

Group dancing as the evolutionary origin of rhythmic entrainment in humans

Steven Brown

Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Keywords:

Rhythm
Entrainment
Dance
Music
Body percussion
Evolution

ABSTRACT

An ecologically-valid approach to the evolutionary origins of rhythmic entrainment in humans has to address not one but two key issues: first, the capacity to *generate* acoustic rhythms, and second, the ability to *entrain* body movements to them. Most research in this area has ignored the first issue altogether and has instead placed all of the emphasis on motor entrainment skills per se. But this begs the question of how auditory rhythms came to be generated in the first place. I discuss evolutionary models that explicitly link the mechanisms of body entrainment to the mechanisms of sound generation. The most plausible models are those in which these processes occur interactively and mutually through group dancing, employing not only visual and haptic cues for entrainment but percussive sounds generated through body movements, most especially locomotor movements. Body percussion during movement creates a link between motor and sensory components of interpersonal entrainment.

1. Who is producing the beat?

The evolution of the novel ability of humans to entrain movements to rhythmic timekeepers is discussed almost exclusively in relation to music, since the ability to “keep the beat” is seen as an evolutionary hallmark of human musical capacity. However, I will argue in this article that this ability should be more properly thought of in the context of dance. In recent years, there has been much discussion about the ability of certain non-human animal species to entrain body movements to the beat of human music (Honing, 2019; Merchant & Honing, 2014; Patel, 2014; Wilson & Cook, 2016). Notable among them is the ability of a cockatoo and parrot to entrain their body movements to the beat of pop music (Patel, Iversen, Bregman, & Schulz, 2009; Schachner, Brady, Pepperberg, & Hauser, 2009), and the training of a sea lion to bob her head to the beat of pop music (Cook, Rouse, Wilson, & Reichmuth, 2013).

However, this interest in animal behavior highlights the caveat that any ecologically-valid discussion of the biological basis of rhythmic entrainment to an acoustic beat has to address not one but two key issues: first, the capacity to *generate* acoustic rhythms to begin with (i.e., to produce the entrainment signal), and second, the ability to *entrain* body movements to such rhythms. Demonstrations that non-human animals can entrain their body movements to the beat of a metronome or to recordings of music do not address the first point and instead place all the focus on entrainment skills alone. In addition, these demonstrations

come with the implicit assumption that entrainment is limited to the acoustic modality, since studies in this area invariably attempt to make a phylogenetic linkage to human musicality. In other words, these models assume that entrainment had to have evolved as a component of musical capacity, rather than some other aspect of social coordination.

The comparison between the capacity to generate metric rhythms and the capacity to entrain movements to them underscores a huge disconnect between these two capacities in the animal world. Virtually all animals engage in locomotion in a regular, metric manner, whether this be through the terrestrial locomotor movements of quadruped and biped mammals, the flying movements of birds, or the swimming movements of fish. Central pattern generators at the level of the spinal cord and brain stem are sufficient to produce alternating activation patterns between the right and left sides of the body, or the simultaneous activation of the two sides of the body in the case of flight (Grillner, 2006; Katz, 2016). Hence, the capacity to generate metric movement patterns seems to be widespread in the animal world.

This stands in stark contrast to the capacity to *synchronize* movements (locomotor or otherwise) to an external acoustic beat. This does indeed seem to be nearly a human novelty, with a few sparse exceptions occurring in the animal world, the best described examples being a cockatoo, parrot, and sea lion, all of them raised in human settings. So, while metric locomotion seems to be widespread, metric entrainment to an external beat appears to be nearly human-specific. Since human music is not part of the evolutionary history of cockatoos, parrots, and

E-mail address: stebro@mcmaster.ca.

<https://doi.org/10.1016/j.newideapsych.2021.100902>

Received 22 March 2021; Received in revised form 20 June 2021; Accepted 30 July 2021
0732-118X/© 2021 Elsevier Ltd. All rights reserved.

sea lions, the capacity of these species to entrain body movements to a musical rhythm raises more questions than it answers. Who is generating the acoustic beat in the wild and how is the sound being generated? To the best of our knowledge, the answer to this question for cockatoos, parrots, sea lions, and the other species that demonstrate the ability to entrain movements to metronome beats or music is “no one”. In other words, there is no evidence that such species produce metric movements or vocalizations in the wild that are sonorant and that can serve as acoustic entrainment-signals for other members of the species to dance to.

In contrast to this, there is a well-described class of non-locomotor behaviors that does show the joint features of being metric and sonorant. These are the communal mating calls of a large number of both invertebrate and vertebrate species, including the stridulatory calls of insects and the vocal advertisement calls of amphibians (Grafe, 1999; Greenfield, 2015; Greenfield & Roizen, 1993; Greenfield & Schul, 2008). The stridulatory sounds of crickets that can be heard on warm summer evenings provide an excellent metronomic beat for crickets to dance to. The problem is that there is no evidence that they do. These calls, generally produced by the male of the species, attract females to the calling male, but they do not induce movements in the female’s body that are entrained to the tempo of the male’s call.

We thus see another significant disconnect in the animal world, namely a large number of species that can produce metric and sonorant calls, but with no evidence that members of these species synchronize body movements to them through dancing. This is the reciprocal problem to the dancing cockatoo that is able to move to an acoustic beat, but does not seem to be able to produce one. In other words, we have an opposition between the “metronomic caller”, who can produce an acoustic beat but who does not (or cannot) dance to one, and the “dancing animal”, who can move to an acoustic beat but who does not (or cannot) produce one in the wild. Only humans appear to have *co-evolved* the capacity to generate metrically-timed sounds with the ability to entrain body movements to them, as seen in the universal phenomenon of dancing to a musical beat.

This capacity for interpersonal coordination in time in humans serves a critical adaptive role in group functionality, such as the promotion of cooperation, altruism, cohesion, and group identity (Anshel & Kipper, 1988; Cirelli, Einarson, & Trainor, 2014; L.; Cross, Turgeon, & Atherton, 2019; Dunbar, Kaskatis, MacDonald, & Barra, 2012; Fischer, Callander, Reddish, & Bulbulia, 2013; Good & Russo, 2016; Hove & Risen, 2009; Kirschner & Tomasello, 2010; Kniffin, Yan, Wansink, & Schulze, 2017; Launay, Tarr, & Dunbar, 2016; Mehr, Krasnow, Bryant, & Hagen, 2021; Pearce, Launay, & Dunbar, 2015; Reddish, Fischer, & Bulbulia, 2013; Rennung & Göritz, 2016; Savage et al., 2021; von Zimmermann, Vicary, Sperling, Orgs, & Richardson, 2018; Weinstein, Launay, Pearce, Dunbar, & Stewart, 2016; Wiltermuth & Heath, 2009). The short-term intervention of synchronized movement and/or vocalizing with others has the long-term benefit of fostering a communitarian sense of belonging to the group and of supporting cooperative endeavors with group members, not least with non-kin. Along these lines, music and dance most likely evolved as a type of territorial display to defend group territories on a year-round basis (Brown, 2000, 2007; Geissmann, 2000a; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Mehr et al., 2021).

Interestingly, many of the communal metronomic callers just mentioned do show a form of rhythmic entrainment, but it is not the kind that cockatoos show. It is instead an entrainment of acoustic calling to acoustic calling. Males of several species demonstrate the ability to phase-lock their calls to the calls of other males (Greenfield & Roizen, 1993; Greenfield, 2005; see Buck and Buck, 1978 for a similar phase-locking mechanism in the communal flashing of fireflies). This is not the cooperative phase locking that underlies the unison singing of *Happy Birthday* in humans, but instead a competitive kind that aims to have the caller be heard slightly ahead of another caller and thereby grab the attention of a potential mate. It has been referred to as a “jamming” mechanism (Greenfield & Roizen, 1993; Sheridan, Brookes,

& Pomiankowski, 1993), but the term here connotes “signal jamming” in the competitive sense, not “musical jamming” in the cooperative sense. Overall, we have evidence in these species of entrainment to acoustic signals, although it is not an entrainment of body movement per se, but instead an entrainment of sonorant signals to other sonorant signals, hence leading to the generation of a type of synchronous chorus, although quite different in mechanism and function from the cooperative chorusing found universally in human cultures (Ravignani, Bowling, & Fitch, 2014).

2. Group dancing as the origin of rhythmic entrainment in humans

In this section, I will examine evolutionary models for the human capacity to entrain body movements to acoustic timekeepers. While this has always been conceptualized as being related to the evolution of music, I see it as being about the evolution of *dance* pure and simple, most especially group dancing. A voluminous literature on finger tapping to metronome beats in humans has failed to make this point, since the work has always been framed as research on “music” or “musical rhythm”. Hence, the connection to dance has been obscured. To me, finger tapping to a metronome beat is nothing if not a simple form of dancing. Likewise, the evolutionary issue of entrainment of body movements to acoustic beats is first and foremost about dance. *There is a genuine need to reframe this issue from one about music to one about dance.*

Before discussing the models, it is important to explain what I mean by music and dance, since these functions will be both distinguished from and intertwined with one another in what follows. In addition, anthropological evidence suggests that cultural definitions of music and dance may be highly interrelated. My goal in presenting evolutionary models here will be in analyzing music and dance behaviorally, rather than culturally. I take the stance that the single defining feature of music is *tonality*, in other words the use of relatively discrete intervals in both production and perception, leading ultimately to scale patterns. This view is in no way grounded in Western heptatonic scales or harmonic progressions, but simply in the idea that music is based on a digitization of pitch-space in some manner, whether that be a monotone chant, a two-note scale, or an Indian raga. In contrast to this, I do not see rhythm as being a defining feature of music at all, but instead a function that is shared between dance, music and metric forms of speech. Rhythm has no necessary connection with pitch and melody. The parameters that are associated with the concept of rhythm in musicology, such as metrical structure, meter types, tempo, and articulation (Jones, 2019), are comparably present in dance, music, and poetry. Only tonality defines music. This is reinforced by the abundance of non-metric musical forms across cultures that are in a free meter (Clayton, 1996).

While music is about the generation of patterned sounds (tonal sounds in particular), dance is likewise about the generation of patterned body movements, either by individuals or groups. Judith Lynne Hanna, in her classic book *To Dance is Human* (1979), defines dancing as culturally-patterned movement that is rhythmic, communicative, and aesthetically valued, in other words as a form of nonverbal communication. Most theorists conceive of dance as a distinct means of engaging in body movement from the functional movement patterns of everyday life, such as locomotor patterns and instrumental actions. This “remoteness from the habitual” (Kreitler & Kreitler, 1972, p. 163) is a key aspect of dance as a conventionalized system of body movements that is highly structured through elaborate choreography compared to everyday movements. It is also found in the fact that dancing tends to occur in ritualized social contexts, that it generally incorporates music, and that it is often carried out by multiple individuals who synchronize their movements with one another. While I will be arguing below that the temporal dimension of dance and music is shared, their spatial dimensions are distinct: physical space in the case of dance and tonal pitch-space in the case of music. Some theorists have pointed to parallel forms of motion in these two spaces, such as when a rising body

movement is choreographed to an ascending melodic contour (Sachs, 1943; von Hornbostel, 1904).

While dance can unquestionably be done by soloists, I do believe that an evolutionary understanding of dance requires that we see it as a form of *cooperative joint action*. Group dancing is one of the most important mechanisms by which people are able to develop what Pacherie (2011:343) describes as “a sense of we-ness at the expense of a well-defined sense of self”. William McNeill (1995) drives home the same point in his book *Keeping Together in Time*, in which he describes how the synchronous movement of groups, such as the marching of soldiers, instills a feeling of unity and common cause for collective actions that is achievable by few other means in human behavior. Dance is an important behavior for achieving intercorporeality between co-participants (Meyer, Streeck, & Jordan, 2017) as well as a sense of kinaesthetic empathy between performers and spectators (Jola, Ehrenberg, & Reynolds, 2012). My aim is not to detract from the importance of the individual in dance, but to highlight that dance’s most human-specific feature is interpersonal synchronization of body movement.

It is important to state that my objective in this article is to provide a phylogenetic, rather than functional, analysis of the origins of rhythmic entrainment. My focus will be on proposing a progression of precursor capacities leading to the current human ability to produce rhythm in music and dance, rather than on justifying the emergence of this trait in functional, Darwinian terms. I abide by the view mentioned above that dance and music are coordinative functions that foster group cohesion and territory maintenance, two functions that tend to co-occur in animal species (Dahlin & Benedict, 2014; Geissmann, 2000a, 2000b; Hall, 2009; Tobias et al., 2016). I have written in detail about why this is a better model than a sexual display account of music and dance (Brown, 2000). The question of why humans developed rhythmic entrainment, but not other species with similar sensorimotor abilities and ecologies, is a topic of intense investigation. In my view, the capacities for synchronous interpersonal movement and chorusing are group-adaptive traits (Smaldino, 2014) that serve as both a reflection and reinforcer of the lifeways of a highly cooperative group-living species like *Homo sapiens*. Rhythmic-entrainment abilities are far more mysterious in the case of animals like cockatoos and sea lions that do not have such social ecologies.

2.1. Voice First and Body First models of entrainment

In order to develop models for the evolutionary origins of rhythmic entrainment in humans, we need to start by distinguishing potential *sound sources* from potential *effectors* for entrainment. The same three general parts of the body can serve as either or both. These are the voice, the upper limbs, and the lower limbs. Sounds can be generated by the vocal tract through phonation, non-phonatory noise (like the “sh” sound), and articulatory movements (e.g., clicks); they can be generated by the hands using one hand (e.g., snapping), two hands (e.g., clapping), or by using the hands to strike a part of the body (e.g., chest beating); and they can be generated by the legs while either standing in place (stomping, jumping) or by moving about, such as through walking, skipping, or moving side to side. While human entrainment often occurs to instrumental musical sounds, I am going to leave melodic instruments out of the present discussion and consider the voice as the main pitch-generating instrument in these models, since it is quite reasonable to assume that the voice was the first pitch-generating instrument in humans (Cross & Morley, 2009; Montagu, 2017; Morley, 2014; Sachs, 1943), just as it is in other animal species like songbirds. However, I will make mention of percussion instruments, and for this we should include Jordania’s (2014) proposal that the hitting together of stones can create a powerful acoustic signal.

Looking now to the parts of the body whose movements can become rhythmically entrained to a sound source, the voice serves as that effector during chorusing. However, it is typically the rest of the body

that synchronizes to a beat during dancing and other standard forms of audiomotor entrainment. Hence, for the purposes of the present discussion, I will mainly consider the voice to be a source factor alone, and consider the upper and lower limbs (as indicators of the body outside of the vocal tract) to be the main effectors that synchronize with beats during entrainment, although I will make mention below of a potential role of vocal-tract percussion in the dance model being developed here. I will also consider a purely voice-based model in which entrainment evolved through chorusing alone, without a consideration for dance.

Before discussing the actual models of entrainment, I would like to consider three general features of the models that will be presented. 1) *Solo vs. group agents*. I will consider two types of models for the agents that engage in entrainment. In a solo model, synchronization occurs at the within-individual level through *self-entrainment*. This results from synergies between the body parts that serve as the sound source and effector, respectively (Clayton, 2012; Latash, Scholz, & Schöner, 2007). In a group model, by contrast, entrainment emerges at the between-individual level through group interaction. This retains the intrapersonal synergies of solo models, but supplements them with interpersonal-coordination mechanisms as well (Keller, Novembre, & Hove, 2014; Sebanz, Bekkering, & Knoblich, 2006). A group model, unlike a solo model, is predicated on social entrainment (Kim, Reifgerst, & Rizzonelli, 2019). 2) *Source/effector relationship*. In an “iso-effector” model, the sound source and the effector that entrains to it are the same type of body part (for example, a hand source for a hand effector), whereas in a “cross-effector” model, they are different body parts (for example, a vocal sound source for a leg effector). 3) *Voice First vs. Body First sound-source scenarios*. In a Voice First model, acoustic rhythms are generated by the voice, and other parts of the body entrain to the beat generated by these vocal rhythms. This is by necessity a cross-effector model. In a Body First model, acoustic rhythms are generated by percussive body sounds, to which entrainment can be achieved either by that same body part (iso-effector model) or by another body part (cross-effector model). I will now examine three evolutionary models for the origins of metric entrainment, beginning with the Solo model.

Solo model. The solo model of entrainment is the model of self-entrainment. There are two variants of it. In the Voice First version, humans evolve the capacity to generate rhythmic sounds with their voice,¹ and acquire a second capacity to entrain the movement of various body parts to one’s own vocalizing, for example tapping one’s foot to one’s rhythmic vocal sounds. This is seen in modern times in North American indigenous cultures where solo singers beat a frame drum in coordination with their vocal rhythm (Browner, 2000). In the Body First version of the solo model, humans evolve the capacity to generate rhythmic percussive sounds with the body, to which other body parts entrain in a cross-effector manner, as in tapping one’s foot to the sound of one’s finger snaps. It is important to note that the iso-effector version of self-entrainment makes no sense, since there is simply the rhythmic movement of a single effector, but no true sense of entrainment, since nothing is being entrained to. Therefore, a solo model has to be a cross-effector model. I present the solo model for completeness sake, but it is not a very plausible explanation for the group-wide coordination that is so prominent in human cultures. A better model of entrainment is a group model, of which there are two versions. Both of them are predicated on the idea that coordinated body movement among two or more individuals is an act of *group dancing*.

Voice First group model. As in the solo version of the Voice First model, humans evolve the capacity to generate rhythmic sounds with their

¹ There are forms of birdsong that are metric and that are performed by solo singers. Likewise, MacNeilage (1998, 2008) proposed that syllable structure in speech had its origin in rhythmic oscillations of the jaw, such as the kind that underlies the babbling of babies. While I favor a body-based model of the origins of rhythm, I cannot rule out the possibility that rhythm began with the vocal tract.

voice, but this now occurs in the form of a *chorus* as a result of mutual entrainment among group members. This rhythmic chorus becomes the sound-source to which other group members synchronize their body movements through external entrainment, making this a cross-effector model, as shown in Fig. 1. The model essentially calls for a differentiation between “musicians” and “dancers” during a given ritual, where musicians are defined as people who generate acoustic beats without necessarily dancing to them, and dancers are defined as people who move to acoustic beats without necessarily generating them. However, all people evolve the two separate capacities to 1) vocalize rhythmically, and 2) entrain body movements to vocal rhythms. As mentioned earlier, the source musicians could be instrumentalists instead of vocalists, but I do believe that this complicates the issue without offering any clear benefit. Note that if the dancers were only entraining body movements to their own vocalizations, then this would simply reduce to the solo version of the Voice First model, whereas what I am presenting here is the group version of it in which the rhythmic sound source evolves through chorusing.

Body First group model. As in the solo version of the Body First model, humans evolve the capacity to generate rhythmic percussive sounds with the body, but this now occurs *interactively* as a result of mutual entrainment among group members during group dancing (Fig. 2). One important outcome of this arrangement, compared to the Voice First group model, is that entrainment occurs using both visual and auditory cues, resulting in both visuomotor and audiomotor entrainment. Experimental finger-tapping studies with humans have shown that the standard tapping advantage for auditory cues over visual cues disappears when the visual stimulus is in motion (Hove, Fairhurst, Kotz, & Keller, 2013). In addition, interpersonal synchronization between a pair of individuals performing body entrainment to a metronome beat is enhanced when the partners can see one another, compared to when they cannot (Miyata, Varlet, Miura, Kudo, & Keller, 2017). This again conforms with the predictions of a group-dancing model that the individuals entraining to one another should be in motion and should be within eyesight of one another.

The basic premise of the Body First model is that *body percussion serves as the sound source for entrainment*. One variant of this model, what I will call the Locomotor model, is an iso-effector group model in which the lower extremity is both the source and effector of entrainment, but where these functions are *spread across group members through mutual interaction during group dancing* (in contrast to a solo self-entrainment model, where an iso-effector mechanism makes no sense). Such a mechanism could be reinforced by attaching simple percussive objects to the body, such as in the form of leggings made with beads or shells. Another variant of the Body First model could be a cross-effector version, where rhythms are generated with the hands, and where

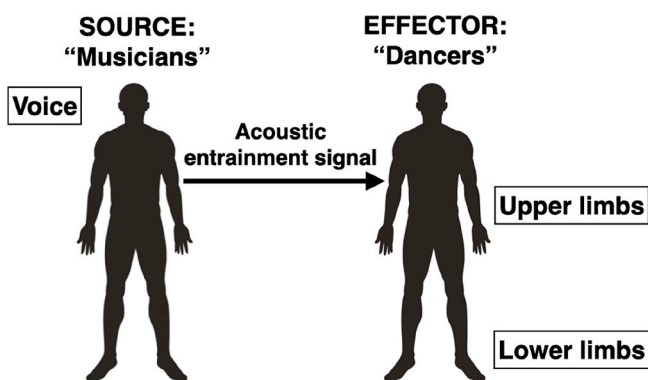


Fig. 1. The Voice First model. In the Voice First model of rhythmic entrainment, rhythmic structure first evolves in the context of music (via vocal chorusing), and dancers secondarily evolve the capacity to entrain to the beat of music using the body.

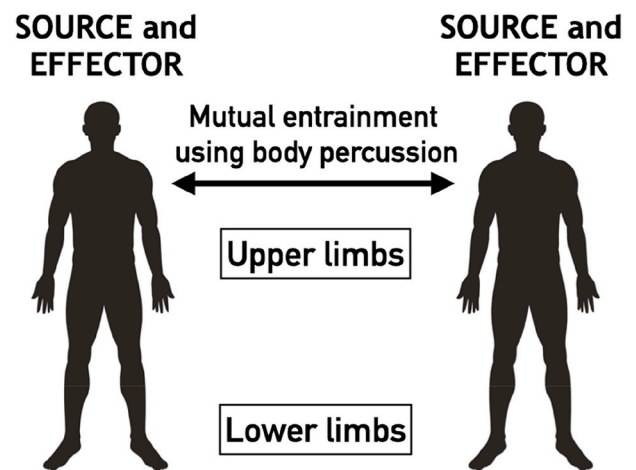


Fig. 2. The Body First model. Body percussion generated through body movements acts as both the source and the effector for entrainment through mutual coupling.

entrainment occurs with the legs (or vice versa).

2.2. The Locomotor model of the origins of rhythmic entrainment

I would like to consider the Locomotor (iso-effector) version of the Body First model as a promising candidate for the origins of human entrainment through group dancing, where the lower extremity is both the sound source and effector of entrainment. Larsson (2014) made a compelling case for the idea that the “predictable sounds of locomotion may stimulate the evolution of synchronized behavior” and that “bipedal gait and the associated sounds of locomotion influenced the evolution of human rhythmic abilities” (p. 1). He pointed out that locomotion typically creates “audible sounds containing a number of qualitatively dissimilar acoustical events: isolated impulsive signals, sliding sounds, crushing sounds, and complex temporal patterns of overlapping impulsive signals” (p. 4). This idea need not apply to locomotion alone, but can be extended to include any kinds of full-body movements that result in contact with the ground, by either standing in place (e.g., jumping, stomping) or moving about. Buttress drumming in chimpanzees is one such example using the limbs to contact tree buttresses (Arcadi, Robert, & Boesch, 1998). So, the model I have in mind has perhaps less to do with locomotion in isolation as with the ability of groups of individuals who are simultaneously engaging in rhythmic full-body movements of all kinds to generate percussive sounds that can be entrained to in an interactive and mutual manner, both acoustically and visually. Walking in pace may be but one example of this (Larsson, Richter, & Ravignani, 2019). This idea conforms with the proposal of Jordania (2014) that the evolution of group-coordinative behaviors like music and dance is related to a “mobbing” mechanism directed at predatory species. This mechanism involves a combination of chorusing and percussive stomping movements as a means of scaring off or confusing predators. A key point of Jordania’s model is that group dancing need not be limited to group-internal effects, but that it can have an impact external to the group in the form of group displays, including effects on both competing human groups and predatory animal species.

Let me elaborate on this group dancing model by considering some of its important features. 1) Bipedal locomotion is a novel human trait and so is dance. The model creates a causal link between these two novel human capacities that employ the lower extremities (Larsson et al., 2019; Mithen, 2005). 2) Many dances are in fact based on walking patterns. While dancing can engage virtually any part of the body – making dance the most complex and diverse type of body movement that humans engage in – locomotor patterns are the foundation for many

forms of dance. Likewise, many dances constrain the upper body in some manner, making the lower body the main effector for entrainment. Sitting dances are the principal exceptions to this (Sachs, 1937). 3) The interactive nature of the model establishes dancing as a group phenomenon from its inception. This is consistent with the fact that group dancing is found universally across human cultures (Sachs, 1937), and that it is far more prominent than solo dancing. 4) The interactive nature of the model also implies that rhythm evolved as a social phenomenon from its inception. This is a point that is easily lost in contemporary discussions of finger tapping and dancing cockatoos. Strogatz (2003), in talking about the mutual mechanisms by which fireflies are able to achieve synchrony in their flashing, says that “everyone is continually sending and receiving signals, shifting the rhythms of others and being shifted by them in turn” (p. 13). There is no question that the evolution of metric entrainment required an interactive arrangement in which “everyone is continually sending and receiving signals” in a mutual manner. 5) The partnering arrangements for dancing afford the opportunity to entrain to both visual and auditory cues in a multisensory manner. In addition, this capitalizes on the preference of visual entrainment for moving objects, rather than static objects (Hove et al., 2013). Hence, the model accounts for the real and present ability of humans to engage in visual entrainment (Wilson & Cook, 2016), a point that is generally overlooked in models of musical entrainment that focus on auditory cues alone. Visual entrainment is akin to a process of imitation, in particular visual mirroring, where the focus is on matching the timing of movements. This is the visual analogue of vocal pitch-matching.

6) Another highly overlooked sensory cue for entrainment is haptic contact. Many forms of multi-person dancing involve body contact, most typically through the upper limbs, affording yet a third cue for entrainment beyond visual and auditory cues. Elliott, Wing, and Welchman (2010) demonstrated that finger tapping to a tactile metronome was nearly as precise as tapping to an auditory metronome beat. Haptic contact is intimately associated with the fact that group dancing occurs in particular spatial configurations, most notably the geometric patterns of lines and circles. While such groupings can be unconnected, it is also quite common for people to establish physical contact through the hands or arms to create connected lines and closed circles. Haptic contact greatly amplifies the coupling strength between human bodies, providing an additional mechanism to support rhythmic entrainment through group dancing (Chauvigné, Walton, Richardson, & Brown, 2019). Lameira, Eerola, and Ravignani (2019), in the first demonstration of synchronous bipedal walking in a pair chimpanzees, revealed the importance of haptic contact between the animals in supporting this behavior. 7) Yet one more type of sensory signal that a group dancing model incorporates is vestibular signals, since the body movements of dancers can be strong stimuli for both the otolithic organs and semi-circular canals. Trainor, Gao, Lei, Lehtovaara, & Harris (2009) demonstrated that electrical stimulation of the vestibular system in the absence of body movement could lead to the perception of specific rhythms associated with the stimulation pattern.

8) A locomotor model provides an endogenous source of a continuous acoustic rhythm in a way that a vocal model does not. The opening of this article talked about how most models of audiomotor entrainment do not consider the origin of the acoustic rhythms that are entrained to. A locomotor model based on percussive interactions with the ground obviates this problem by providing a clear source of an acoustic rhythm that can be entrained to (Larsson, 2014; Larsson et al., 2019). 9) Along these same lines, duple rhythms are by far the dominant rhythms in world musics – including dance musics – (Brown & Jordania, 2013; Savage, Brown, Sakai, & Currie, 2015), and a locomotor model provides a clear source for the regular alternation between two isodurational beats. 10) The model is consistent with the ethnographic observation that dancers in indigenous cultures very often create body percussion not just from their movements per se but by attaching percussive objects to their body (see below), such as by wearing leggings, holding rattles in

the hands, or attaching objects to their regalia that make sounds when the dancer moves. To the extent that beads might be involved in these implements, it is clear that beads and beading are extremely ancient in human cultural evolution (Bednarik, 2015; Kuhn & Stiner, 2007). 11) The model also conforms with the ethnographic observation that people typically do not sing melodically while they dance, and that dancers are generally distinct from the musicians who generate melodies through singing or who play melodic instruments, although it is not uncommon for dancers in indigenous cultures to dance while playing percussion-based instruments, such as rattles or frame drums. This might suggest that the Voice First model is a better explanation for the distinction between musicians and dancers in human cultures. However, my problem with this model is that it requires a two-step evolution: first the ability to generate vocal rhythms, followed next by the ability to entrain body movements to them. The Body First model is more parsimonious in that the mechanism of mutual entrainment using body percussion allows for sound generation and motor entrainment to co-evolve. 12) To the extent that dance does have a connection with instrumental music, it is far more frequently connected with percussion music than with melodic music, most commonly drumming. A Body First model of entrainment based on body percussion argues that dancers were the first percussionists and probably the first percussion instruments as well. According to this view, a distinct class of percussion musicians later evolved by shifting the primordial body-percussion mechanisms – which included objects attached to the leg, held in the hand, or attached to one’s clothing, in addition to body-percussive mechanisms such as clapping and stomping – to individuals dedicated to playing percussion instruments in the absence of dancing.

Body percussion is highlighted here not just as an ancient feature of an evolutionary model but as a prominent contemporary feature of dance throughout the world, from indigenous cultures to large-scale cultures. Dancers attach objects to their bodies or to their regalia that allow them to make sounds as they move. This is seen in the leggings attached to the bodies of dancers in many cultures, the sequins and coins that are attached to the belts and bras of belly dancers, the shoe taps of Flamenco and tap dancers, and the hand-held rattles, shakers and frame drums of native North American dancers and well beyond. This is in addition to the sounds that dancers can generate with their bodies alone through stomping, clapping, and vocalizing.

To cite one example, traditional Aztec dancers in Mexico wear leggings (called *chachayotes*) containing seeds of the ayoyotl tree that make a loud clank with each and every step that the dancers take. Such dance performances are accompanied by people playing large barrel drums with mallets. A dancer wearing *chachayotes* around his/her legs generates an acoustic rhythm in a manner that can be indistinguishable from the rhythm generated by the person beating a drum with a mallet. The dancer’s entire body becomes a percussion instrument, and this leads to a true blurring of the distinction between dancers and musicians, as well as that between dance and music. There is a tendency to call the person who plays the drum a “musician” and the person who moves with percussion attached to their body a “dancer”, but in many cases the two are producing the identical acoustic rhythms. *Dance and music are simply two different manners of creating rhythmic patterns.* In the case of sonorant dances like Aztec dancing, they are in fact the same manner.

I can imagine the Voice First and Body First group models coming together to synergistically establish a sound source for mutual entrainment. Just as the Body First model takes advantage of body percussion, so too a Voice First model, while not requiring melodious pitch production, may work perfectly well with *vocal-tract percussion*. This may derive from the sounds of respiration or other sounds that can be generated by the vocal tract (Larsson, 2014; Larsson & Abbott, 2018), such as grunts, snorts and the percussive sounds that can be generated by the oral effectors. This could also include simple voiced sounds, such as tones, hums, or screams, as Jordania’s (2014) mobbing model suggests. The indigenous Maori people of New Zealand have a ceremonial war dance called *haka* that combines loud vocalizations (grunts, cries) with

percussive body movements (slapping, stomping), in addition to facial effects such as tongue protrusion (Youngerman, 1974). The function of such dances is to scare off enemies and to prevent or substitute for real, all-out fighting.

The potential synergy between vocal percussion and body percussion in the evolution of rhythmic entrainment might capitalize on the rhythmic coupling between locomotion and respiration in many mammals, birds, and fish (Stickford & Stickford, 2014). For example, many quadruped species take one stride per breath. Evidence for this locomotor/respiratory coupling (LRC) is controversial in humans, being mainly observed in trained runners, but much less so in non-runners. Hoffmann, Torregrosa, and Bardy (2012) demonstrated that human LRC can be stabilized when people on a stationary bicycle synchronize their cycling to the sound of a metronome tone at their preferred walking tempo, compared to no entrainment signal. This result shows that entraining to an external signal that is rhythmic is a means of stabilizing LRC in humans. One would expect that *vocalization* should have an even stronger impact on LRC than rhythm perception alone, since vocalization is directly linked to the process of respiration. Along these lines, Miyata and Kudo (2014) showed that vocalizing “ta” while performing flexion movements at several different joints in a standing posture led to a reduction in the variability of the movements, compared to performing the same body movements without vocalizing. While this study did not look at LRC per se, it did reveal a mutual stabilization between rhythmic vocalization and rhythmic body movement. With regards to the evolutionary model presented here, rhythmic vocalizing could have served not only as a sound source for entrainment – along with body percussion – but could have enhanced LRC, which itself may have provided additional visceral cues for entrainment. In addition, this could have increased the physiological efficiency of the activity by potentially reducing oxygen consumption during the task (Hoffmann et al., 2012).

While I cannot rule out a solo self-entrainment scenario using a vocal percussion mechanism – for example, entraining body movement to a rhythmically-produced vocal source as a mating display by individual males – this would in no way explain the universal capacity for *interpersonal* rhythmic entrainment in human cultures. Rhythmic entrainment is not only about creating synergies between the effectors of a single body, but about creating interpersonal synergies between interacting individuals (Riley, Richardson, Shockey, & Ramenzoni, 2011), due to the clear adaptive effect that this interpersonal coupling has on group cohesion and social cooperation (see references above). In addition, from an ethnographic standpoint, there are far more examples of dances in which the dancers do not vocalize than those in which they do, even in indigenous cultures. Body percussion is much more prevalent than vocal percussion among dancers, although a Huli bird dance from Papua New Guinea is one example where dancers sing simple repetitive melodies while dancing (Knauff, 1985). The same is true of the percussive vocal sounds made by Maori dancers during the *haka*. Vocalizing clearly does occur during dancing in certain cultures, but it is much less common than the use of body percussion or instrumental percussion.

2.3. Dance first or music first for the evolution of rhythm?

This discussion highlights the fact that the evolutionary models just presented offer opposite predictions about how rhythm came to appear in music as compared to dance. In the Voice First group model, the entrainment mechanism evolved in music first – as derived from the mutual entrainment of chorusing – and dance later co-opted this mechanism so as to permit either solo or group dancing to music. In the Body First group model, the entrainment mechanism evolved in dance first – as derived from the mutual entrainment of group dancing – and music later co-opted this entrainment mechanism so as to permit vocal chorusing and ensemble instrumental performance through mutual interaction. While the pitch dimension of music is clearly derived from

the voice, I believe that it is an open question as to whether the rhythm dimension is as well, not least since, as mentioned above, rhythm has no necessary connection with pitch and melody. The parameters that are associated with the concept of rhythm are just as plausibly accounted for by a body model as a voice model. As mentioned, the bipedal locomotor cycle offers a source of an endogenous duple-meter oscillator that has no counterpart in non-human primate vocalizations. Therefore, I think that it is far more reasonable to talk about the origins of rhythm in dance than in music, even though dance in contemporary times seems to be always “done to” music, hence implying an evolutionary prioritization of music over dance. I have argued here that body-percussion mechanisms obviate the problem of having to assume that a pre-established musical capacity had to be in place before dance was able to evolve in humans.

Fig. 3 presents an evolutionary model of the origins of rhythm that reflects the ideas presented in this paper. The precursor consists of mechanisms of mutual entrainment that are non-metric, of which there are abundant examples in animals (Bailey, 2003; Dahlin & Benedict, 2014; Geissmann, 2000a; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Haimoff, 1986; Hall, 2009; Ravnani et al., 2014; Wilson & Cook, 2016). Another possible source could be metric behaviors that are self-paced, such as locomotion. I propose that the primary adaptive change was the evolution of a rhythmic form of *mutual* entrainment, based on the locomotor Body First model employing multisensory acoustic/visual/haptic/vestibular mechanisms of interpersonal coupling during group dancing, perhaps supplemented by simple vocal percussion. This would establish a social mechanism for entraining with other people in a rhythmic manner. However, it needs to be pointed out that all existing evolutionary models of rhythmic entrainment attempt to explain not this phenomenon of mutual entrainment, but instead *external* metric entrainment, as in finger tapping to a metronome or dancing to recorded music, in which people synchronize aurally to an inflexible external signal. In the model presented in Fig. 3, I propose that external metric entrainment (e.g., dancing to music) evolved as a secondary offshoot of the principal adaptation for mutual metric entrainment, rather than the reverse.

Fig. 4 further elaborates on this process by specifying the evolutionary relationship between dance and music. The model argues that the system of mutual metric entrainment that emerged through the Body First group-dancing mechanism established a *joint origin of dance and rhythmic music*. If so, how does musical melody fit into this evolutionary scheme? I propose that music’s tonal system for creating vocal melodies using scaled pitches evolved independently of this rhythmic system and later *co-opted dance’s system of metric mutual entrainment*, uniting it with its chorusing mechanism so as to generate rhythmic forms of chorusing from what were presumably poorly-coordinated choruses at their origin (Brown, 2007, 2017). The rhythmic scaffold provided by the dance

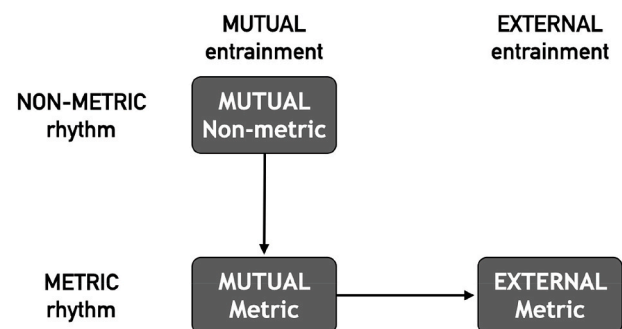


Fig. 3. Evolutionary model of entrainment. In this evolutionary scheme, the capacity for mutual entrainment through group dancing is the primary adaptation for coordination. The system for mutual entrainment in group dancing acts as a joint precursor for dancing and rhythmic forms of music. The capacity for external entrainment to timekeepers like musical beats evolves secondary to this.

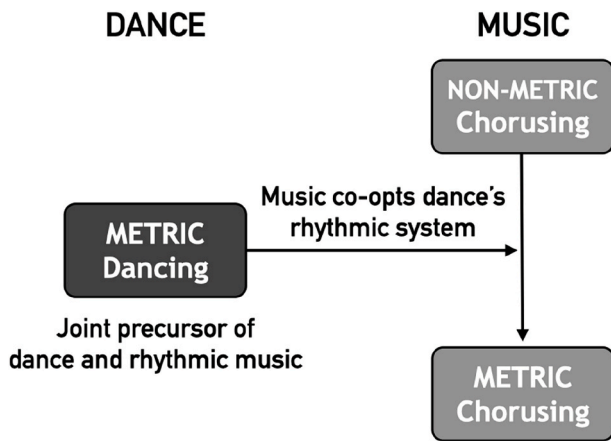


Fig. 4. Music co-opts dance's rhythmic system.

mechanism would now permit the insertion of vocal pitches – rather than body movements – into the time-slots created by the metrical structure. This idea might show similarities with Bücher's (1924) classic proposal that song and poetry derived their metricality from rhythmic forms of body movement occurring during communal labor, leading to the emergence of work songs accompanying such activities (described in Meyer-Kalkus, 2007). The model presented here contradicts the dominant perspective in the field that external entrainment is specifically about music (rather than dance), and argues instead that dance rhythm preceded musical rhythm in human evolution.

Overall, I propose a progression by which 1) rhythmic entrainment evolved in the multisensory and multi-effector context of group dancing

using body-percussion mechanisms via an evolutionary stage that was the joint precursor of dance and percussion music, 2) this was later supplemented by dedicated individuals playing percussion instruments but who did not dance, and 3) it was only later supplemented by musical ensembles made up of vocalists and melodic instruments, creating the conditions for external entrainment of dance movements to music, as is the general case in modern times. It is important to point out that this model is different in kind from "rhythm first" models of music, which argue that – within the specific context of music evolution – music's rhythmic features evolved before its tonal/melodic features (see Montagu, 2017; Seifert, 2018). I am instead arguing for a rhythm-first model of dance, not music, by positing that rhythmic entrainment evolved first in the context of dance and was later co-opted by music to permit integrated chorusing. Whether the stage of music evolution at which this co-opting occurred was pre-tonal or post-tonal is a separate issue for consideration. I would speculate that it occurred at the post-tonal stage – supporting a "pitch first" model from the perspective of music itself – in part because the world is full of chorusing types that are tonal/melodic but that are poorly coordinated (Jordania, 2006). However, I consider this to be an open question in need of further exploration.

The end result of this scenario is the parallel evolution of dance and music, with their reciprocal emphases on rhythm and melody, respectively (Fig. 5). Rhythm evolves through the "body" route and has a privileged connection with percussion instruments (including the body's own percussion), whereas melody evolves through the "vocal" route and has a privileged connection with melodic instruments, which serve as surrogates for the voice. In the evolutionary model presented here, I argue that music co-opts dance's rhythmic system to create the capacity for rhythmic chorusing. As a result, music evolves as a "dual coordination system" (Brown, 2017) in which interpersonal coordination comes about using the combination of the rhythmic properties of entrainment

	DANCE	MUSIC
Effector	Body	Voice
Modulated parameter	Rhythm	Pitch/melody
Space of coordination	Physical space	Tonal pitch space
Musical instruments	Percussion instruments	Melodic instruments
Representational system	Pantomimic gesture	Language/speech

Fig. 5. Parallel evolution of dance and music.

and the tonal properties of chorusing.

The figure also shows that dance and music have distinct connections with the mechanisms of narrative re-creation. In particular, dance's body route takes advantage of iconic gesturing through pantomime, whereas music's vocal route develops a deep connection with language and speech via songs with words. The end result of this constellation of features, as Radcliffe-Brown (1922) pointed out, is that dance allows people to move as if with one body, and music allows people to vocalize as if with one voice. Dance and music are the two arts of interpersonal coordination, taking advantage of the coordinative potential of the body and voice, respectively. The oppositions shown in Fig. 5 might be seen as being overly dichotomous. However, they highlight salient differences between dance and music as sensorimotor and communicative systems.

2.4. Did vocal imitation and rhythmic entrainment co-evolve?

Patel (2014) proposed the interesting hypothesis that the evolution of vocal learning (VL) and external metric entrainment – or what he calls beat-based entrainment (BBE) – are linked phylogenetically and mechanistically, hence co-evolving in a number of species, including humans. While there is some evidence to support this hypothesis, there are a number of reasons to be skeptical of this idea (Cook et al., 2013). As Wilson and Cook (2016) point out, “the ability to entrain the timing of a given behavior need not depend on a great deal of voluntary control over the content of that behavior” (p. 1653). Vocal learning and motor entrainment are highly distinct processes, most likely underlain by different neural systems. I will briefly discuss eight important differences between these two processes.

1) *Timing*. BBE is synchronous, while VL is sequential. BBE is about synchronizing with an ongoing beat, whereas VL is about reproducing a previously-heard sequence after it has appeared. In other words, these two phenomena occur in strikingly different time domains. This makes BBE far more oriented towards prediction, expectation, and adaptive timing than VL is (Keller et al., 2014; Large, 2001; Repp & Keller, 2008). 2) *Meter*. BBE is almost always based on metric rhythms, in accordance with the strong need for predictability in timing, whereas VL does not have this requirement and is often based on non-metric rhythms, as seen during speech learning. 3) *Pattern*. VL is about replicating a pattern, such as the pitches of a melodic sequence, whereas BBE has no requirement for pattern and simply requires a beat. That beat can be as pattern-less as a white noise pulse. As mentioned in the next point, it need not even be acoustic. 4) *Sensory modality*. VL is restricted to the auditory modality whereas BBE can occur with visual and haptic signals as well (Chauvigné et al., 2019; Elliott et al., 2010; Repp & Penel, 2002). 5) *Effector*. VL is restricted to the voice, whereas BBE can potentially engage any part of the body, the voice included. In fact, it often engages multiple parts of the body simultaneously, as seen in the synergies of dance movements. This difference is compounded by Point 1 above about timing differences such that VL is about reproducing an acoustic sequence with the voice (only) after the sequence has been presented, whereas BBE is about synchronizing multiple parts of the body with an ongoing beat that can be acoustic, visual and/or haptic. The former is exemplified by an impressionist mimicking a celebrity's manner of speaking, while the latter is exemplified by a person dancing a tango to an occurrent beat. Clearly, these are quite different processes.

6) *Neural control*. While the neuroscience of VL and BBE is still under investigation, it is worth pointing out that these two processes reflect trademark differences between two important subcortical motor systems. The basal ganglia are involved in sequence learning, while the cerebellum is involved in motor adaptation (Shmuelof & Krakauer, 2011). As Patel points out, there is abundant evidence in songbirds that the basal ganglia are involved in the process of song-sequence learning. They are also involved in vocal imitation in human neuroimaging studies (Belyk, Pfordresher, Liotti, & Brown, 2016). Likewise, there is equally strong evidence for the role of the cerebellum in adaptation of the motor system to ongoing sensory signals through its ability to reduce

prediction error (Kornysheva & Schubotz, 2011). Hence, VL and BBE nicely exemplify signature differences between the basal ganglia and cerebellum, respectively, in controlling sequentiality (VL) versus simultaneity (BBE). 7) *Neural pathways*. While there is good evidence that the arcuate fasciculus is central to VL in humans (Bernal & Ardila, 2009; Loui, Alsop, & Schlaug, 2009), there is currently no evidence implicating this system in BBE in humans. In addition, it is reasonable to think of the audiovocal pathway, of which the arcuate fasciculus is a part, as a specific brain network that expanded in humans to mediate vocal learning (Belyk & Brown, 2017; Rilling et al., 2008), analogous to the song network of birds. Such a system functions to take perceived acoustic information, such as the vocalizations of conspecifics, and generate a motor command to replicate the form of that acoustic signal with the voice. It is unlikely that this same system would mediate tapping one's foot to a beat. In other words, the neural system for vocal learning requires a highly specific audiovocal linkage for vocal imitation, whereas BBE implicates a far more general multi-sensory and multi-effector system. 8) *Behavioral dissociations*. There is mounting evidence for behavioral dissociations between deficits in vocal imitation and rhythm processing. Conditions like poor-pitch singing (Pfordresher & Brown, 2007) and congenital amusia (Dalla Bella, Giguère, & Peretz, 2009) demonstrate deficits in the ability to vocally imitate acoustic sequences without loss of the ability to process rhythm. Likewise, there is a phenomenon of “beat deafness” in which people have a deficit in entrainment to a beat but have relatively preserved pitch processing (Phillips-Silver et al., 2011). These dissociations are further evidence that the neural systems for vocal learning/imitation and entrainment are, at least in large part, distinct. While I would not deny that shared sensorimotor resources are engaged by VL and BBE, I think it best to consider these as distinct processes in both the time and pattern/tonal domains, as well as to realize that only humans have evolved the ability to unite these two processes in the species-specific form of melodic unison chorusing.

2.5. A role for mentalizing in motor entrainment

While rhythmic entrainment has only ever been considered as a *sensorimotor* phenomenon about oscillators coupling to one another in time, I believe that we need to include a cognitive component to entrainment as well, one that includes a prominent role for mentalizing in this process (Abe et al., 2019; Baimel, Birch, & Norenzayan, 2018; Baimel, Severson, Baron, & Birch, 2015). The mutual adjustments that underlie entrainment with another person, such as those occurring during a couple dance (Kimmel, 2012, 2019), require that we actively monitor our partner's intentions and emotions, as well as our ongoing relationship with them in order to predict how they will react to our actions. Entraining with another person is very much a social behavior, and so it involves both the positive and negative emotional appraisals associated with any form of social interaction.

The role of mentalizing in entrainment has been underappreciated compared to sensorimotor mechanisms, such as the adjustment of micro-timing. Because mentalizing is a pivotal concern in the evolution of human cooperative behavior (Baron-Cohen, 1999; Hare, 2011; Kellogg & Evans, 2019; MacLean, 2016), it needs to be incorporated into the study of rhythmic entrainment, which is a social behavior just like any other. In a functional MRI study of joint action, Chauvigné, Belyk, and Brown (2018) examined physical entrainment between a participant and an experimenter while they performed rhythmic bimanual movements at their own pace. The results showed that rhythmic entrainment of the partners as co-equals recruited components of the mentalizing network of the brain, such as the temporoparietal junction and posterior cingulate cortex, suggesting that synchronizing with a person in time involves an ongoing process of predictive mentalizing about their intentions and reactions, as well as a coupling between the mentalizing and motor systems.

3. Proto-dancing: body convergence during conversation as a potential fossil

The current model of rhythmic entrainment through group dancing introduces a significant visual component that models of audiomotor entrainment alone do not consider. I would like to suggest that a modern-day fossil of the capacity for visuomotor entrainment might be found in the universal proclivity of interlocutors to converge in their body movements and posturing during conversation. This phenomenon is one part of a multimodal suite of coordinative behaviors that occur during conversation, which includes a convergence of facial expression, speech prosody, phonology, word selection, and syntactic construction (Duran & Fusaroli, 2017; Gaziv, Noy, Liron, & Alon, 2017; Manson, Bryant, Gervais, & Kline, 2013; Pickering & Garrod, 2004; Waciewicz, Żywicznyński, & Chiera, 2017). However, I will focus on gestural convergence since it is the closest thing to a dance that occurs in conversation.

Body convergence is a form of mirroring – and thus gestural imitation – that occurs reliably during affiliative social interactions. For example, Gaziv et al. (2017) provided evidence for synchronized side-to-side sway of the torso when familiar dyads engaged in face-to-face conversation while standing. A second major type of coordination was mutual stillness, in which the interlocutors synchronized by being still together. While I do not consider gestural convergence to be a dance per se, it might be a fossil of the underlying mirroring capacity that gave rise to the human capacity for rhythmic entrainment in the Body First model. To the extent that gestural mirroring is indeed important for the origins of rhythmic entrainment according to a dance-based model, then this process invokes functional systems like the mirror neuron system that have been implicated in the gestural origins of linguistic communication (Arbib, 2012), in which complex imitation played a key role.

Body convergence during conversation differs from group dancing in two important respects. First, it is done unconsciously, rather than being an explicit attempt by the individuals to entrain to one another. Second, its rhythm is non-metric, rather than having a regular metrical structure. Therefore, the transition from this putative precursor to full-fledged rhythmic entrainment would require two major innovations. First, it would require a transition from unconscious to conscious coupling and hence to the explicit and mutual goal of entraining to another person through mechanisms of joint action and shared agency. Second, it would require a transition from non-metric (and thus poorly coordinated) timing to metric (and thus well-coordinated) movement patterns of the participants. I would speculate that, akin to the conversational dyads in Gaziv et al.'s (2017) study, a simple way to initiate a transition to metricality would be for people to lock-step their side-to-side locomotor movements while standing in a face-to-face configuration. The next section mentions a report of lock-step dyadic locomotion by a pair of chimpanzees (Lameira et al., 2019) – although in a conga-line configuration – thereby lending plausibility to this idea. Evolutionarily, this might also be facilitated by a change in the social arrangement of entrainment from dyads to larger groupings that could engender a greater degree of coupling and rhythmicity than dyads alone, leading ultimately to the well-coordinated group dances of modern-day cultures, both small and large. Flash mobs in which 20,000 or more people move in synchrony with one another is a contemporary manifestation of an ancient, but newly-evolved, capacity of humans.

3.1. Dance-like behaviors in chimpanzees

As a final phylogenetic point about the origins of entrainment, let us consider the evidence for dance-like behaviors in non-human primates, and examine the evidence for potential evolutionary continuity of entrainment skills. I am not aware of any ethological evidence in non-human primates for the intentional generation of regular acoustic rhythms in the wild. Given this absence of a rhythmic sound-source for

entrainment, the question then becomes the standard one of “dancing animal” models regarding whether non-human primates have a latent ability to entrain body movements to human-generated acoustic stimuli. Hattori, Tomonaga, and Matsuzawa (2013) reported a case study of a female chimpanzee who engaged in spontaneous tapping to an auditory rhythm on an electronic keyboard after having been trained to perform this task in the lab without auditory cuing. Closer to dance, Hattori and Tomonaga (2020) demonstrated spontaneous whole-body movements by chimpanzees to recordings of piano sounds, involving body movement, head movement, and hand clapping. Entrainment to the beat of the music was most apparent in bipedal posture, and was observed more in males than females. Demonstrations such as these of latent audiomotor abilities in primates, just as in some birds and non-primate mammals, is important evolutionarily, but still begs the question of where the acoustic rhythms come from in the wild, a conundrum that I raised at the opening of this article. In addition, studies that present music through recordings are not able to address the social, interpersonal roots of entrainment.

A more promising primate approach is found in a case study reported in Lameira et al. (2019), which described synchronized, yet self-paced, bipedal walking by an affiliative pair of captive female chimpanzees. This behavior took the form of a “conga line” in which one animal was walking in front of the other, and in which rhythmic walking movements of the two animals occurred in lock step, just as in human military marching (McNeill, 1995). Online videos of this behavior reveal a striking finding: it involves a clear *haptic* coupling between the two animals. During synchronous walking, the follower animal has her left hand on the back of the front animal, hence creating a haptic channel for entrainment. In principle, the back animal can also employ visual cues for entraining to the front animal. This in-line walking arrangement also creates a foundation for thinking about the distinction between leading and following in many forms of joint action in human behavior (Fairhurst, Janata, & Keller, 2014; Sacheli, Tidoni, Pavone, Aglioti, & Candidi, 2013; Sebanz et al., 2006). This is reinforced by the fact that one animal was always the front animal across different bouts of the behavior (Lameira et al., 2019), although a Grainger causality analysis of relative timing revealed a bidirectional causal relation between the two animals during half of the bouts, whereas the back animal actually led in about a quarter of the bouts.

This synchronous, in-phase walking behavior accords perfectly with the perspective of the Locomotor Body First model presented here. In particular, it demonstrates that the achievement of synchrony in the body movements of two agents has *absolutely no requirement for an external auditory signal to entrain to*. Entrainment occurs in a self-paced manner through mutual entrainment alone, and relies on haptic and visual cues internal to the dyad. Whether locomotor and/or respiratory sounds contribute to the synchrony of this chimpanzee dyad is not known. What the study of Lameira et al. (2019) suggests is that haptic coupling may have been the earliest means for producing rhythmic entrainment between two agents in non-human primates, something that a music model is incapable of explaining. In addition, the fact that this joint behavior is restricted to the bipedal manner of locomotion in chimpanzees might have implications for the evolution of both bipedalism and dance. This behavior is consistent with a gradualist hypothesis of the origins of acoustically-driven motor entrainment (Merchant & Honing, 2014; Podlipniak, 2021), rather than one in which entrainment evolved as a single event (Mithen, 2005).

Online videos of this chimpanzee dyad reveal that the pair engages in a great deal of body swaying when they sit down next to each other, again fitting in with the “proto-dance” hypothesis described here. When the animals are sitting, their swaying is not synchronous (i.e., it is not in phase), but neither is there any haptic coupling between them during these times, again suggesting that haptic coupling may be the primary cue for producing whole-body entrainment in chimpanzees. What is still missing, from an evolutionary standpoint, is how auditory cues come into play here. The fundamental tenet of the group dancing model

presented in this article is that the source of such sound generation should be *internal* to the entraining agents, rather than external to them (as posited in “dancing animal” models). I contend that the diverse forms of body percussion that are generated during joint movement could provide a key source for such auditory cues.

4. Conclusions

Most evolutionary models of the human capacity for rhythmic entrainment place their focus on audiomotor entrainment without giving any consideration for where the rhythmic acoustic stimulus came from to begin with, whether in humans or other animals. I have presented a model that unifies the sound-source and effector for entrainment via interpersonal coupling through group dancing with body percussion. The two key features of the model are that 1) body percussion served as the primary acoustic sound-source in the evolution of motor entrainment, and 2) rhythmic entrainment emerged interpersonally as an act of mutual entrainment through group dancing, rather than as an individual-level act of external entrainment to music. The Body First model proposed here does not require a pre-existing capacity to produce rhythmic music, but instead capitalizes on the acoustic properties of both body and vocal percussion. Such interpersonal coupling of movement in both time and space would promote affiliation and cooperation not just in dyads but in larger social groups. I argued that rhythm in music was inherited from an entrainment capacity that initially evolved in the context of group dancing. A potential fossil of the precursor-capacity for entrainment might be found in the modern-day convergence of body sway and posturing that occurs between conversationalists and even between affiliative chimpanzee dyads. This convergence, while unconscious and non-metric, might have provided the underlying visuo-motor mirroring capacity necessary for body-based entrainment in time and space. Overall, I have argued that a model of the evolution of rhythmic entrainment grounded in group dancing provides far more explanatory richness than models based on finger tapping or animal entrainment to human music.

Author statement

I am the sole author of this work. I do not have any financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

Acknowledgments

This work was supported by a grant from the Social Sciences and Humanities Research Council (SSHRC) of Canada (grant number 435-2017-0491). I thank Matz Larsson and Joseph Jordania for critical reading of the manuscript and for suggestions for improvement. I thank the three anonymous reviewers for their constructive and insightful comments on an earlier version of the manuscript and for directing me to literature that I was not aware of.

References

- Abe, M. O., Koike, T., Okazaki, S., Sugawara, S. K., Takahashi, K., Watanabe, K., et al. (2019). Neural correlates of online cooperation during joint force production. *NeuroImage*, *191*, 150–161. <https://doi.org/10.1016/j.neuroimage.2019.02.003>
- Anshel, A., & Kipper, D. A. (1988). The influence of group singing on trust and cooperation. *Journal of Music Therapy*, *25*, 145–155.
- Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: Oxford University Press.
- Arcadi, A. C., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, *39*(4), 505–518. <https://doi.org/10.1007/BF02557572>
- Bailey, W. J. (2003). Insect duets: Underlying mechanisms and their evolution. *Physiological Entomology*, *28*, 157–174. <https://doi.org/10.1046/j.1365-3032.2003.00337.x>
- Baimel, A., Birch, S. A. J., & Norenzayan, A. (2018). Coordinating bodies and minds: Behavioral synchrony fosters mentalizing. *Journal of Experimental Social Psychology*, *74*, 281–290. <https://doi.org/10.1016/j.jesp.2017.10.008>
- Baimel, A., Severson, R. L., Baron, A. S., & Birch, S. A. J. (2015). Enhancing “theory of mind” through behavioral synchrony. *Frontiers in Psychology*, *6*, 670. <https://doi.org/10.3389/fpsyg.2015.00870>
- Baron-Cohen, S. (1999). Evolution of a theory of mind? In M. C. Corballis, & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 261–277). Oxford: Oxford University Press.
- Bednarik, R. G. (2015). The significance of the earliest beads. *Advances in Anthropology*, *5*, 51–66. <https://doi.org/10.4236/aa.2015.52006>
- Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience & Biobehavioral Reviews*, *77*, 177–193. <https://doi.org/10.1016/j.neubiorev.2017.03.014>
- Belyk, M., Pfordresher, P. Q., Liotti, M., & Brown, S. (2016). The neural basis of vocal pitch imitation in humans. *Journal of Cognitive Neuroscience*, *28*, 621–635.
- Bernal, B., & Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain*, *132*, 2309–2316. <https://doi.org/10.1093/brain/awp206>
- Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. In F. Tonneau, & N. S. Thompson (Eds.), *Perspectives in ethology. 13: Behavior, evolution and culture* (pp. 231–281). New York: Plenum Publishers.
- Brown, S. (2007). Contagious heterophony: A new theory about the origins of music. *Musicae Scientiae*, *11*, 3–26. <https://doi.org/10.1177/102986490701100101>
- Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, *8*, 1894. <https://doi.org/10.3389/fpsyg.2017.01894>
- Browner, T. (2000). Making and singing pow-wow songs: Text, form, and the significance of culture-based analysis. *Ethnomusicology*, *44*(2), 214. <https://doi.org/10.2307/852530>
- Brown, S., & Jordania, J. (2013). Universals in the world’s musics. *Psychology of Music*, *41*(2), 229–248. <https://doi.org/10.1177/0305735611425896>
- Bücher, K. (1924). *Arbeit und Rhythmus* (6th ed.). Leipzig: Emmanuel Reinicke.
- Chauvigné, L. A. S., Belyk, M., & Brown, S. (2018). Taking two to tango: fMRI analysis of improvised joint action with physical contact. *PLoS One*, *13*, e0191098. <https://doi.org/10.1371/journal.pone.0191098>
- Chauvigné, L. A. S., Walton, A., Richardson, M. J., & Brown, S. (2019). Multi-person and multisensory synchronization during group dancing. *Human Movement Science*, *63*, 199–208. <https://doi.org/10.1016/j.humov.2018.12.005>
- Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*, 1003–1011. <https://doi.org/10.1111/desc.12193>
- Clayton, M. R. L. (1996). Free rhythm: Ethnomusicology and the study of music without metre. *Bulletin of the School of Oriental and African Studies*, *59*(2), 323–332.
- Clayton, M. (2012). What is entrainment? Definition and applications in musical research. *Empirical Musicology Review*, *7*(1–2), 49–56. <https://doi.org/10.18061/1811/52979>
- 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*, 412–427. <https://doi.org/10.1037/a0032345>
- Cross, I., & Morley, I. (2009). The evolution of music: Theories, definitions and the nature of the evidence. In S. Malloch, & C. Trevarthen (Eds.), *Communicative musicality* (pp. 61–81). Oxford: Oxford University Press.
- Cross, I., Turgeon, M., & Atherton, G. (2019). How moving together binds us together: The social consequences of interpersonal entrainment and group processes. *Open Psychology*, *1*, 273–302. <https://doi.org/10.1515/psych-2018-0018>
- Dahlin, C. R., & Benedict, L. (2014). Angry birds need not apply: A perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology*, *120*, 1–10. <https://doi.org/10.1111/eth.12182>
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, *126*, 414–424.
- Dunbar, R. I. M., Kaskatis, K., MacDonald, I., & Barra, V. (2012). Performance of music elevates pain threshold and positive affect: Implications for the evolutionary function of music. *Evolutionary Psychology*, *10*(4), 688–702. <https://doi.org/10.1177/147470491201000403>
- Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil’s advocate: Interpersonal coordination in deception and disagreement. *PLoS One*, *12*(6), e0178140. <https://doi.org/10.1371/journal.pone.0178140>
- Elliott, M. T., Wing, A. M., & Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *European Journal of Neuroscience*, *31*, 1828–1835. <https://doi.org/10.1111/j.1460-9568.2010.07205.x>
- Fairhurst, M. T., Janata, P., & Keller, P. E. (2014). Leading the follower: An fMRI investigation of dynamic cooperativity and leader-follower strategies in synchronization with an adaptive virtual partner. *NeuroImage*, *84*, 688–697. <https://doi.org/10.1016/j.neuroimage.2013.09.027>
- Fischer, R., Callander, R., Reddish, P., & Bulbulia, J. (2013). How do rituals affect cooperation? An experimental field study comparing nine ritual types. *Human Nature*, *24*, 115–125. <https://doi.org/10.1007/s12110-013-9167-y>
- Gaziv, G., Noy, L., Liron, Y., & Alon, U. (2017). A reduced-dimensionality approach to uncovering dyadic modes of body motion in conversations. *PLoS One*, *12*(1), e0170786. <https://doi.org/10.1371/journal.pone.0170786>
- Geissmann, T. (2000a). Duet songs of the siamang, *Hylobates syndactylus*: I. Structure and organisation. *Primate Report*, *56*, 33–60.
- Geissmann, T. (2000b). Gibbon songs and human music from an evolutionary perspective. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 103–123). Cambridge, MA: MIT Press.

- Good, A., & Russo, F. A. (2016). Singing promotes cooperation in a diverse group of children. *Social Psychology*, 47, 340–344. <https://doi.org/10.1027/1864-9335/a000282>
- Grafé, T. U. (1999). A function of synchronous chorusing and a novel female preference shift in an anuran. *Proceedings of the Royal Society B: Biological Sciences*, 266, 2331–2336. <https://doi.org/10.1098/rspb.1999.0927>
- Greenfield, M. D. (2015). Signal interactions and interference in insect choruses: Singing and listening in the social environment. *Journal of Comparative Physiology: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 201, 143–154. <https://doi.org/10.1007/s00359-014-0938-7>
- Greenfield, M. D., & Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature*, 364, 618–620. <https://doi.org/10.1038/364618a0>
- Greenfield, M. D., & Schul, J. (2008). Mechanisms and evolution of synchronous chorusing: Emergent properties and adaptive functions in *Neconocephalus katydid* (Orthoptera: Tettigoniidae). *Journal of Comparative Psychology*, 122, 289–297. <https://doi.org/10.1037/0735-7036.122.3.289>
- Grillner, S. (2006). Biological pattern generation: The cellular and computational logic of networks in motion. *Neuron*, 52, 751–766. <https://doi.org/10.1016/j.neuron.2006.11.008>
- Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, 14, 21–51.
- Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the loud calls of primates, lions, hyenas, and wolves. *Musicae Scientiae*, 13, 291–320. <https://doi.org/10.1177/1029864909013002131>
- Haimoff, E. H. (1986). Convergence in the duetting of monogamous Old World primates. *Journal of Human Evolution*, 15, 51–59. [https://doi.org/10.1016/S0047-2484\(86\)80065-3](https://doi.org/10.1016/S0047-2484(86)80065-3)
- Hall, M. L. (2009). A review of vocal duetting in birds. *Advances in the Study of Behavior*, 40, 67–121. [https://doi.org/10.1016/S0065-3454\(09\)40003-2](https://doi.org/10.1016/S0065-3454(09)40003-2)
- Hanna, J. L. (1979). *To dance is human: A theory of nonverbal communication*. Chicago: University of Chicago Press.
- Hare, B. (2011). From hominoid to hominid mind: What changed and why? *Annual Review of Anthropology*, 40, 293–309. <https://doi.org/10.1146/annurev-anthro-081309-145726>
- Hattori, Y., & Tomonaga, M. (2020). Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proceedings of the National Academy of Sciences*, 117(2), 936–942. <https://doi.org/10.1073/pnas.1910318116>
- Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, 3, 1566. <https://doi.org/10.1038/srep01566>
- Hoffmann, C. P., Torregrosa, G., & Bardy, B. G. (2012). Sound stabilizes locomotor-respiratory coupling and reduces energy cost. *PLoS One*, 7, e45206. <https://doi.org/10.1371/journal.pone.0045206>
- Honing, H. (2019). *The evolving animal orchestra: In search of what makes us musical*. Cambridge, MA: MIT Press.
- von Hornbostel, E. M. (1904). *Melodischer Tanz: Eine musikpsychologische studie*. Reprinted with English translation as *Melodic dance: A musico-psychological study*. In K. P. Wachsmann, D. Christensen, & H.-P. Reinecke (Eds.), *Hornbostel opera omnia* (pp. 204–215). The Hague: Martinus Nijhoff.
- Hove, M. J., Fairhurst, M. T., Kotz, S. A., & Keller, P. E. (2013). Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *NeuroImage*, 67, 313–321. <https://doi.org/10.1016/j.neuroimage.2012.11.032>
- Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27, 949–961. <https://doi.org/10.1521/soco.2009.27.6.949>
- Jola, C., Ehrenberg, S., & Reynolds, D. (2012). The experience of watching dance: Phenomenological-neuroscience duets. *Phenomenology and the Cognitive Sciences*, 11, 17–37. <https://doi.org/10.1007/s11097-010-9191-x>
- Jones, M. R. (2019). *Time will tell: A theory of dynamic attending*. Oxford: Oxford University Press.
- Jordania, J. (2006). *Who asked the first question? The origins of human choral singing, intelligence, language and speech*. Tbilisi: Logos.
- Jordania, J. (2014). *Tigers, lions and humans: History of rivalry, conflict, reverence and love*. Tbilisi: Logos.
- Katz, P. S. (2016). Evolution of central pattern generators and rhythmic behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150057. <https://doi.org/10.1098/rstb.2015.0057>
- Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130394. <https://doi.org/10.1098/rstb.2013.0394>
- Kellogg, R. T., & Evans, L. (2019). The ensemble hypothesis of human cognitive evolution. *Evolutionary Psychological Science*, 5, 1–12. <https://doi.org/10.1007/s40806-018-0159-3>
- Kimmel, M. (2012). Intersubjectivity at close quarters: How dancers of Tango Argentino use imagery for interaction and improvisation. *Cognitive Semiotics*, 4, 75–137. https://doi.org/10.1007/978-94-017-9404-6_47
- Kimmel, M. (2019). A cognitive theory of joint improvisation: The case of Tango Argentino. In V. Midgelow (Ed.), *The Oxford handbook of improvisation in dance* (pp. 563–591). Oxford: Oxford University Press.
- Kim, J. H., Reiferger, A., & Rizzonelli, M. (2019). Musical social entrainment. *Music & Science*, 2, 1–17. <https://doi.org/10.1177/2059204319848991>
- Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31, 354–364. <https://doi.org/10.1016/j.evolhumbehav.2010.04.004>
- Knauff, B. M. (1985). Ritual form and permutation in New Guinea: Implications of symbolic process for socio-political evolution. *American Ethnologist*, 12, 321–340. <https://doi.org/10.1525/ae.1985.12.2.02a00080>
- Kniffin, K. M., Yan, J., Wansink, B., & Schulze, W. D. (2017). The sound of cooperation: Musical influences on cooperative behavior. *Journal of Organizational Behavior*, 38, 372–390. <https://doi.org/10.1002/job.2128>
- Kornysheva, K., & Schubotz, R. I. (2011). Impairment of auditory-motor timing and compensatory reorganization after ventral premotor cortex stimulation. *PLoS One*, 6, e21421. <https://doi.org/10.1371/journal.pone.0021421>
- Kreiter, H., & Kreiter, S. (1972). *Psychology of the arts*. Durham, NC: Duke University Press.
- Kuhn, S. L., & Stiner, M. C. (2007). Body ornamentation as information technology: Towards an understanding of the significance of early beads. In P. Mellars, K. Boyle, O. Bar-Yosef, & C. Stringer (Eds.), *Rethinking the human revolution: New behavioural and biological perspectives on the origin and dispersal of modern humans* (pp. 45–54). Cambridge, UK: McDonald Institute for Archaeological Research.
- Lameira, A. R., Eerola, T., & Ravignani, A. (2019). Coupled whole-body rhythmic entrainment between two chimpanzees. *Scientific Reports*, 9, 18914. <https://doi.org/10.1038/s41598-019-55360-y>
- Large, E. W. (2001). Periodicity, pattern formation, and metric structure. *Journal of New Music Research*, 30, 173–185. <https://doi.org/10.1076/jnmr.30.2.173.7113>
- Larsson, M. (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Animal Cognition*, 17, 1–14. <https://doi.org/10.1007/s10071-013-0678-z>
- Larsson, M., & Abbott, B. W. (2018). Is the capacity for vocal learning in vertebrates rooted in fish schooling behavior? *Evolutionary Biology*, 45, 359–373. <https://doi.org/10.1007/s11692-018-9457-8>
- Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal steps in the development of rhythmic behavior in humans. *Music & Science*, 2, 1–14. <https://doi.org/10.1177/2059204319892617>
- Latash, M. L., Scholz, J. P., & Schönner, G. (2007). Toward a new theory of motor synergies. *Motor Control*, 11, 276–308. <https://doi.org/10.1123/mcj.11.3.276>
- Launay, J., Tarr, B., & Dunbar, R. I. M. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology*, 122, 779–789. <https://doi.org/10.1111/eth.12528>
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, 29, 10215–10220. <https://doi.org/10.1523/JNEUROSCI.1701-09.2009>
- MacLean, E. L. (2016). Unraveling the evolution of uniquely human cognition. *Proceedings of the National Academy of Sciences*, 113, 6348–6354. <https://doi.org/10.1073/pnas.1521270113>
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499–546.
- MacNeilage, P. F. (2008). *The origin of speech*. Oxford: Oxford University Press.
- Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6), 419–426. <https://doi.org/10.1016/j.evolhumbehav.2013.08.001>
- McNeill, W. H. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.
- Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*.
- Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7, 274. <https://doi.org/10.3389/fnins.2013.00274>
- Meyer-Kalkus, R. (2007). Work, rhythm, dance: Prerequisites for a kinaesthetics of media and arts. In J. M. Krois, M. Rosengren, A. Steidele, & D. Westerkamp (Eds.), *Embodiment in cognition and culture* (pp. 165–181). Amsterdam: John Benjamins Publishing Company.
- Meyer, C., Streeck, J., & Jordan, J. S. (2017). Introduction. In C. Meyer, J. Streeck, & J. S. Jordan (Eds.), *Intercorporeality: Emerging socialities in interaction* (pp. xv–xlix). Oxford: Oxford University Press.
- Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind and body*. London: Weidenfeld & Nicolson.
- Miyata, K., & Kudo, K. (2014). Mutual stabilization of rhythmic vocalization and whole-body movement. *PLoS One*, 9, e115495. <https://doi.org/10.1371/journal.pone.01115495>
- Miyata, K., Varlet, M., Miura, A., Kudo, K., & Keller, P. E. (2017). Modulation of individual auditory-motor coordination dynamics through interpersonal visual coupling. *Scientific Reports*, 7, 16220. <https://doi.org/10.1038/s41598-017-16151-5>
- Montagu, J. (2017). How music and instruments began: A brief overview of the origin and entire development of music, from its earliest stages. *Frontiers in Sociology*, 2, 8. <https://doi.org/10.3389/fsoc.2017.00008>
- 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. <https://doi.org/10.4436/JASS.92008>
- Pacherie, E. (2011). The phenomenology of joint action: Self-agency versus joint-agency. In A. Seemann (Ed.), *Joint attention: New developments* (pp. 343–389). Cambridge, MA: MIT Press.
- Patel, A. D. (2014). The evolutionary biology of musical rhythm: Was Darwin wrong? *PLoS Biology*, 12, e1001821. <https://doi.org/10.1371/journal.pbio.1001821>
- 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, 827–830. <https://doi.org/10.1016/j.cub.2009.03.038>

- Pearce, E., Launay, J., & Dunbar, R. I. M. (2015). The ice-breaker effect: Singing mediates fast social bonding. *Royal Society Open Science*, 2, 150221. <https://doi.org/10.1098/rsos.150221>
- Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception*, 25, 95–115.
- Phillips-Silver, J., Toivianen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., et al. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, 49, 961–969. <https://doi.org/10.1016/j.neuropsychologia.2011.02.002>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190. <https://doi.org/10.1017/S0140525X04000056>
- Podlipniak, P. (2021). The role of canalization and plasticity in the evolution of musical creativity. *Frontiers in Neuroscience*, 15, 607887. <https://doi.org/10.3389/fnins.2021.607887>
- Radcliffe-Brown, A. (1922). *The andaman islanders: A study in social anthropology*. London: Cambridge University Press.
- Ravignani, A., Bowling, D., & Fitch, W. T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. *Frontiers in Psychology*, 5, 118. <https://doi.org/10.3389/fpsyg.2014.01118>
- Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's dance together: Synchrony, shared intentionality and cooperation. *PLoS One*, 8, e71182. <https://doi.org/10.1371/journal.pone.0071182>
- Renning, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony: A meta-analysis. *Zeitschrift Für Psychologie*, 224, 168–189. <https://doi.org/10.1027/2151-2604/a000252>
- Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, 27, 423–456. <https://doi.org/10.1016/j.humov.2008.02.016>
- Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1085–1099.
- Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2, 38. <https://doi.org/10.3389/fpsyg.2011.00038>
- Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, 11, 426–428. <https://doi.org/10.1038/nn2072>
- Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research*, 226(4), 473–486. <https://doi.org/10.1007/s00221-013-3459-7>
- Sachs, C. (1937). *World history of the dance*. New York: W.W. Norton & Company.
- Sachs, C. (1943). *The rise of music in the ancient world: East and West*. New York: W.W. Norton & Company.
- 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., et al. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*.
- 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, 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Seifert, U. (2018). Relating the evolution of music-readiness and language-readiness within the context of comparative neuroprimatology. *Interaction Studies*, 19(1–2), 86–101. <https://doi.org/10.1075/is.17035.sei>
- Sheridan, L., Brookes, M., & Pomiankowski, A. (1993). Jamming for sex. *Current Biology*, 3, 3–4.
- Shmuelof, L., & Krakauer, J. W. (2011). Are we ready for a natural history of motor learning? *Neuron*, 72, 469–476. <https://doi.org/10.1016/j.neuron.2011.10.017>
- Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *Behavioral and Brain Sciences*, 37, 243–295. <https://doi.org/10.1017/S0140525X13001544>
- Stickford, A. S. L., & Stickford, J. L. (2014). Ventilation and locomotion in humans: Mechanisms, implications, and perturbations to the coupling of these two rhythms. *Springer Science Reviews*, 2, 95–118. <https://doi.org/10.1007/s40362-014-0020-4>
- Strogatz, S. (2003). *Sync: How order emerges from chaos in the universe, nature, and daily life*. New York: Hyperion.
- Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., & Nakagawa, S. (2016). Territoriality, social bonds, and the evolution of communal signaling in birds. *Frontiers in Ecology and Evolution*, 4, 74. <https://doi.org/10.3389/fevo.2016.00074>
- Trainor, L. J., Gao, X., Lei, J., Lehtovaara, K., & Harris, L. R. (2009). The primal role of the vestibular system in determining musical rhythm. *Cortex*, 45, 35–43. <https://doi.org/10.1016/j.cortex.2007.10.014>
- Waciewicz, S., Żywicznyński, P., & Chiera, A. (2017). An evolutionary approach to low-level conversational cooperation. *Language Sciences*, 63, 91–104. <https://doi.org/10.1016/j.langsci.2017.01.005>
- 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>
- Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin & Review*, 23, 1647–1659. <https://doi.org/10.3758/s13423-016-1013-x>
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, 20, 1–5. <https://doi.org/10.1111/j.1467-9280.2008.02253.x>
- Youngerman, S. (1974). Maori dancing since the eighteenth century. *Ethnomusicology*, 18(1), 75–100.
- von Zimmermann, J., Vicary, S., Sperling, M., Orgs, G., & Richardson, D. C. (2018). The choreography of group affiliation. *Topics in Cognitive Science*, 10, 80–94. <https://doi.org/10.1111/tops.12320>