.pone.0002226>
- Stewart, C.-B. (1993). The powers and pitfalls of parsimony. *Nature*, *361*(6413), 603–607. <https://doi.org/10.1038/361603a0>
- Storr, A. (1992). *Music and the mind*. Ballantine Books.
- Tomlinson, G. (2015). *A million years of music: the emergence of human modernity*. The MIT Press.
- Tomlinson, G. (2023). Musical meaning in transspecies perspective: A semiotic model. In E. H. Margulis, P. Loui, & D. Loughridge (Eds.), *The science-music borderlands: Reckoning with the past and imagining the future* (pp. 39–56). The MIT Press. <https://doi.org/10.7551/mitpress/14186.003.0007>
- Wallmark, Z. (2019). Semantic crosstalk in timbre perception. *Music & Science*, *2*, Article 2059204319846617. <https://doi.org/10.1177/2059204319846617>
- Wallmark, Z., & Allen, S. E. (2020). Preschoolers' crossmodal mappings of timbre. *Attention, Perception, & Psychophysics*, *82*(5), 2230–2236. <https://doi.org/10.3758/s13414-020-02015-0>
- Wallmark, Z., Iacoboni, M., Deblieck, C., & Kendall, R. A. (2018). Embodied listening and timbre: Perceptual, acoustical, and neural correlates. *Music Perception*, *35*(3), 332–363.

<https://doi.org/10.1525/mp.2018.35.3.332>

- Wan, C. Y., Rüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception, 27*(4), 287–295.
<https://doi.org/10.1525/mp.2010.27.4.287>
- Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics, 25*(3), 449–466.
<https://doi.org/10.1017/S0142716404001213>
- Warren, P., & Calhoun, S. (2021). Intonation. In J. Setter & R.-A. Knight (Eds.), *The Cambridge handbook of phonetics* (pp. 209–236). Cambridge University Press.
<https://doi.org/10.1017/9781108644198.009>
- 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>
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
<https://doi.org/10.1093/oso/9780195122343.001.0001>
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences, 102*, 6543 LP – 6549.
<https://doi.org/10.1073/pnas.0501844102>
- Whitehead, H., Laland, K. N., Rendell, L., Thorogood, R., & Whiten, A. (2019). The reach of gene-culture coevolution in animals. *Nature Communications, 10*(1), Article 2405.
<https://doi.org/10.1038/s41467-019-10293-y>
- Wilson, E. O. (2013). *The social conquest of Earth*. Liveright Publishing Corp.
- Wong, P. C. M., Chandrasekaran, B., & Zheng, J. (2012). The derived allele of ASPM is associated with lexical tone perception. *PLOS ONE, 7*(4), Article e34243.
<https://doi.org/10.1371/journal.pone.0034243>
- Wong, P. C. M., Kang, X., Wong, K. H. Y., So, H.-C., Choy, K. W., & Geng, X. (2020). ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances, 6*(22), Article eaba5090.
<https://doi.org/10.1126/sciadv.aba5090>
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422.
<https://doi.org/10.1038/nn1872>
- Wurz, S. (2009). Interpreting the fossil evidence for the evolutionary origins of music. *Southern African Humanities, 21*, 395–417.
- Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10*(7), Article e1001372. <https://doi.org/10.1371/journal.pbio.1001372>
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547–558.
<https://doi.org/10.1038/nrn2152>
- Zbikowski, L. M. (2002). *Conceptualizing music: cognitive structure, theory, and analysis*. Oxford University Press.
- 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.

Zlatev, J., Wacewicz, S., Żywiczyński, P., & van de Weijer, J. (2017). Multimodal-first or pantomime-first? *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems.*, 18(3), 465–488. <https://doi.org/10.1075/is.18.3.08zla>

Zlatev, J., Żywiczyński, P., & Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, 5(2), 156–174. <https://doi.org/10.1093/jole/lzaa006>