

Sweet Participation: The Evolution of Music as an Interactive Technology

Music & Science
Volume 5: 1–15
© The Author(s) 2022
DOI: 10.1177/20592043221084710
journals.sagepub.com/home/mns

Dor Shilton

Abstract

Theories of music evolution rely on our understanding of what music is. Here, I argue that music is best conceptualized as an interactive technology, and propose a coevolutionary framework for its emergence. I present two basic models of attachment formation through behavioral alignment applicable to all forms of affiliative interaction and argue that the most critical distinguishing feature of music is entrained temporal coordination. Music's unique interactive strategy invites active participation and allows interactions to last longer, include more participants, and unify emotional states more effectively. Regarding its evolution, I propose that music, like language, evolved in a process of collective invention followed by genetic accommodation. I provide an outline of the initial evolutionary process which led to the emergence of music, centered on four key features: technology, shared intentionality, extended kinship, and multilevel society. Implications of this framework on music evolution, psychology, cross-species and cross-cultural research are discussed.

Keywords

human evolution, interaction, entrainment, attachment, communication, mimesis, social evolution, ritual

Submission date: 22 November 2021; Acceptance date: 13 February 2022

Introduction

One of the most significant biases of Western music scholarship has been the treatment of music as a reified auditory stimulus, something we listen to rather than interact with (Cross, 2012; Turino, 2008). Research has therefore tended to separate sound from movement, listening from active participation, and the aesthetic experience from a social one. It remained conceptually grounded in the concert hall, where audiences listen to compositions written and performed by professionals—a state of affairs largely preserved in today's aural consumption of recorded music. This bias has significantly affected music evolution research. Perhaps its biggest influence is the enduring emphasis on ultimate function, which stems, I argue, from the perceived plausibility of the null hypothesis—that music is, evolutionarily speaking, useless. Having been confined to the status of a mere aesthetic object, music was suggested to be a “byproduct,” and the ability to perceive, produce and enjoy it the result of auditory sensitivities that evolved for other purposes such as auditory scene analysis and language (Pinker, 1997).¹ This argument, and the neo-Darwinian logic in which it was couched, dominated the field for the last two decades. Research

consequently focused on the various ancestral contexts in which music might be functional: male or male-coalition displays (e.g., Merker, 1999; Miller, 2000), coalition displays more generally (e.g., Hagen and Bryant, 2003), infant care (e.g., Dissanayake, 1999) and social bonding (e.g., Dunbar, 2012). More recently, two contrasting articles both rejected the byproduct hypothesis, but still dealt primarily with the question of function: Mehr et al. (2021) argued that music served as a credible signal in the context of both coalition displays and infant care, while Savage et al. (2021) argued for social bonding as an overarching function, operating in the context of infant care, mating and group cohesion.

There is, however, a growing awareness that this approach to music evolution is overly reductive (Killin, 2016; Savage et al., 2021; Tomlinson, 2015). Four related problems can be identified. First, music cannot be clearly isolated from

Cohn Institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Israel

Corresponding author:

Dor Shilton, Cohn Institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Tel-Aviv 69978, Israel.
Email: dorshilt@mail.tau.ac.il



other forms of communication: it is part of a greater communicative toolkit and is interwoven into different communicative registers and rituals (Cross, 2014). Second, it overlooks the relationship of musical behavior to the unique social niche created and inhabited by humans for the past 2 million years (Shilton et al., 2020). Third, it often implies a strictly neo-Darwinian framework, which does not consider the complexities of developmental plasticity, niche construction and genetic accommodation (Killin, 2016; Tomlinson, 2015). Fourth, it is becoming clear that even in ancestral conditions, music or music-like behaviors would have had multiple and diverse purposes, making the search for a single function unfeasible. This problem was noted by Savage et al. (2021), who consequently proposed social bonding as an umbrella explanation, akin to the notion that "vision is for seeing." However, since the function of social bonding is not unique to music, but shared with many other forms of interaction (e.g., conversation, play), the unique properties of music remain to be clarified. A different approach to the question of function would be to ask what is at the core of music as an interactive strategy, distinguishing it—conceptually if not practically—from other forms of communication. This approach can also potentially clarify why music is preferred over alternative interactive strategies in some contexts, and why it is used differently in different societies.

Here, I expand on Savage et al. (2021), and propose a complementary framework which hinges on the treatment of music as an interactive technology. I present two general theories of affiliative interaction, and argue that music elaborates on a basic process of intentional, gestural and emotional alignment through which attachment to individuals and groups is formed, and which is common to all human communication. I argue that music is distinguished primarily by its focus on accurate temporal alignment, which is achieved through the co-construction of a stable periodic framework. I enumerate six distinctive outcomes of this interactive strategy, which enable musical interactions to include more participants, be more compelling and engaging, and last longer. Next, I argue that music, like tool making and language, evolved first as a social technology and was only later genetically accommodated. Finally, I outline the process by which humans became more dependent on one another and on socially created technologies (Shilton et al., 2020). I argue that loose temporal alignment was critical for consolidating attachment between infants and their multiple caregivers, and that it was later extended towards other types of relationships. Accurate temporal alignment, which underlies musical interaction, was borne out of the need to extend attachment to a larger network of kin and non-kin, and is suggested to have been particularly important during regular group gatherings, and for the formation of multilevel societies.

Music as an Interactive Technology

The conceptualization of music primarily as a form of interaction is informed by evidence from four domains of

research. First, the anthropological and fieldwork-based study of music across cultures, which has firmly established that music is deeply embedded in social life and is often a participatory activity (Blacking, 1973; Feld, 1984; Lewis, 2013; Merriam, 1964; Nettle, 2015; Savage et al., 2015; Turino, 2008). Second, experiments in social psychology demonstrating how musical interactions support empathy, bonding and prosociality (Mogan et al., 2017; Pearce et al., 2015; Rabinowitch et al., 2013; Weinstein et al., 2016). Third, the psychological study of music perception, which points to the critical importance of embodied anticipation, and consequently suggests that music listening is essentially active, and can in fact be construed as "covert performance" (Cannon & Patel, 2021; Cross, 2010; Huron, 2006; Koelsch et al., 2019; Patel & Iversen, 2014; Vuust & Frith, 2008). And finally, the study of musical behavior in animals, which finds no shortage of species with complex individual vocalizations, but very few with tightly coordinated or entrained group vocalizations, suggesting flexible coordination as the most important bottleneck for musical behavior (Ravignani et al., 2014; Schachner et al., 2009). The word "music" will therefore denote an interactive process throughout the paper, and could be replaced with "music making," "musical interactions" or simply "musicking."²

Affiliative Interactions

To understand music as a type of interaction, we first need to ground it in a more general theory of affiliative interactions and their effects. On a psychological level, affiliative interactions involve bibehavioral synchrony, a phylogenetically ancient process that underlies the selective formation of attachment bonds in mammals (Feldman, 2017). Bibehavioral synchrony refers to the coordination of behavioral and biological processes between attachment partners, comprising of behavioral synchrony (coordination and alignment of gaze, touch, movement and vocalizations), heart rate coupling, endocrine fit and brain-to-brain synchrony. The behavioral component can be described as a bi-directional signal, in which the correspondence between attachment partners reliably indicates attention and affiliation. Behavioral synchrony activates the mammalian attachment system, which evolved originally in the context of consolidating mother-infant bonds and is underpinned by dopamine and oxytocin crosstalk in the striatum (Feldman, 2017). It is essential in alloparental species to flexibly create emotional bonds with multiple caregivers. In humans, it is observed across all types of social bonds: parental, romantic, peer and conspecific (Feldman, 2017).

On a microsociological level, interaction rituals have been suggested as the most basic model of copresent human interactions (Collins, 2004; Goffman, 1959). Interaction rituals are defined as situations when two or more persons share an attentional focus and an emotional state, and align their communicative expressions in both form and periodicity. When successful, they have three main outcomes: group solidarity, positive

emotional energy (as well as emotional contagion), and shared symbols and values. Group solidarity is a form of attachment that extends towards a group rather than a specific individual. Emotional energy ranges from enthusiasm and high motivation to depression and apathy. Successful interaction rituals elevate emotional energy, making participants more driven and exuberant. The shared symbols are those objects of shared attention that connect the group, and can consequently transform into sacred objects. Shared values are often embedded in the ritual itself and are derived from the laws of behavior guiding the ritual (e.g., Lewis, 2013).

Participation in rituals is often costly, reliably displaying the commitment of participants to the sacred symbols and social rules which govern the ritual (Henrich, 2009; Wen et al., 2020), while the emotional energy generated during the ritual intensifies it (Collins, 2004). Whereas biobehavioral synchrony explains the neuroendocrine basis of attachment formation between dyads (originally, mother and infant), interaction ritual theory explores how the basic rules of dyadic interaction can be expanded to a group, and how these rituals produce shared symbols and values.

Interaction rituals do not necessarily succeed. They can also fail, and in their failure, point to disparities and dissatisfactions. As such, they also reliably test the strength of personal relationships and group identity. We can clearly recognize when our interaction partners are inattentive or unresponsive, when there is no shared “rhythm” or mutual understanding. The same is true of group rituals, where failure to participate can bring to the fore hidden disputes (e.g., Oloa-Bilola, 2017, pp. 198–199). Interaction rituals are therefore both producers and reliable tests of social solidarity.

Examples of interaction rituals abound—indeed, Collins (2004) considers them to be quite ubiquitous in social life. To take one very well-known example, let us consider the presidential inauguration ceremony in the United States. After a period of heated disagreement, political parties and their supporters need to align behind the elected candidate. Crowds gather in the Capitol and watch together the new president become symbolically one with the state by reciting the oath of office under the nation’s flag and in front of the iconic Capitol dome. The ritual enacts the power relations and organizational principles it is meant to uphold: the platform of leaders stands above the citizens, and both watch the newly empowered head of state proclaim his allegiance to the state’s basic laws. The audience co-constructs and invigorates the ceremony by its sheer size and uniformity, whether expressed as silent mutual focus or through enthusiastic applause. The emotional energy generated during the ceremony revitalizes the commitment of participants to the symbols and values embedded in it. Just as personal attachments need to be affectively affirmed routinely, so does a person’s sense of belonging to a larger community. It cannot be merely stated—it must be *felt*.

Musical Interactions

Both biobehavioral synchrony and interaction rituals are widely applicable across various types of human

interactions. The alignment of attention, emotion, and the shape and rhythm of gestures and vocalizations constitutes the interaction engine that underlies all human communication (Levinson, 2006). While it involves alignment on several levels, some can be prioritized over others, in service of different mutual goals. Conversations, for example, prioritize the alignment of imagined referential meaning (Dor, 2015), and mutual focus is set primarily on encoding and decoding instructive messages, though other types of alignment are occurring simultaneously.

Because all communication relies on a basic interaction engine, music can sometimes be difficult to clearly demarcate as a pattern of interaction. Speech can often sound “musical,” with distinctive pitch movement and rhythmic organization. Conversations can demonstrate a relatively high degree of periodic and tonal alignment between participants (Hawkins, 2014; Robledo et al., 2016), and the back-channel communication that scaffolds them can be more simultaneous and rhythmic, include more participants and result in greater involvement from listeners (Bavelas et al., 2002; Wiessner, 2014). Clear distinctions, therefore, appear to be more a matter of cultural constructs than of any hard cognitive boundaries. Different cultures slice up the communicative spectrum in different ways, utilizing pragmatically and normatively the vocal and gestural channels to serve different purposes (Cross, 2015; Everett, 2012; Lewis, 2014; Seeger, 1987; Senft, 2018). Lewis (2009), for example, describes several interactive categories defined by the Mbendjele, which incorporate musical elements to different degrees: from the hushed, secretive and monotonic “*ya miso minai*” (speech of four eyes), through the louder, more song-like women’s talk (*besime ya baito*), to *gano*, a form of storytelling which involves spoken narrative, song and rhythmic entrainment, and *massana*, a full-blown communal song and dance. It is reasonable to expect similar or even greater entanglements of speech, song and movement in our evolutionary past, as hominins were venturing into the more complex use of multimodal communication and honing their skills in each modality—the vocal modality, in particular (Levinson & Holler, 2014; Mithen, 2005).

That said, the property which seems to distinguish music most consistently is the degree of temporal and tonal alignment between participants (Robledo et al., 2021). The focus of attention is not on some external object or some displaced event participants are trying to imagine together—it is on the participants’ rhythmic, gestural and vocal coordination. If language can be thought of as an extension of the extrinsic component of interactions—the mutual reference to external objects—music is an extension of the intrinsic component of interactions—the embodied alignment which affirms the affiliative relationship between participants and allows for cooperative interactions (Whiteman, 2020).

Musical interactions are therefore a special case of biobehavioral synchrony or interaction ritual, in which action is entrained within a stable periodic framework, instead of

being more flexibly coordinated. The shared periodic framework is constituted through the repeated articulation of pitched or unpitched rhythmic patterns (regular sequences of inter-event durations, often experienced as related to one another by simple ratios; Polak et al., 2018). From these, regular pulses are abstracted, according to which vocal utterances, percussion and movement can be coordinated (Clayton et al., 2020; Jones, 2016). The use of the vocal modality prioritizes rhythmic structure and salient pitch movement, mostly through discrete pitch changes. The latter also enables frequency and spectral alignment between multiple voices (Savage et al., 2021). Timelines, metric structures and pitch classes are all elaborations of these basic features, and extend the ability of humans to coordinate and diversify their contributions.

The periodicity of music is prioritized because it is the most foundational resource for creating vocal and gestural alignment between multiple participants. If we are to consider the meaning of a signal as the desired response of its receiver, then the cyclicity of music carries one of its most basic meanings: it invites participation. This is supported by studies showing that repetition of sound stimuli makes them feel more musical (Simchy-Gross & Margulis, 2018), and perhaps best illustrated by the speech-to-song illusion, in which the repetition of a speech phrase re-orientates listeners away from decoding its linguistic meaning, to a focus on reproducing its rhythmic and tonal structure (Deutsch et al., 2011). Repetitive rhythmic patterns are, in that sense, like gaze following and pointing for referential communication. Both are invitations to attend and respond jointly to a shared experience. A pointing finger means: “look at this, with me”; a repetitive rhythmic sequence means: “*embody* this, with me.”

Of course, periodicity is not always as foundational. Musical interactions can have different levels of stratification (the extent to which some participants control and lead the ritual), from the mostly leveled polyphony of BaYaka spirit plays to the ornate recitation of scripture by a single Hazzan in synagogues (Lewis, 2013; Slobin, 2002; Turino, 2008). The more stratified a ritual is, the less emphasis is expected on the structuring of a shared periodic framework enabling wide participation. Instead, the focus will be on elaborate performance techniques, capturing the attention of participants while preventing more active involvement. From the perspective offered here, the focus on a single performer is more similar to oratory, while the shared immersion in non-periodic sound—be it vocal, orchestral or ambient—is a categorically different experience.

These distinguishing features of music raise a question: Why did entrained temporal coordination evolve when less accurate forms of alignment are apparently sufficient to create attachment bonds? Several distinctive outcomes of music’s unique interactive strategy can be identified:

1. Musical interactions increase substantially the potential number of participants, and were

demonstrated to be effective for social bonding in groups of over 200 people (Launay et al., 2016; Weinstein et al., 2016). Speech usually involves one speaker at a time, which limits the ability of others to fully participate, and may be abused by dominant individuals. This problem exacerbates as group size increases, as does the problem of assessing temporal alignment between participants, upon which attachment formation depends. A stable periodic framework allows for a clear convergence between many participants, and bypasses the limitations of turn-taking by allowing everyone to participate at the same time.

2. Musical interactions strengthen the effects of emotional contagion, a process in which automatic bodily mimicry results in emotional convergence (Hatfield et al., 1994). Music, as Langer (1957) wrote, reflects the “morphology of feeling,” a proposition affirmed by the consistent association between musical form (e.g., tempo, pitch, timbre) and certain emotional states (Juslin, 2019; Juslin & Laukka, 2003). Enacting that form (and, to a lesser extent, listening to it) changes the participants’ emotional state in a bottom-up process, and can result in such an overwhelming state of convergence that participants feel as though they merged into a larger, collective body (Lewis, 2014).
3. Musical interactions can engage participants for a very long time. Among Mbendjele and Suya, for example, ritual singing can last for several hours and even days (Lewis, 2013; Seeger, 1987). The Natural History of Song ethnographic corpus, which contains coded ethnographic texts mentioning singing from 60 traditional societies, lists 168 episodes of singing which lasted for 1–10 hours, and 31 which lasted for more than 10 hours (Mehr et al., 2019). Long durations may be partially due to how music alters time perception: listening to music can give the impression that time moves faster or slower or even disappears completely (Schäfer et al., 2013).
4. Musical interactions can accommodate trancing, defined by Becker (2004, p. 43) as “a bodily event characterized by strong emotion, intense focus, the loss of the strong sense of self, usually enveloped by amnesia and a cessation of inner language”. By focusing primarily on bodily action in the present moment, performers feel more intensely their most basic sense of self as it exists in the here and now, and less their autobiographical self, which connects that present feeling with a personal past and future (Damasio, 1999). This profound change in conscious experience is often associated with religious practices, as it enables performers to embody or commune with spirits, partially explaining the cross-cultural use of music for communication with the supernatural (Nettl, 2015).

5. Musical interactions allow participants to unite on a more basic, physical level, establishing a floating intentionality that can be vital for temporarily curbing disputes and handling precarious social situations (Cross, 2009).
6. Musical interactions often demand higher levels of exertion, resulting in greater opioid release which enhances social bonding as well as inducing a general feeling of euphoria (Tarr et al., 2015).

All of the above make music a highly potent technology of engagement (Shilton et al., 2020), capable of producing higher levels of emotional contagion and emotional energy, and creating a stronger sense of group solidarity and a firmer devotion to shared symbols and values. Its ability to extend a shared copresent and embodied experience to tens and even hundreds of participants suggests that the invention of musical interactions was intimately connected to changes in human social organization, in particular, the creation of larger and more stable parties, and the accommodation of larger temporary aggregations.

To summarize, music can create a narrowly focused, present-oriented, shared intentional space that diffuses and defuses existing tensions (if only temporarily), removes individual boundaries and creates a larger, unified self. It bypasses the partial limitation of turn-taking in conversations by foregrounding the backchannel, allowing multiple individuals to contribute more evenly at the same time, and producing an important leveling effect. It also increases substantially the number of people who can participate in a shared intentional space. Musical interactions are inherently affiliative, emphasizing the relational dimension of communication, and focusing more on the interaction itself (an intrinsic rather than extrinsic purpose), and more on the body and the present moment. Group musical interactions are consequently an important feature of social gatherings across cultures, particularly those involving the supernatural—propitiation of spirits, initiations, healing and mourning—but also in other contexts related to group coordination, such as work, recreation, and games (Feld, 1984; Lewis, 2013; Mehr et al., 2019; Merriam, 1964; Nettl, 2015).

If we attempt to provide a general explanation of music parallel to “vision is for seeing,” but in the sense explored here—focusing on process rather than purpose—we arrive at an apparent tautology: music is for musicking. While used in a variety of contexts, music has a single interactive strategy: the tight coordination of vocal and corporal gestures in time. It shares with other forms of interaction the utility of creating social bonds and shared symbols but can be far more powerful than others in generating group solidarity and a shared emotional state because it unites the basic embodied experience of multiple participants. This bridge that music creates between embodiment and social co-production was aptly captured by Langer (1957, p. 199): “Music,” she wrote, “is our myth of the inner life.”

Culturally Driven Evolution

Music was initially conceptualized as a technology to differ it from a biological adaptation (Patel, 2008; Pinker, 1997). This is a false dichotomy. Like tool-making technologies, music evolved through cumulative innovation sustained by social learning, and—along with other communication technologies—influenced biological evolution (Dor, 2015; Killin, 2016; Patel, 2018; Tomlinson, 2015). This section aims to explain how evolutionary adaptations may stem from behavioral, developmental, and cultural changes, in a phenotype-first mode of evolution in which “genes are followers, not leaders” (West-Eberhard, 2003, p. 20). More specifically, it adds to the existing literature on music and niche construction (Killin, 2016; Tomlinson, 2015) the concepts of plasticity and genetic accommodation, which have yet to be integrated more fully into the emerging coevolutionary framework (though see Podlipniak, 2017).

The basic process of phenotype-first evolution has been summarized by West-Eberhard (2005). We start with a varied population of developmentally plastic organisms. Environmental changes are then met with variable developmental responses, which consist of new combinations of phenotypic traits. Given the persistence of these environmental changes, there is a consistent selection of the most adaptive responsive phenotypes. This may result in genetic accommodation, in which genetic variations that support the phenotypic adaptations are selected. Given the genetic complexity of natural populations, genetic accommodation does not necessarily require new mutations—it is often likely that standing variation will be sufficient to accommodate new phenotypic responses.

Genetic accommodation may increase or decrease the plasticity of a given trait (Dor & Jablonka, 2010; Schlichting & Wund, 2014). Genetic assimilation, which decreases the plasticity of a trait by making it less dependent on environmental inputs (a process also known as canalization), is one type of genetic accommodation and was famously demonstrated in *Drosophila* by Waddington (1953). Alternatively, fluctuating environments may result in selection for increased plasticity, where context-dependent responses are more suitable (Jablonka, 2017). Genetic accommodation is especially relevant when considering behavioral adaptations that are based on novel neural associations. Avital and Jablonka (2000, pp. 330–333) have suggested the “assimilate-stretch” principle, in which a behavioral sequence is canalized (becoming less dependent on learning and environmental inputs), simplifying the learning process and freeing up cognitive resources that can later be used for the sophistication of that behavior. Learning is thus guided by predispositions, while remaining open-ended.

In humans, cumulative culture adds another dimension. Accurate social learning—to which humans are predisposed—provides another channel of inheritance, which can result in environmental cues being sustained for longer periods (Jablonka & Lamb, 2005; Laland et al.,

2000; Tennie et al., 2009). Selection is then guided by a culturally constructed niche, in a process that essentially turns the gene-first neo-Darwinian view on its head: culture is the fountainhead from which changes in cognition and physiology arise, changes that may eventually become genetically accommodated.

The earliest example of technology transforming human evolution is probably that of tool making and its influence on the human hand. The production of sharpened stone tools originated over 3 million years ago (Harmand et al., 2015), and became more systematic over time. As early as 2 million years ago, the human hand acquired several traits which improved the precision grip abilities underlying stone tool production and use: a bigger thumb-to-fingers ratio, increased thumb robusticity, musculature and opposition efficiency, and broad fingertips (Karakostis et al., 2021; Key et al., 2018; Richmond et al., 2016). A much more recent example of gene-culture coevolution is the relationship between dairying and lactase persistence. There is clear evidence that the cultural and behavioral adaptations related to dairying preceded the correspondent genetic changes by several thousand years (Burger et al., 2020; Gerbault et al., 2011). Studies of lactase persistence also demonstrate how a variety of changes to regulatory pathways can lead to a single phenotypic trait—with different single nucleotide polymorphisms implicated for different populations (Ingram et al., 2009; Ségurel & Bon, 2017).

Dor and Jablonka (2000, 2010, 2014) have written extensively about the relevance of this process for the evolution of language. They argue for a culturally-driven coevolutionary process, in which interactive exploration resulted in communicative innovations—themselves reliant upon individual plasticity—which was later genetically accommodated. Their crucial point is that language was borne out of social processes rather than starting with changes in individual cognition: cultural invention preceded and guided biological, neurophysiological, and genetic adaptation. “First we invented language,” they write, “then language changed us.” (Dor & Jablonka, 2014, p. 16)

Languages are often shaped by the communicative demands of different social environments. For example, languages that are more exoteric—with larger speaker populations, greater geographical spread, and more contact with other languages—have simpler morphologies and larger phonological inventories (Lupyan & Dale, 2010; Nettle, 2012). These characteristics appear to be shaped by the communicative pressures of exoteric groups, in which more frequent interactions between strangers and a greater proportion of second language adult learners require simpler and more systematic language structures. Following the same logic, Dor and Jablonka suggest core properties of language have emerged to meet the demands of a changing social environment of increasing cooperativity and codependence, and were constructed through the use of a more limited mimetic communication system (Donald, 1991; Dor, 2015). Music, I argue, was likewise constructed, though it is still unclear what social demands—beyond

codependence—stimulated the development of its unique interactive strategy.

Behavioral innovations are reliant upon individual plasticity, which is amply provided by the exceptionally large human brain. Dor and Jablonka (2014) illustrate this with the fascinating example of human echolocators: blind people who have learned to perceive their physical environment by making clicking sounds and listening to their echoes (Thaler & Goodale, 2016). Functional magnetic resonance imaging (fMRI) studies show that in these individuals, areas of the brain normally related to visual processing are recruited for the processing of sounds, even producing retinotopic-like maps in the primary visual cortex (Norman & Thaler, 2019; Thaler et al., 2011). Plastic reorganization was also demonstrated in deaf people, with fMRI studies showing auditory cortex activation during the discrimination of temporally complex visual stimuli (Bola et al., 2017). Cross-species studies of the capacity to interact with language and music also demonstrate the importance of plasticity. Several great apes and a single exceptional parrot have learned language-like systems of communication (Patterson & Cohn, 1990; Pepperberg, 1999; Savage-Rumbaugh & Lewin, 1994). A sulphur-crested cockatoo has learned to dance (Patel et al., 2009), and a California sea lion was trained to accurately bob her head to an isochronous beat (Cook et al., 2013). The extent of neural plasticity in these and many other cases makes it all the more likely that individual adaptations to changing social environments need not rely on novel mutations.

Genetic accommodation seems to fit the evolution of musicality (the biological capacity to engage in music) because of the latter’s partial modularity, early ontogeny and long evolutionary history. Beat induction appears to play no role in linguistic communication, while fine and relative pitch processing play much more subtle ones. Congenital amusia, a condition impairing fine pitch discrimination, appears in relative isolation from speech disorders, and has a much smaller effect on language processing (though more pronounced in the case of tonal languages) than on music processing (Liu et al., 2012; Peretz, 2016). Studies in newborn infants reveal the very early ontogeny of beat perception (Winkler et al., 2009) and of right hemisphere dominance (Perani et al., 2010). Moreover, music is at least as old as the earliest musical instruments (approximately 40,000 years, Buisson, 1990; Conard et al., 2009), and almost certainly much older (d’Errico et al., 2003). The longer a selective environment persists, the more likely it is that the plastic responses to it will be genetically accommodated. If 8–11,000 years of dairy farming resulted in genetic accommodation, it is highly reasonable that tens or even hundreds of thousands of years of musical interactions would result in a similar process.

The Evolution of Musical Interactions

In the following, I introduce chronologically four key features of human evolution, that together laid the foundations

for the social construction of music as an interactive technology. Each is based on well-established paleoanthropological evidence, though arguments about their influence on more nuanced interactive dynamics are necessarily conjectural. First, the technological niche, improving motor and emotional control and establishing human reliance on tools and on the social creation of technologies; second, shared intentionality and social learning, which resulted in humans experiencing their world as a shared intentional space; third, extended kinship, which expanded the creation of attachment bonds through temporal alignment into a wide array of human relationships; and fourth, multilevel society and the creation of larger aggregations, in which entrainment could have played a significant role. Taken together, these evolutionary processes created the prerequisites of musical behavior—fine motor control, multimodal communication, shared intentionality, and bonding behavior extending toward non-kin and groups (see also Killin, 2017, 2018; Tomlinson, 2015). They also profoundly changed human social organization, extending kinship networks and creating the need for communication technologies capable not only of sharing information in a bigger network but of establishing attachment and trust within it.

Technology

The first sharpened stone tools used by hominins predate the emergence of *Homo erectus* by more than a million years (Harmand et al., 2015; McPherron et al., 2010). During that time, hand morphology and musculature had changed substantially, with the most prominent divergent traits suggesting it coevolved with toolmaking. A growing reliance on tools also coevolved with domain-general and domain-specific changes in brain structure. Overlap in brain areas involved in both tool-use and speech suggest improvements in manual dexterity could also involve finer motor control in the facial and vocal modalities (Stout & Chaminade, 2012). Furthermore, even the earliest Oldowan tools (2.6 mya) were suggested to have been reproduced through behavior copying rather than individual reinvention (cf. Donald, 1991; Stout et al., 2019). All of the above point to the early start and lengthy development of the human technological niche, marked by an increasing reliance on tools for foraging, which drove the complexification of those tools by means of social learning, which then selected for improved social learning and motor and emotional control (Shilton et al., 2020; Stout & Khreisheh, 2015). It is the first instance of an important pattern in human evolutionary history: a growing reliance on a certain technology drives the improvement of that technology, which then further increases the reliance on it, triggering more improvements, and so on. This evolutionary spiral is relevant not just to stone tool technologies, but to communication and interactive technologies as well (Dor, 2015). The technological niche thus set the stage for the evolution of music by improving fine motor control and social learning skills and engendering the

evolutionary spiral dynamic that would underpin the evolution of complex communication technologies.

The technological niche also means that humans have been deft percussionists and have inhabited a distinctive soundscape for millions of years. Both stone tool production and the extraction of marrow involve precise striking and are staggeringly ancient. Rhythmicity aside, it can be safely assumed that percussive activities were accurately executed and that the accompanying sounds were intimately familiar to ancient hominins. It was experimentally demonstrated that Aurignacian-type flint blades can produce distinct pitches, resulting in consistent use-wear patterns (Cross & Blake, 2008). Pitched and unpitched percussion in echo-rich environments can also produce quite impressive sound effects, with several examples of such use in the Paleolithic (Dams, 1985). While percussion is unlikely to be entrained in the context of tool making and marrow extraction, the endurance of this activity for millions of years means it was a proximate target for later explorations of entrained joint action.

Shared Intentionality

The emergence of *Homo erectus* marked a major advance in the social learning of skills, as indicated by the more complex Acheulian technology, the hunting of large mammals, and the rapid migration out of Africa and across Eurasia. What form did these new modes of social transmission take, and what were their affective, cognitive and communicative requirements? Explicit teaching is unlikely to have been an important factor at any time period, as it is rare even among modern foragers (Shilton, 2019). Instead, social learning seems to occur through the shared experience of foraging and tool-making activities. Novices spend time with experienced adults, observing them as they hunt, forage, make tools and locate raw materials. Through gaze following and attention guiding gestures, they learn to attend to the same external cues, e.g. the tracks and calls of different animals, the look and texture of suitable lithic and organic materials, or the spot at which it is best to strike a stone core. Both in real-time with adults, and in jest with inexperienced peers, novices practice and gradually attain the necessary perceptual-motor skills.

These activities require, at the very least, habitual group foraging, a good degree of social motivation, a theory of mind, and a communicative repertoire enabling shared intentionality. Mimetic communication, a multimodal and representational toolkit that includes bodily and manual gestures, facial expressions, vocalizations and mimicry, was critical to accommodate these new forms of social learning (Donald, 1991; Shilton, 2019). Functionally limited to the here and now, mimesis allowed the diverse types of cooperation which became essential to human subsistence. Most importantly, the regular practice of social learning steadily transformed the environment into an intersubjective space, in which one is increasingly aware of the

attention and intent of others, and increasingly motivated to seek out this information.

Large mammal hunting, the extraction of lithic raw materials, and the reduction of large cores for the production of bifacial stone tools were all likely practiced predominantly by males, involving substantial risks and often requiring considerable physical strength. Hunting, in particular, seems to require active cooperation, as in the absence of deadly projectile weapons the capturing and killing of large, prime-aged mammals is unlikely to have been achieved alone (Bunn & Gurtoy, 2014). According to Sterelny (2020), these indicate the presence of male coalitions, and make it likely that early *Homo* had *Pan*-like residential patterns, with subadult female dispersal. Sterelny further suggests that it was this residential pattern that explains the apparent stasis in lithic technology observed between approximately 1.7 and 0.9 Ma. As long as males remained in their natal groups, innovations in Acheulian bifacial technology could not disperse beyond it. Whether or not the stasis was related to male vs. female dispersal, it has been shown by several studies that fluid dispersal (both males and females leaving their natal residence) and extensive kin recognition, common among modern foragers, are essential for cumulative culture (Dyble, 2018; Migliano et al., 2017, 2020). The change to the fluid residence is related to another major factor in human evolution: alloparenting, and the dramatic expansion of kinship relationships.

Extended Kinship

Sometime during erectine evolution, infant care became—like foraging—a cooperative effort (Hrdy, 2009). Extended altricial periods, most likely related to the brain size increase, meant that mothers needed plenty of assistance. Matrilocal residence and female coalitions consequently became more common. Stable pair bonds—perhaps a correlate of male coalitions and the leveling effect of deadly weapons—did not only add another provisioning parent but also increased substantially the size of kin networks (Chapais, 2017). Female coalitions would have also resulted in decreased intragroup and intergroup violence—as suggested by the different rates of violence in chimpanzees and bonobos—extending the kin networks even further (Chapais, 2017; Furuichi, 2011; Stanford, 2018). Human sociality went through a profound change, and the added dimension of kinship had its impact on the evolution of interactions.

As previously mentioned, behavioral synchrony—the matching of gaze, movement and vocalization—was essential for the flexible creation of attachment bonds between infants and multiple caregivers. As infant care became a more distributed task, so did the proficiency in this form of interaction, which foregrounds the alignment between attachment partners. Both male and female infants would engage in behavioral synchrony for long developmental periods, making it quite plausible that behavioral synchrony

would radiate from early caregiving interactions to pair and peer interactions in later life. It is also important to remember the fluid nature of copresence in caregiving situations, which were not strictly dyadic. Inexperienced caregivers would be present and watching more experienced ones, and infants would occasionally move from the arms of one to another. This fluidity not only enabled the social learning of caregiving skills but also provided a possible arena for entrainment between multiple participants, thus expanding attachment formation from the dyad to the group.

Behavioral synchrony thus became an important component of the human communicative toolkit. Humans were not just guiding each other's attention but were creating affiliative relationships through attentional, gestural and emotional alignment. Adding an affective component to human relationships increased social motivation and, consequently, the amount of active time spent in social interactions—certainly beyond the 10%–20% of active time characteristic of primates (Fuentes, 2021).

Biobehavioral synchrony allowed for kinship to become more flexible—to be socially constructed rather than biologically mandated. By frequently interacting in this manner, unrelated individuals could feel like kin, and a variety of relationships could tap into the neuroendocrine core of the mother–infant attachment bond. In modern humans kinship is extended to non-biological kin, as well as other-than-human persons like animals, forests and spirits—making it one of the key relational principles through which humans perceive their place in the world (Bird-David, 1999), and enabling a significant increase in social complexity.

Multilevel Society

Extended kinship affected all levels of hominin social organization: mother-infant dyads expanded to include more caregivers temporary small parties grew into larger, more stable bands; and dispersed, fission-fusion communities turned into federated societies (Grueter et al., 2012; Layton et al., 2012). All levels were scaled up, and a new midlevel social entity had emerged: the band. In modern foragers the band normally comprises of about 30 individuals (though numbers can vary substantially) who are organized as families, engage in cooperative foraging and childcare, and regularly assemble each night at a camp site (Layton et al., 2012). Archeological evidence suggests the residential camp had emerged about 400,000 years ago, and perhaps even sooner (Goren-Inbar et al., 2018; Kuhn & Stiner, 2019). This new form of sustained copresence was the primordial soup in which new forms of interaction and joint action were experimented with in multi-participant settings. Various situations could call for coordinated group action, including predator deterrence, coalition displays and play. Predator deterrence through group music making was documented among several hunter-gatherers (Knight & Lewis, 2017). The joint production of loud,

heterogenous sounds seems particularly crucial during dark hours, when humans are most vulnerable to large nocturnal carnivores (and have been for millions of years; Knight and Lewis, 2017; Packer et al., 2011). The compelling nature of music, as well as its influence on participants' sense of time, makes it a highly viable strategy for extending the duration and intensity of group-coordinated predator deterrence. Another platform for joint action was coalition displays, whether by males in territorial boundaries or by females against dominant males (Power, 2014). The regular use of fire extended the possibilities of nighttime interactions, which in modern foragers differ substantially from those engaged during the day (Shimelmitz et al., 2014; Wiessner, 2014). While daytime interactions are more economically focused and involve the separation of the band into several parties, the night unites everyone in a single location, with little productive work to be done, and in a somewhat enchanted atmosphere. The shift from extrinsic to intrinsic purpose is therefore quite natural, as interactions are aimed less at achieving a clear target and more at strengthening the connection between group members.

Regular aggregations are another feature of multilevel societies likely to have been of importance to the emergence of musical interactions. Common to many modern foragers, and often corresponding to seasonal variation in resource availability, regular aggregations can involve hundreds of people and were described as times of intense sociality (Lee, 1972; Mauss & Beuchat, 1979; Shott, 2004). The distinct features of music seem uniquely poised to resolve some of the challenges of aggregated residence. Interpersonal entrainment can join tens and even hundreds of participants in a single copresent interaction, engendering a sense of belonging and general euphoria. These positive emotional effects, along with music's floating intentionality, can help mitigate the greater potential for disputes in large groups (Lee, 1972). Looking at the different levels of social organization—from residential units to temporal, large-scale aggregations—it seems reasonable that the greater the number of participants (and the less familiar participants are with one another), the more formalized group interactions need to be. Thus, band gatherings and seasonal aggregations seem to be the most crucial platforms for the ritualization of entrainment-based interactions.

Conclusion and Future Directions

The framework presented here considers musical interactions to be a type of biobehavioral synchrony on the psychological level, and interaction ritual on the sociological level. Music is considered as part of a larger communicative toolkit and is most consistently characterized by entrained coordination of movement, percussion and vocalization, based on a shared periodic framework. It emerged after the unique features of human sociality had been established, within the context of shared intentionality, extended kinship and an interaction engine based on mimesis.

This framework coheres with some of the fundamental perspectives offered by other theories of music evolution. It is most in line with the social bonding hypothesis (Savage et al., 2021), and helps clarify why social bonding seems to be more relevant to music than language, despite the fact that both serve that purpose. Since music focuses more on the interactive event itself and less on external objects (present or imagined), it more obviously, and often more powerfully, contributes to social bonding. Biobehavioral synchrony clarifies further the strong relationship of music to infant care and mate bonding, while the emphasis on joint action through entrainment fits the focus on coalitions (Hagen & Bryant, 2003). The concept of costly signaling is relevant primarily to the reliable indication of attachment and commitment to a single partner or to a group through loose temporal alignment or entrainment. The complexity and cultural specificity of ritual behaviors—including musical performance—indicate it is primarily an in-group-directed display, as only in-group members can judge whether they are performed correctly (cf. Mehr et al., 2021). Finally, the understanding of music as a more emotionally effective copresent interaction explains why, in societies in which such interactions are increasingly unimportant for the maintenance of alliances and institutions, and in which music is experienced mostly outside the context of social gatherings, the byproduct hypothesis should seem particularly appealing.

The conceptualization of music as a form of biobehavioral synchrony has two important implications. First, it resolves the apparent circularity problem of the social bonding hypothesis raised by Pinker (1997) and rearticulated by Mehr et al. (2021). Musical interactions rely and elaborate on the baseline bonding effects of biobehavioral synchrony. They, therefore, had bonding as well as rewarding effects from the moment humans started exploring them as an interactive possibility.

Second, it potentially offers a different interpretation of studies of reward system activation in response to music listening (Salimpoor et al., 2011; Salimpoor & Zatorre, 2013). Results from such studies are often interpreted within the framework of musical expectation theory (Huron, 2006; Meyer, 1956), which focuses on music's ability to generate salient expectations, combined with the framework of prediction error theory, which predicts positive reward to be generated when outcomes are better than expected, though what that means for music is unclear (Cheung et al., 2019; Salimpoor et al., 2015; Schultz, 2017). However, this interpretation fails to convincingly account for the clear preference of familiar music by most listeners (Madison & Schiöde, 2017; Pereira et al., 2011), since fully predicted outcomes are not meant to trigger a reward. While the expectations generated by music are quite certain to play a critical role, it is possible working within the framework of attachment theory—which involves dopamine, oxytocin, opioid and endocannabinoid activity in the same brain areas—will provide more fruitful interpretations than theories of musical affect relying solely on prediction error.

The emphasis given here to social co-production and plasticity suggests that cross-species experiments should not just test baseline abilities, but also explore how social motivation and long-term learning can influence the extent to which different species can participate in musical interactions. If music was social from its inception, it makes sense that it should be studied as a plastic response to a form of social interaction. This is very likely what happened in the case of “Snowball,” the dancing cockatoo (Patel et al., 2009), and could probably be reproduced with a paradigm similar to that used by Pepperberg (1999). This experimental approach has the potential to produce insights on baseline abilities of different species, developmentally sensitive periods, and correspondent changes in behavior, physiology and neuroanatomy.

Finally, this framework suggests a positive correlation between temporal alignment and affiliative interaction will be observed across various forms of communication and diverse human cultures. Similar to Savage et al. (2021), it suggests that studies of musical practice across cultures will find a greater prevalence overall of group, participatory music making over solo performances. Levels of participation should correlate with emphasis on rhythm, and vary based on ritual stratification, which itself may depend on social organization. The hypothesized relationship between music and midlevel social structures also suggests musical activity should be more common during the night, and strongly associated with regular aggregations. Altogether, these diverse forms of evidence are expected to deepen our understanding of music as an interactive technology.

Acknowledgements

I am grateful to Ian Cross, Eva Jablonka, Daniel Dor, Anton Killin, Aniruddh Patel, David Huron, Nikki Moran and Jin Hyun Kim for their helpful suggestions and comments on previous versions of this manuscript.


Action Editor

Elizabeth Tolbert, John Hopkins University, The Peabody Institute.

Peer Review

Nikki Moran, University of Edinburgh, Reid School of Music. Jin Hyun Kim, Humboldt-Universität zu Berlin, Institute for Musicology and Media Science.

ORCID iD

Dor Shilton  <https://orcid.org/0000-0002-6503-3693>

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

The author received no financial support for the research, authorship, and/or publication of this article.

Notes

1. It is intriguing that Pinker is the most recognized with this position, as it was Pinker himself who argued against the byproduct view of language held by Chomsky and Gould (Pinker & Bloom, 1990). Just as Chomsky tore language from its social context—characterizing it as a tool for thought—so did Pinker with music, characterizing it primarily as a tool for pleasure.
2. The term musicking as used by Small (1998) referred to all activities related to organizing and performing a musical event. Here, I refer just to the actual music making and those engaged in it, mostly as performers but also as a sufficiently engaged audience.

References

- Avital, E., & Jablonka, E. (2000). *Animal traditions: behavioural inheritance in evolution* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511542251>
- Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: the role of gaze. *Journal of Communication*, 52(3), 566–580. <https://doi.org/10.1111/j.1460-2466.2002.tb02562.x>
- Becker, J. (2004). *Deep listeners: music, emotion, and trancing*. Indiana University Press.
- Bird-David, N. (1999). “Animism” revisited: personhood, environment, and relational epistemology. *Current Anthropology*, 40(S1), S67–S91. <https://doi.org/10.1086/200061>
- Blacking, J. (1973). *How musical is man?* University of Washington Press.
- Bola, L., Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., & Szwed, M. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proceedings of the National Academy of Sciences*, 114(4), E600–E609. <https://doi.org/10.1073/pnas.1609000114>
- Buisson, D. (1990). Les flûtes paléolithiques d’Isturitz (Pyénées-atlantiques). *Bulletin de La Société Préhistorique Française*, 87(10/12), 420–433. JSTOR. <https://doi.org/10.3406/bspf.1990.9925>
- Bunn, H. T., & Gurtov, A. N. (2014). Prey mortality profiles indicate that early pleistocene homo at olduvai was an ambush predator. *Quaternary International*, 322–323, 44–53. <https://doi.org/10.1016/j.quaint.2013.11.002>
- Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y., Žegarac, A., Hofmanová, Z., Winkelbach, L., Reyna-Blanco, C. S., Bieker, V., Orschiedt, J., Brinker, U., Scheu, A., Leuenberger, C., Bertino, T. S., Bollongino, R., Lidke, G., ... Wegmann, D. (2020). Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 3,000 years. *Current Biology*, 30(21), 4307–4315.e13. <https://doi.org/10.1016/j.cub.2020.08.033>

- Cannon, J. J., & Patel, A. D. (2021). How beat perception Co-opts motor neurophysiology. *Trends in Cognitive Sciences*, 25(2), 137–150. <https://doi.org/10.1016/j.tics.2020.11.002>
- Chapais, B. (2017). From chimpanzee society to human society: bridging the kinship Gap. In M. N. Muller, R. W. Wrangham, & D. R. Pilbeam (Eds.), *Chimpanzees and human evolution* (pp. 427–463). Harvard University Press. <https://doi.org/10.4159/9780674982642-012>
- Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory Cortex activity. *Current Biology*, 29(23), 4084–4092. e4. <https://doi.org/10.1016/j.cub.2019.09.067>
- Clayton, M., Jakubowski, K., Eerola, T., Keller, P. E., Camurri, A., Volpe, G., & Alborn, P. (2020). Interpersonal entrainment in music performance. *Music Perception*, 38(2), 136–194. <https://doi.org/10.1525/mp.2020.38.2.136>
- Collins, R. (2004). *Interaction ritual chains*. Princeton University Press. <https://doi.org/10.1515/9781400851744>
- Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, 460(7256), 737–740. <https://doi.org/10.1038/nature08169>
- 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>
- Cross, I. (2009). The evolutionary nature of musical meaning. *Musicae Scientiae*, 13(2_suppl), 179–200. <https://doi.org/10.1177/1029864909013002091>
- Cross, I. (2010). Listening as covert performance. *Journal of the Royal Musical Association*, 135(S1), 67–77. <https://doi.org/10.1080/02690400903414848>
- Cross, I. (2012). Music as a social and cognitive process. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 313–328). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199553426.003.0033>
- Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, 42(6), 809–819. <https://doi.org/10.1177/0305735614543968>
- Cross, I. (2015). Music, speech and meaning in interaction. In C. Maeder & M. Reybrouck (Eds.), *Music, analysis, experience: New perspectives in musical semiotics* (pp. 19–30). Leuven University Press. <https://doi.org/10.2307/j.ctt180r0s2>
- Cross, I., & Blake, E. C. (2008). Flint tools as portable sound-producing objects in the upper palaeolithic context: An experimental study. In R. P. Paardekooper, P. Cunningham, & J. Heeb (Eds.), *Experiencing archaeology by experiment* (pp. 1–19). Oxbow Books.
- Damasio, A. R. (1999). *The feeling of what happens: body and emotion in the making of consciousness*. Vintage.
- Dams, L. (1985). Palaeolithic lithophones: descriptions and comparisons. *Oxford Journal of Archaeology*, 4(1), 31–46. <https://doi.org/10.1111/j.1468-0092.1985.tb00229.x>
- d’Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A.-M., Soressi, M., Bresson, F., Maureille, B., Nowell, A., Lakarra, J., Backwell, L., & Julien, M. (2003). Archaeological evidence for the emergence of language, symbolism, and music—an alternative multidisciplinary perspective. *Journal of World Prehistory*, 17(1), 1–70. <https://doi.org/10.1023/A:1023980201043>
- Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, 129(4), 2245–2252. <https://doi.org/10.1121/1.3562174>
- Dissanayake, E. (1999). Antecedents of the temporal arts in early mother-infant interaction. In S. Brown, B. Merker, & C. Wallin (Eds.), *The origins of music* (pp. 389–410). The MIT Press. <https://doi.org/10.7551/mitpress/5190.003.0027>
- Donald, M. (1991). *Origins of the modern mind: three stages in the evolution of culture and cognition*. Harvard University Press.
- Dor, D. (2015). *The instruction of imagination: language as a social communication technology*. Oxford University Press.
- Dor, D., & Jablonka, E. (2000). From cultural selection to genetic selection: a framework for the evolution of language. *Selection*, 1(1–3), 33–56. <https://doi.org/10.1556/Select.1.2000.1-3.5>
- 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. (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>
- Dyble, M. (2018). The effect of dispersal on rates of cumulative cultural evolution. *Biology Letters*, 14(2), 20180069. <https://doi.org/10.1098/rsbl.2018.0069>
- Everett, D. L. (2012). *Language: The cultural tool* (1st ed). Pantheon Books.
- Feld, S. (1984). Sound structure as social structure. *Ethnomusicology*, 28(3), 383. <https://doi.org/10.2307/851232>
- Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, 21(2), 80–99. <https://doi.org/10.1016/j.tics.2016.11.007>
- Fuentes, A. (2021). Searching for the “Roots” of masculinity in primates and the human evolutionary past. *Current Anthropology* 62(S23), S13–S25. <https://doi.org/10.1086/711582>
- Furuichi, T. (2011). Female contributions to the peaceful nature of bonobo society. *Evolutionary Anthropology: Issues, News, and Reviews*, 20(4), 131–142. <https://doi.org/10.1002/evan.20308>
- Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D. M., & Thomas, M. G. (2011). Evolution of

- lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566), 863–877. <https://doi.org/10.1098/rstb.2010.0268>
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday; /z-wcorg.
- Goren-Inbar, N., Alperson-Afil, N., Sharon, G., & Herzlinger, G. (2018). *The acheulian site of gesher benot Ya 'aqov volume IV: The lithic assemblages*. Springer International Publishing. <https://www.springer.com/gp/book/9783319740508>
- Grueter, C. C., Chapais, B., & Zinner, D. (2012). Evolution of multilevel social systems in nonhuman primates and humans. *International Journal of Primatology*, 33(5), 1002–1037. <https://doi.org/10.1007/s10764-012-9618-z>
- 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>
- Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., Boës, X., Quinn, R. L., Brenet, M., Arroyo, A., Taylor, N., Clément, S., Daver, G., Brugal, J.-P., Leakey, L., Mortlock, R. A., Wright, J. D., Lokorodi, S., Kirwa, C., ... Roche, H. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature*, 521(7552), 310–315. <https://doi.org/10.1038/nature14464>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion* (1. publ). Cambridge Univ. Press.
- Hawkins, S. (2014). Situational influences on rhythmicity in speech, music, and their interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130398. <https://doi.org/10.1098/rstb.2013.0398>
- Henrich, J. (2009). The evolution of costly displays, cooperation and religion. *Evolution and Human Behavior*, 30(4), 244–260. <https://doi.org/10.1016/j.evolhumbehav.2009.03.005>
- Hrdy, S. B. (2009). *Mothers and others*. Harvard University Press.
- Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. The MIT Press. <https://doi.org/10.7551/mitpress/6575.001.0001>
- Ingram, C. J. E., Raga, T. O., Tarekegn, A., Browning, S. L., Elamin, M. F., Bekele, E., Thomas, M. G., Weale, M. E., Bradman, N., & Swallow, D. M. (2009). Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *Journal of Molecular Evolution*, 69(6), 579–588. <https://doi.org/10.1007/s00239-009-9301-y>
- Jablonka, E. (2017). The evolution of linguistic communication: piagetian insights. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New perspectives on human development* (pp. 353–370). Cambridge University Press. <https://doi.org/10.1017/CBO9781316282755.019>
- Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT Press.
- Jones, M. R. (2016). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 125–142). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198722946.013.13>
- Juslin, P. N. (2019). *Musical emotions explained: unlocking the secrets of musical affect*. Oxford University Press.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Karakostis, F. A., Haeufle, D., Anastopoulou, I., Moraitis, K., Hotz, G., Tourloukis, V., & Harvati, K. (2021). Biomechanics of the human thumb and the evolution of dexterity. *Current Biology*, 31(6), 1317–1325.e8. <https://doi.org/10.1016/j.cub.2020.12.041>
- Key, A., Merritt, S. R., & Kivell, T. L. (2018). Hand grip diversity and frequency during the use of lower palaeolithic stone cutting-tools. *Journal of Human Evolution*, 125, 137–158. <https://doi.org/10.1016/j.jhevol.2018.08.006>
- Killin, A. (2016). Rethinking music's status as adaptation versus technology: a niche construction perspective. *Ethnomusicology Forum*, 25(2), 210–233. <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, 205920431775197. <https://doi.org/10.1177/2059204317751971>
- Knight, C., & Lewis, J. (2017). Wild voices: mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology*, 58(4), 435–453. <https://doi.org/10.1086/692905>
- Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, 23(1), 63–77. <https://doi.org/10.1016/j.tics.2018.10.006>
- Kuhn, S. L., & Stiner, M. C. (2019). Hearth and home in the middle pleistocene. *Journal of Anthropological Research*, 75(3), 305–327. <https://doi.org/10.1086/704145>
- Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, 23(1), 131–146. <https://doi.org/10.1017/S0140525X00002417>
- Langer, S. K. (1957). *Philosophy in a new key: A study in the symbolism of reason, rite and art*. Harvard University Press.
- Launay, J., Tarr, B., & Dunbar, R. I. M. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology*, 122(10), 779–789. <https://doi.org/10.1111/eth.12528>
- Layton, R., O'Hara, S., & Bilsborough, A. (2012). Antiquity and social functions of multilevel social organization among human hunter-gatherers. *International Journal of Primatology*, 33(5), 1215–1245. <https://doi.org/10.1007/s10764-012-9634-z>
- Lee, R. B. (1972). The intensification of social life among the! kung bushmen. In B. Spooner (Ed.), *Population growth: anthropological implications* (pp. 343–350). MIT Press.
- Levinson, S. C. (2006). On the human “interaction engine. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: culture, cognition and interaction* (pp. 39–69). Berg.
- Levinson, S. C., & Holler, J. (2014). The origin of human multimodal communication. *Philosophical Transactions of the*

- Royal Society B: Biological Sciences*, 369(1651), 20130302. <https://doi.org/10.1098/rstb.2013.0302>
- Lewis, J. (2009). As well as words: Congo pygmy hunting, mimicry, and play. In R. Botha & C. Knight (Eds.), *The cradle of language* (pp. 236–256). Oxford University Press.
- Lewis, J. (2013). A cross-cultural perspective on the significance of music and dance to culture and society. In M. A. Arbib (Ed.), *Language, music, and the brain* (pp. 45–66). The MIT Press. <https://doi.org/10.7551/mitpress/9780262018104.003.0002>
- Lewis, J. (2014). Bayaka pygmy multi-modal and mimetic communication traditions. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language* (pp. 77–91). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199665327.003.0007>
- Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: evidence from mandarin speakers. *PLOS ONE*, 7(2), e30374. <https://doi.org/10.1371/journal.pone.0030374>
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1), e8559. <https://doi.org/10.1371/journal.pone.0008559>
- Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: implications for the appreciation and aesthetics of music. *Frontiers in Neuroscience*, 11, <https://doi.org/10.3389/fnins.2017.00147>
- Mauss, M., & Beuchat, H. (1979). *Seasonal variations of the eskimo: A study in social morphology*. Routledge & Kegan Paul.
- McPherron, S. P., Alemseged, Z., Marean, C. W., Wynn, J. G., Reed, D., Geraads, D., Bobe, R., & Béarat, H. A. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at dikika, Ethiopia. *Nature*, 466(7308), 857–860. <https://doi.org/10.1038/nature09248>
- 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. Cambridge Core. <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(6468). <https://doi.org/10.1126/science.aax0868>
- Merker, B. (1999). Synchronous chorusing and the origins of music. *Musicae Scientiae*, 3(1_suppl), 59–73. <https://doi.org/10.1177/10298649000030S105>
- Merriam, A. P. (1964). *The anthropology of music (6. Paperback print)*. Northwestern Univ. Press.
- Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.
- Migliano, A. B., Battiston, F., Viguier, S., Page, A. E., Dyble, M., Schlaepfer, R., Smith, D., Astete, L., Ngales, M., Gomez-Gardenes, J., Latora, V., & Vinicius, L. (2020). Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution. *Science Advances*, 6(9), eaax5913. <https://doi.org/10.1126/sciadv.aax5913>
- Migliano, A. B., Page, A. E., Gómez-Gardeñes, J., Salali, G. D., Viguier, S., Dyble, M., Thompson, J., Chaudhary, N., Smith, D., Strods, J., Mace, R., Thomas, M. G., Latora, V., & Vinicius, L. (2017). Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour*, 1(2), 0043. <https://doi.org/10.1038/s41562-016-0043>
- Miller, G. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. Heinemann.
- Mithen, S. J. (2005). *The singing neanderthals: The origins of music, language, mind and body*. Weidenfeld & Nicolson.
- Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, 72, 13–20. <https://doi.org/10.1016/j.jesp.2017.03.009>
- Nettl, B. (2015). *The study of ethnomusicology: thirty-three discussions (third edition)*. University of Illinois Press.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1597), 1829–1836. <https://doi.org/10.1098/rstb.2011.0216>
- Norman, L. J., & Thaler, L. (2019). Retinotopic-like maps of spatial sound in primary 'visual' cortex of blind human echolocators. *Proceedings of the Royal Society B: Biological Sciences*, 286(1912), 20191910. <https://doi.org/10.1098/rspb.2019.1910>
- Oloa-Bilola, C. (2017). The egalitarian body. A study of aesthetic and emotional processes in massana performances among the Mbendjéle of the Likouala region (Republic of Congo) [Doctoral thesis, UCL (University College London)]. <https://discovery.ucl.ac.uk/id/eprint/1522643/>.
- Packer, C., Swanson, A., Ikanda, D., & Kushnir, H. (2011). Fear of darkness, the full moon and the nocturnal ecology of African lions. *PLoS ONE*, 6(7), e22285. <https://doi.org/10.1371/journal.pone.0022285>
- 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.
- 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. <https://doi.org/10.3389/fnins.2014.00057>
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 19(10), 827–830. <https://doi.org/10.1016/j.cub.2009.03.038>
- Patterson, F. G. P., & Cohn, R. H. (1990). Language acquisition by a lowland gorilla: koko's first ten years of vocabulary development. *Word*, 41(2), 97–143. <https://doi.org/10.1080/00437956.1990.11435816>
- Pearce, E., Launay, J., & Dunbar, R. I. M. (2015). The ice-breaker effect: singing mediates fast social bonding. *Royal Society*

- Open Science*, 2(10), 150221. <https://doi.org/10.1098/rsos.150221>
- Pepperberg, I. M. (1999). *The alex studies: cognitive and communicative abilities of grey parrots*. Harvard University Press.
- Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., & Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, 107(10), 4758–4763. <https://doi.org/10.1073/pnas.0909074107>
- Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS ONE*, 6(11), e27241. <https://doi.org/10.1371/journal.pone.0027241>
- Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences*, 20(11), 857–867. <https://doi.org/10.1016/j.tics.2016.09.002>
- Pinker, S. (1997). *How the mind works*. Norton.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4), 707–727. <https://doi.org/10.1017/S0140525X00081061>
- Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. *Frontiers in Neuroscience*, 11, <https://doi.org/10.3389/fnins.2017.00542>
- Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: a comparative study of tapping synchronization. *Music Perception*, 36(1), 1–23. <https://doi.org/10.1525/mp.2018.36.1.1>
- Power, C. (2014). The evolution of ritual as a process of sexual selection. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language* (pp. 196–207). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199665327.003.0007>
- Rabinowitch, T.-C., Cross, I., & Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. *Psychology of Music*, 41(4), 484–498. <https://doi.org/10.1177/0305735612440609>
- Ravignani, A., Bowling, D. L., & Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in Psychology*, 5, 1118. <https://doi.org/10.3389/fpsyg.2014.01118>
- Richmond, B. G., Roach, N. T., & Ostrofsky, K. R. (2016). Evolution of the early hominin hand. In T. L. Kivell, P. Lemelin, B. G. Richmond, & D. Schmitt (Eds.), *The evolution of the primate hand: anatomical, developmental, functional, and paleontological evidence* (pp. 515–543). Springer New York. https://doi.org/10.1007/978-1-4939-3646-5_18
- Robledo, J. P., Hawkins, S., Cornejo, C., Cross, I., Party, D., & Hurtado, E. (2021). Musical improvisation enhances interpersonal coordination in subsequent conversation: motor and speech evidence. *PLOS ONE*, 16(4), e0250166. <https://doi.org/10.1371/journal.pone.0250166>
- Robledo, J. P., Hawkins, S., Cross, I., & Ogden, R. A. (2016). Pitch-interval analysis of “periodic” and “aperiodic” Question + Answer pairs. *Proc. Speech Prosody*, 2016, 1071–1075. <https://doi.org/10.21437/SpeechProsody.2016-220>
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, 14(2), 257–262. <https://doi.org/10.1038/nn.2726>
- Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences*, 19(2), 86–91. <https://doi.org/10.1016/j.tics.2014.12.001>
- Salimpoor, V. N., & Zatorre, R. J. (2013). Neural interactions that give rise to musical pleasure. *Psychology of Aesthetics, Creativity, and the Arts*, 7(1), 62–75. <https://doi.org/10.1037/a0031819>
- 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*, 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. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, 44, e59. Cambridge Core. <https://doi.org/10.1017/S0140525X20000333>
- Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The Ape at the brink of the human mind*. John Wiley & Sons.
- Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19(10), 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>
- Schäfer, T., Fachner, J., & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology*, 4, <https://doi.org/10.3389/fpsyg.2013.00508>
- Schlichting, C. D., & Wund, M. A. (2014). Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution*, 68(3), 656–672. <https://doi.org/10.1111/evo.12348>
- Schultz, W. (2017). Reward prediction error. *Current Biology*, 27(10), R369–R371. <https://doi.org/10.1016/j.cub.2017.02.064>
- Seeger, A. (1987). *Why suyá sing: A musical anthropology of an amazonian people*. Cambridge Univ. Press.
- Ségurel, L., & Bon, C. (2017). On the evolution of lactase persistence in humans. *Annual Review of Genomics and Human Genetics*, 18(1), 297–319. <https://doi.org/10.1146/annurev-genom-091416-035340>
- Senft, G. (2018). Pragmatics and anthropology: The trobriand Islanders’ ways of speaking. In C. Ilie & N. R. Norrick (Eds.), *Pragmatics & beyond New series* (Vol. Vol. 294, pp. 185–211). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.294.09sen>
- Shilton, D. (2019). Is language necessary for the social transmission of lithic technology? *Journal of Language Evolution*, 4(2), 124–133. <https://doi.org/10.1093/jole/lzz004>
- Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human social evolution: self-domestication or self-control? *Frontiers in Psychology*, 11, 134. <https://doi.org/10.3389/fpsyg.2020.00134>
- Shimelmitz, R., Kuhn, S. L., Jelinek, A. J., Ronen, A., Clark, A. E., & Weinstein-Evron, M. (2014). ‘Fire at will’: The

- emergence of habitual fire use 350,000 years ago. *The Role of Freshwater and Marine Resources in the Evolution of the Human Diet, Brain and Behavior*, 77, 196–203. <https://doi.org/10.1016/j.jhevol.2014.07.005>
- Shott, M. (2004). Hunter-gatherer aggregation in theory and evidence: The north American paleoindian case. In G. Crothers (Ed.), *Hunter-Gatherers in theory and archaeology* (pp. 68–102). Southern Illinois University, Center for Archaeological Investigations.
- Simchy-Gross, R., & Margulis, E. H. (2018). The sound-to-music illusion: repetition can musicalize nonspeech sounds. *Music & Science*, 1, 205920431773199. <https://doi.org/10.1177/2059204317731992>
- Slobin, M. (2002). *Chosen voices: The story of the American cantorate* (1st pbk. ed). University of Illinois Press.
- Small, C. (1998). *Musicking: The meanings of performing and listening*. University Press of New England.
- Stanford, C. B. (2018). *The new chimpanzee: A twenty-first-century portrait of our closest kin*. Harvard University Press.
- Sterelny, K. (2020). Innovation, life history and social networks in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1803), 20190497. <https://doi.org/10.1098/rstb.2019.0497>
- Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1585), 75–87. <https://doi.org/10.1098/rstb.2011.0099>
- Stout, D., & Khreisheh, N. (2015). Skill learning and human brain evolution: an experimental approach. *Cambridge Archaeological Journal*, 25(4), 867–875. <https://doi.org/10.1017/S0959774315000359>
- Stout, D., Rogers, M. J., Jaeggi, A. V., & Semaw, S. (2019). Archaeology and the origins of human cumulative culture: a case study from the earliest oldowan at gona, Ethiopia. *Current Anthropology*, 60(3), 309–340. <https://doi.org/10.1086/703173>
- Tarr, B., Launay, J., Cohen, E., & Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology Letters*, 11(10), 20150767. <https://doi.org/10.1098/rsbl.2015.0767>
- Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2405–2415. <https://doi.org/10.1098/rstb.2009.0052>
- Thaler, L., Arnott, S. R., & Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE*, 6(5), e20162. <https://doi.org/10.1371/journal.pone.0020162>
- Thaler, L., & Goodale, M. A. (2016). Echolocation in humans: an overview: echolocation in humans. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(6), 382–393. <https://doi.org/10.1002/wcs.1408>
- Tomlinson, G. (2015). *A million years of music: The emergence of human modernity*. Zone Books. <https://doi.org/10.2307/j.ctt17kk95h>
- Turino, T. (2008). *Music as social life: The politics of participation*. University of Chicago Press.
- Vuust, P., & Frith, C. D. (2008). Anticipation is the key to understanding music and the effects of music on emotion. *Behavioral and Brain Sciences*, 31(5), 599–600. <https://doi.org/10.1017/S0140525X08005542>
- Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution*, 7(2), 118–126. <https://doi.org/10.1111/j.1558-5646.1953.tb00070.x>
- Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I. M., & Stewart, L. (2016). Singing and social bonding: changes in connectivity and pain threshold as a function of group size. *Evolution and Human Behavior*, 37(2), 152–158. <https://doi.org/10.1016/j.evolhumbehav.2015.10.002>
- Wen, N. J., Willard, A. K., Caughy, M., & Legare, C. H. (2020). Watch me, watch you: ritual participation increases in-group displays and out-group monitoring in children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1805), 20190437. <https://doi.org/10.1098/rstb.2019.0437>
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences*, 102(Supplement 1), 6543–6549. <https://doi.org/10.1073/pnas.0501844102>
- Whiteman, R. (2020). Structuring social relationships: Music-making and group identity [Doctoral thesis, Cambridge University]. <https://www.repository.cam.ac.uk/handle/1810/310994>.
- Wiessner, P. W. (2014). Embers of society: firelight talk among the Ju/'hoansi bushmen. *Proceedings of the National Academy of Sciences*, 111(39), 14027–14035. <https://doi.org/10.1073/pnas.1404212111>
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, 106(7), 2468. <https://doi.org/10.1073/pnas.0809035106>