



Searching for roots of entrainment and joint action in early musical interactions

Jessica Phillips-Silver^{1*} and Peter E. Keller^{2,3*}

¹ International Laboratory for Brain, Music and Sound, Montreal, Quebec, Canada

² Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

³ MARCS Institute, University of Western Sydney, Australia

Edited by:

Robert J. Zatorre, McGill University, Canada

Reviewed by:

Robert J. Zatorre, McGill University, Canada

Edward W. Large, Florida Atlantic University, USA

Erin Hannon, University of Nevada Las Vegas, USA

*Correspondence:

Jessica Phillips-Silver, International Laboratory for Brain, Music and Sound, Pavillon 1430 Boulevard Mont-Royal, H2V 4P3, QC, Montréal, Canada.

e-mail: jessica.phillips-silver@umontreal.ca;

Peter E. Keller, Max Planck Research Group "Music Cognition & Action," Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

e-mail: keller@cbs.mpg.de

When people play music and dance together, they engage in forms of musical joint action that are often characterized by a shared sense of rhythmic timing and affective state (i.e., temporal and affective entrainment). In order to understand the origins of musical joint action, we propose a model in which entrainment is linked to dual mechanisms (motor resonance and action simulation), which in turn support musical behavior (imitation and complementary joint action). To illustrate this model, we consider two generic forms of joint musical behavior: chorusing and turn-taking. We explore how these common behaviors can be founded on entrainment capacities established early in human development, specifically during musical interactions between infants and their caregivers. If the roots of entrainment are found in early musical interactions which are practiced from childhood into adulthood, then we propose that the rehearsal of advanced musical ensemble skills can be considered to be a refined, mimetic form of temporal and affective entrainment whose evolution begins in infancy.

Keywords: music, joint action, entrainment, ensemble skills, development, dance

INTRODUCTION

What joint human behavior reveals greater coordination in time and affect – or entrainment – than music? When individuals join together in almost any musical behavior, ranging from a simple tune shared between a young child and caregiver, to a rhythmically complex performance of a Cuban Jazz band (**Figure 1**), their joint action is characterized by entrainment. The term entrainment (from the Middle French *entraîner*, v., to draw, drag) is used across diverse domains to refer generally to spatiotemporal coordination between two or more individuals, often in response to a rhythmic signal (e.g., Phillips-Silver et al., 2010). Used in musical contexts (including dance), entrainment typically refers to the rhythmic synchronization as a product of individuals' interaction (Clayton et al., 2005) – that is, rhythmic coordination with simple (e.g., 1:1 in-phase or anti-phase) or more complex (e.g., 2:3 or 3:4 polyrhythmic) phase relations.

One component of entrainment is thus temporal. Temporal entrainment to music can be observed at hierarchical levels of metrical periodicity in the body (Leman and Naveda, 2010; Naveda and Leman, 2010; Toiviainen et al., 2010) and brain (Large and Jones, 1999; Large and Palmer, 2002; London, 2004; Jones, 2009; Nozaradan et al., 2011). People often experience temporal musical entrainment in the automatic, even uncontrollable head-bobbing or toe-tapping that occurs when we listen to musicians play (or the covert activation of motor areas of the brain in the absence of overt

movement), while the height of temporal entrainment might be seen in the complex rhythmic timing and exchange between partners or ensemble members in music or dance (Keller et al., 2007; Keller, 2008; Goebel and Palmer, 2009) as portrayed by the band in **Figure 1**.

A second component of entrainment derives from the mutual sharing of an affective state between individuals. Authors have even spoken of a kind of affective "synchrony," although we prefer the term entrainment in order to distinguish the temporal and affective components while encapsulating the essence of both. Therefore, for the purpose of this paper we will refer to both temporal and affective entrainment. Affective entrainment involves the formation of interpersonal bonds (see Feldman, 2007) and is related to the pleasure in moving the body to music and being in time with others (known in the vernacular as "groove"; Pressing, 2002; Madison, 2006; Janata et al., 2012). Janata et al. (2012) have put forth that temporal and affective components of entrainment in music might be inextricably linked, as for example, the quality of sensorimotor synchronization (temporal) can predict the degree of experienced groove (affective), and affiliation (Hove and Risen, 2009). In searching for the biological roots of temporal and affective entrainment, early musical interactions between infants and their caregivers might provide the first experiences in entrainment (e.g., Malloch and Trevarthen, 2009) and thus provide a foundation for the development of skills in musical joint action (**Figure 1**).



FIGURE 1 | Illustrations of temporal and affective entrainment. The skills of musical ensembles such as a Cuban Jazz band embody temporal entrainment and musical joint action. Affective entrainment

can be seen in early forms of joint action between an infant and caretaker, and might provide a foundation for the development of musical skills.

In joint action in general, multiple individuals coordinate their movements in space and time, with a goal to communicate, or to effect a change in the environment (see Sebanz et al., 2006). Here we take on the term “coordination” commonly used in this literature because movements between co-actors might be coupled in either a synchronized or complementary fashion, in time, or in space. Studies of the spatiotemporal aspects of interpersonal joint action reveal that coordination in joint action can be emergent or planned (Knoblich et al., 2011). Emergent coordination occurs spontaneously due to automatic processes that are grounded in perception and action couplings in the brain (Knoblich et al., 2011). In a non-musical example, when two conversational partners are engaged, they show a coordination of body sway that is unintentional and does not require planned action (Shockley et al., 2003). Unintentional coordination of body sway can also occur between duetting musicians (Keller and Appel, 2010), an example that will be discussed in more detail later. The emergent process can also be described in terms of concepts from dynamical systems theory (e.g., coupled oscillators; see Schmidt and Richardson, 2008; Oullier and Kelso, 2009; Riley et al., 2011). In contrast, planned coordination requires shared representations of the intended outcome of the joint action and a plan for each individual’s own role in the joint action. This can be seen in a variety of activities in which multiple individuals engage, including spoken conversation, or carrying a heavy object together (e.g., Clark, 1996, 2005; Knoblich and Sebanz, 2008; Bekkering et al., 2009; Sebanz and Knoblich, 2009). In music, both emergent (automatic head-bobbing) and planned (playing and dancing along) coordination reflect entrainment and can exert simultaneous influences on the individuals’ actions.

How do we experience emergent and planned coordination in musical interactions where entrainment involves demands on temporal synchronization and complex rhythmic timing? To understand how these forms of musical joint action arise requires an excursion into several different research fields. The same terminology is occasionally used to different ends in these fields, and part of our task consists of an attempt to reconcile this with the aim of merging these diverse fields on common ground (see **Box 1** for a glossary of our terminology and definitions). We begin in the next section by identifying two primary generic modes of musical joint action, chorusing and turn-taking, and their underlying

entrainment mechanisms. Then in the section entitled “Ensemble Skills in Musical Joint Action” we consider how these mechanisms are crystallized in specific cognitive and motor skills that support musical ensemble performance. We then discuss in the section entitled “The Beginnings of Musical Joint Action” some of the potential ontogenetic roots for the cognitive and motor systems required in musical joint action. In the final section, we speculate on a possible mimetic relationship between developmental trajectories in social coordination skills and musical ensemble skills.

COORDINATION MODES IN MUSICAL JOINT ACTION: CHORUSING AND TURN-TAKING

Chorusing and turn-taking are canonical modes of joint action that exemplify emergent and planned coordination. Chorusing occurs when communicative signals (sounds or movements) produced by separate individuals make simultaneous and roughly equal contributions to the joint action as a whole (as in synchronous chorusing observed in several species; Merker, 2000). For example, in multi-part music, monophonic (unison), and homophonic (chordal) textures can be considered to be instances of chorusing. In dance, the “chorus line” and the “corps de ballet” provide definitive examples of chorusing. Turn-taking involves the ordering of communicative signals produced by separate individuals in such a way that there is little temporal overlap (i.e., serial organization) or, when there is overlap, information in the signal produced by only one individual has priority at any given time (hierarchical organization). Musical turn-taking occurs in antiphonal “call and response” textures (as in Afro-American blues and gospel music; Williams-Jones, 1975; Waterman, 1999) and, more generally, when instruments or dancers exchange roles, taking turns at leading and accompanying, such as happens in a Baroque fugue.

The distinction between chorusing and turn-taking is based on the temporal relations between actions of interacting individuals. Musical joint actions, however, have behaviorally relevant spatial, as well as temporal, components. Spatiotemporal movements shape dance forms and define the embodiment of metrical hierarchies (Leman, 2007). In a musical melody, the pitch intervals and contours define trajectories through mental representations of pitch space that can correspond cross-modally to physical space (Rusconi et al., 2006; Lidji et al., 2007; Keller and Koch, 2008; Eitan

Box 1 | A glossary of terms as they pertain to the concepts of musical joint action in this paper. Terms are listed in the order they appear in text.

Entrainment: The spatiotemporal coordination between two or more individuals, often in response to a rhythmic signal (e.g., Phillips-Silver et al., 2010). Examples include playing music or dancing together, in which temporal entrainment is particularly important, and infant–caregiver interactions, in which affective entrainment is often the primary goal (see **Figure 1**). Temporal and affective components of entrainment often coincide in musical joint action, and might be inextricably linked (Janata et al., 2012).

Joint action: Social interaction wherein multiple individuals coordinate their behaviors in space and time to communicate or to effect a change in the environment (see Sebanz et al., 2006). Two types of joint action, emergent coordination and planned coordination, are described below.

Emergent coordination: A type of joint action that occurs by spontaneous, automatic processes that are grounded in links between perception and action.

Planned coordination: A type of joint action that, in addition to basic entrainment, requires shared representations of the intended outcome of the joint action, such as in musical ensemble playing.

Chorusing: The production by separate individuals of simultaneous sounds or movements that signal joint action, such as in a chorus of voices, the beat of a drum circle, or a dance chorus line.

Turn-taking: The production by separate individuals of alternating or dovetailing sounds or movements that signal joint action, such as in “call and response” or musical fugue.

Imitation: An intentional or unintentional spatiotemporal mirroring of the actions (or their effects) of one individual by another. Examples occur in a choral unison or a group’s entrainment to a musical pulse.

Complementary joint action: An action achieved by partners with a common goal, and a systematic but non-identical spatiotemporal relation between the partners’ actions. An example occurs in hierarchical turn-taking as heard in the trading of licks in a Jazz band.

Motor resonance: The automatic, bottom-up, exogenously driven activation of movement-related brain areas (e.g., premotor and sensorimotor cortex). Examples include an infant’s non-conscious mimicry of the posture or affective expression of an adult or another infant.

Action simulation: The controlled, top-down activation of sensory and movement-related brain areas, that does not necessarily rely upon exogenous stimulation. Examples include the mental imagery of a sequence of musical sounds or dance movements.

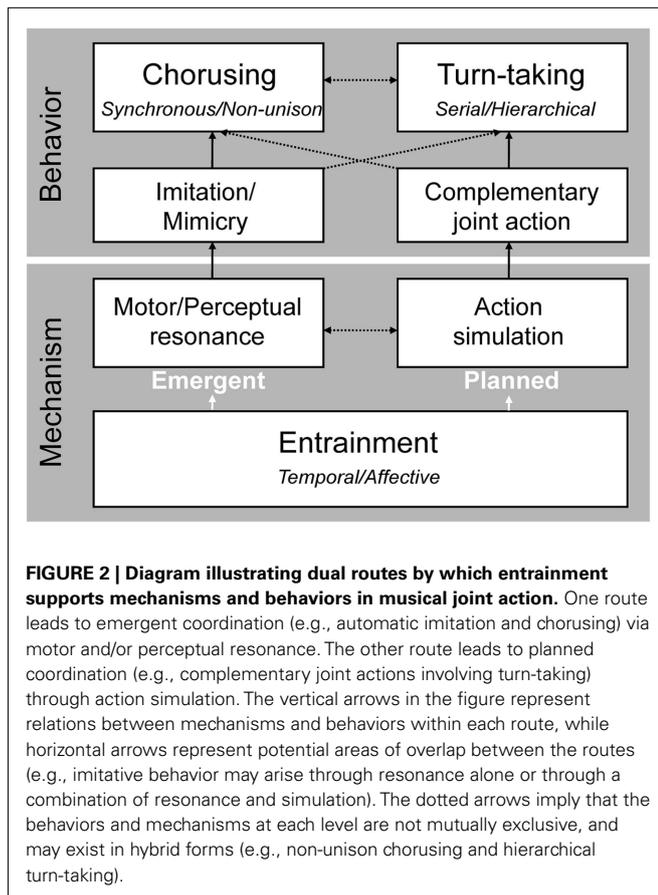
Mimesis: An aesthetic representation or recreation of a physical reality. We propose an example of mimesis in the process of musical practice and expertise, which recalls and builds upon the stage of learning affective and temporal entrainment during infancy (see **Figure 1**).

and Timmers, 2010). To account for such spatiotemporal relations, chorusing and turn-taking can be classified with respect to two further general classes of action: imitation and complementary joint action (**Figure 2**).

Imitation entails one individual mirroring the spatiotemporal features of another’s movements or their effects (see Brass and Heyes, 2005), often in a goal-directed manner (Sebanz et al., 2006). Imitation may occur intentionally, such as when a musician or dancer incorporates a pre-existing sequence in a variation on a theme, or unintentionally, as when two duetting players inadvertently copy each other’s body sway or the effects of that sway on expressive parameters of the sound that they are producing (Keller and Appel, 2010). Imitation may occur at varying time lags, ranging from short intervals that give the impression of simultaneity (as in unison chorusing, or moving bodies or instruments to a common musical pulse) to longer intervals at which the model behavior and the copy do not overlap (serial turn-taking). Complementary joint action, which requires more sophisticated partnership skills, occurs when spatiotemporal features of one individual’s behavior are different from, but systematically related to, those of another individual (see Bekkering et al., 2009). This is exemplified in non-unison chorusing and hierarchical turn-taking.

We assume that entrainment supports imitation and complementary joint action through two brain mechanisms that capitalize on links between perception and action. Specifically, imitation is mediated by motor resonance while complementary joint action requires internal simulation processes that may involve mental imagery. Motor resonance occurs when brain regions that control movement – including motor, premotor, and sensorimotor cortices – are activated automatically by the observation of another’s movements (Rizzolatti and Craighero, 2004). This process may lead to non-conscious mimicry, where individuals unwittingly adopt the facial expressions, mannerisms, or postures of interaction partners (Chartrand and Bargh, 1999). Motor resonance may contribute to the perception of musical pulse (van Noorden and Moelants, 1999; Grahn and Brett, 2007; Chen et al., 2008; Large, 2008) and, in both music and dance, may promote readiness for temporally coordinated, planned action.

Motor resonance evoked during the perception of music and dance does not only function in the service of action, but may also modulate aesthetic responses (Calvo-Merino et al., 2008; Kornysheva et al., 2010; Cross and Ticini, 2011). Motor representations of practiced actions result in stronger motor resonance in adults as well as in infants (Calvo-Merino et al., 2006; van Elk et al., 2008), which suggests that motor resonance is tuned by experience and,



when combined with perceptual resonance (Schütz-Bosbach and Prinz, 2007a), can contribute to action simulation.

Action simulation, as conceived here, is richer than motor resonance to the extent that it involves the activation of sensory, in addition to movement-related, brain regions (cf. Schütz-Bosbach and Prinz, 2007a). Specifically, action simulation occurs when sensorimotor neural processes that resemble those associated with executing an action are engaged in the absence of overt movement (see Gallese et al., 2004; Wilson and Knoblich, 2005; Decety and Grèzes, 2006). Such covert activity may be triggered by observing – or by merely imagining – an action or its effects, for example, body movements in the case of dance (Cross et al., 2006), or tones in the case of music (Keller, 2008). This highlights what we believe to be a fundamental distinction between motor/perceptual resonance and action simulation: While resonance is driven exogenously and automatically (i.e., preattentively) by the perception of external events, simulation may be generated endogenously in the absence of external stimuli (cf. Grush, 2004) and may be modulated by attention.

Action simulation is a mark of expertise, as it is mediated by experience-based associations between sensory and motor processes (Hauelsen and Knösche, 2001; Baumann et al., 2005; Bangert et al., 2006; Zatorre et al., 2007). Auditory–motor simulation is, therefore, especially strong in musicians (Lahav et al., 2007) and visuo-motor simulation is particularly potent in dancers (Calvo-Merino et al., 2005, 2006; Cross et al., 2006). It has been

proposed that action simulation facilitates the understanding of others' intentions and affective states, as well as playing a role in predicting an observed action's immediate outcome and future course (e.g., Wilson and Knoblich, 2005; Decety and Grèzes, 2006; Schütz-Bosbach and Prinz, 2007b; Sebanz and Knoblich, 2009). In musical joint action, predictions based on simulation can facilitate interpersonal coordination in the context of musical textures characterized by chorusing and turn-taking (e.g., Keller et al., 2007; Keller and Appel, 2010).

Instances of unison chorusing – for example, singing a tune such as Auld Lang Syne, or moving one's body in matched fashion with other individuals as in a group folk dance – can be considered to be forms of imitation to the extent that co-actors adopt leader or follower roles. Spatiotemporal interpersonal coordination at this level can especially reveal the contribution of motor resonance triggered exogenously by perceptual input, in addition to the simulation, and action-planning required to execute the musical unison. A complex example that illustrates the contribution of multiple cognitive–motor processes is the Afro-Brazilian Congado – a musical ritual of religious significance that entails the simultaneous performance of separate but proximal communal groups of singers, percussionists, and dancers. A rigorous analysis of Congado performances (Lucas et al., 2011) has shown that inter-group entrainment is influenced by proximity and visual contact, similarity of tempo, and intention (i.e., whether proximal groups resist entrainment), factors which reveal contributions of emergent entrainment and imitation, as well as simulation and action-planning in planned coordination. Finally, during turn-taking, the serial or hierarchical ordering of signals produced by different individuals (e.g., when “trading licks” in jazz or exchanging melody and accompaniment roles in chamber music) constitutes complementary joint action. Effective coordination during turn-taking therefore requires the participating individuals to predict the spatiotemporal trajectories of each other's actions via simulation.

The cognitive and motor demands of chorusing and turn-taking vary as a function of the degree to which these modes of musical joint action require imitation or complementary joint action. The scheme in Figure 2 captures this through the use of arrows to delineate potential links between the two levels of behavior (chorusing/turn-taking and imitative/complementary) and entrainment-based mechanisms (motor/perceptual resonance and action simulation). Note that musical joint action may adopt hybrid forms that house elements of both imitative and complementary joint action (e.g., when one improvising musician imitates another's rhythm while altering the pitches). In the next section we consider in more detail the cognitive and motor processes that serve advanced forms of musical joint action, such as skilled ensemble performance, through links to basic mechanisms of entrainment, motor resonance, and action simulation.

ENSEMBLE SKILLS IN MUSICAL JOINT ACTION

Musicians coordinate their actions with precision and flexibility. To engage in such planned coordination, whether it entails turn-taking or chorusing, ensemble musicians in many musical traditions invest considerable time into rehearsing together in order to form shared representations of the ideal sound (see

Davidson and King, 2004; Ginsborg et al., 2006; Davidson, 2009). According to Vesper et al. (2010), the “minimal architecture” supporting generic joint action includes shared task representations and processes related to action monitoring, prediction, and behavioral modifications that simplify coordination. The latter “coordination smoothers” involve modulations of an individual’s own behavior that render the action more predictable and make it easier for others to coordinate with it (Vesper et al., 2010). In music performance, exaggerated movements associated with breathing, body sway, and ancillary performance gestures such as head nods (Goebel and Palmer, 2009; Keller and Appel, 2010; Keller, forthcoming) may serve as smoothers. Once shared goals are established, these musicians rely upon three core cognitive–motor ensemble skills that are based on entrainment, and allow each individual to coordinate with the actions of co-performers in real time (see Keller, 2008).

The most fundamental ensemble skill is adaptive timing, or adjusting the timing of one’s movements in order to maintain synchrony in the face of unintended temporal perturbations (e.g., by other players) and intended tempo changes in the music. Adaptive timing is mediated by temporal error correction mechanisms that enable internal timekeepers (i.e., neural oscillations at timescales relevant to musical meter) to remain coupled, or entrained in stable phase relations with external signals (Repp, 2005; Large, 2008; Repp and Keller, 2008). Temporal error correction processes that operate automatically support emergent coordination, while others that require attention may be invoked in contexts requiring planned coordination, such as when an ensemble musician intentionally adapts to the expressively motivated tempo changes of a co-performer (Keller, 2008). Research on adaptive timing has yielded evidence for assimilation in interpersonal action timing during dyadic sensorimotor synchronization. This has been observed in joint finger-tapping tasks that require basic in-phase coordination akin to chorusing (Konvalinka et al., 2010), as well as in tasks that require individuals to produce movements in alternation, which involves turn-taking. In these studies, cross-correlation analyses revealed interpersonal dependencies in the timing of co-actors’ finger taps that were suggestive of mutual assimilation. Temporal assimilation may be a form of imitation that facilitates ensemble cohesion by making multiple individuals sound collectively as one.

A more cognitively advanced skill is prioritized integrative attending, that is, simultaneous attention to one’s own actions (high priority) and those of others (lower priority) while monitoring the integrated ensemble sound (Keller, 1999, 2001). Prioritized integrative attending thus relies both on an individual’s ability to divide attention between different sound sources, and on the group’s joint attention skills (to the extent that multiple individuals must attend to the aggregate structure that results from their coordinated actions). This variety of joint attention is particularly important in music characterized by complex turn-taking. For example, to perform the interlocking rhythms of Central African music, according to Nketia (1962), each player must have a general awareness of the resulting aggregate pattern, as well as the “knack” for coming in at the right moment. It has been proposed that the dynamic allocation of attentional resources that is required for prioritized integrative attending relies on entrainment (Keller, 2008).

Specifically, the coupling of internal timekeepers to periodicities associated with the music’s metric structure may provide a temporal schema that allows attention to be modulated in a manner that is optimal for monitoring multiple levels of the musical texture concurrently (Keller, 1999; London, 2004; Jones, 2009). Evidence that such metric schemas facilitate prioritized integrative attending has been found in studies in which musicians are required to memorize one instrumental part, as well as the aggregate structure, of multi-part rhythm patterns (Keller and Burnham, 2005). For example, in one experiment listeners were required simultaneously to memorize a target (high priority) part and the overall aggregate structure (resulting from the combination of parts) of short percussion duets. The key finding was that recognition memory for both aspects of each duet was influenced by how well the target part and the aggregate structure could be accommodated within the same metric framework.

The third ensemble skill is anticipatory imagery, which involves the use of mental imagery in planning the production of one’s own sounds and predicting upcoming sounds of co-performers. Auditory or motor imagery may thus assist in anticipating others’ actions with a view to optimizing ensemble cohesion (Keller, 2008). In support of this hypothesis, it has been shown that individual differences in synchrony between pianists in duos are positively correlated with performance on a task designed to measure the vividness of anticipatory auditory imagery (Keller and Appel, 2010). With regard to the mechanisms underlying this relationship, it has been claimed that anticipatory imagery relies on entrainment-based sensorimotor coupling to drive internal models that trigger mental images during action simulation (Keller, 2008). These internal models, which allow sensorimotor transformations between bodily states and events in the immediate environment to be represented in the brain, are harnessed during action simulation to generate predictions about an action’s future course (Wolpert et al., 1998, 2003). The accuracy and stability of sensorimotor synchronization in the context of real and virtual instances of interpersonal coordination have been shown to depend on temporal prediction abilities (Wöllner and Cañal-Bruland, 2010; Pecenka and Keller, 2011), which are, in turn, positively correlated with temporal imagery abilities (Pecenka and Keller, 2009a,b).

Thus a suite of cognitive and motor skills, operating in concert, supports successful joint action in musical contexts via entrainment-based mechanisms that enable diverse forms of chorusing and turn-taking behavior. This is particularly evident in the performances of expert ensembles, such as the Cuban Jazz band depicted in **Figure 1**. Some of these ensemble skills, in addition to coordination smoothers and shared task representations, can be founded on principles of joint action learned very early in life (**Figure 1**). We now turn to the early foundations of musical joint action.

THE BEGINNINGS OF MUSICAL JOINT ACTION

Joint action as a general form of coordinated social action begins with the earliest relationship. Affective entrainment in social contexts seems to be a predisposition of the infant and mother (Trehub, 2003), as in the “intuitive parenting” that helps to regulate the infant’s emotional state and guide learning in his pre-verbal

environment (Papoušek, 1996b). One of the highlighted aspects of the mother–infant interaction is its musical nature. Musical qualities of typical infant-directed speech include slow and regular tempo, repetition, exaggerated prosody and the accompaniment of movement and gesture. These features may serve as coordination smoothers, as well as functioning to engage the infant's attention (Fernald, 1991), to enable emotional communication (Trainor et al., 2000; Trainor, 2008), and to promote language acquisition (Papoušek, 1996a; Kuhl, 2004).

The infant is responsive to such expressive musical characteristics in the mother's (or other caretaker's) displays, and also participates in “communicative musicality” (Malloch and Trevarthen, 2009; the infant's behavior has also been referred to as “protomusical,” e.g., Cross, 2003), for example, in predicting the timing of the mother's expressions and producing responses that are to some degree temporally coordinated (e.g., Crown et al., 2002). Jaffe et al. (2001) define temporal coordination as a form of interpersonal contingency, and they describe interactional (non-periodic) rhythms which allow an infant and parent to each predict the timing pattern of the other's behavior. This ability to predict timing is necessary in order to eliminate the time lag of a sensorimotor reaction (Merker et al., 2009), and is considered to be essential to bonding within the dyad (Jaffe et al., 2001). In a common game played by adults to engage infants, violation of expectation in timing is used to make the infants laugh, also supporting infants' ability for predictive timing (Stern, 2007). Infants' production of musical actions, as in sung tones or rhythmic movements, begins to emerge during the first year (perhaps especially with encouragement) and can be considered in the context of the communication of the pre-verbal infant (Trehub, in press). Yet before full musical actions emerge, the infants' responsiveness to musicality and the timing of their responses may build upon a repertoire of non-verbal vocal and gestural expressions (Eckerdal and Merker, 2009).

Infants thus engage in a kind of “sympathetic conversation” with their mothers, the timing of which enables them to anticipate and relish in the mothers' expressive displays, and causes them distress if timing is mismatched (see Trevarthen and Aitken, 2001). Temporally coordinated actions can be simultaneous, dovetailing, or alternating (Feldman, 2007), though these interactions are not typically strictly periodic, and are stochastic and bidirectional in organization (Cohn and Tronick, 1988). When mother–infant interactions take the form of chorusing, they have been referred to as “coaction” (Dissanayake, 2000). In this context, the behaviors that infants display with their caregivers include social gaze, facial expressions, and vocal behaviors (by 3 months of age), gesture, and shared attention to objects (after 6 months of age; Feldman, 2007). The roughly unison or matched forms of interaction between infants and their caregivers might be attributed to motor resonance (e.g., Meltzoff and Decety, 2003), and reflect emergent coordination behavior. The function of this emergent coordination is to establish affective entrainment, cooperation, and bonding between the infant and her caregiver (Feldman, 2007; Feldman et al., 2007).

Turn-taking in infancy has been studied primarily in the context of observing spoken conversation, and indicates the importance of the development of social cognition and attention. Infants shift gaze or attention as they observe videos of adults engaged in

conversational turn-taking (von Hofsten et al., 2005), and gaze shifts become increasingly predictive with age (Bakker et al., 2011). By 4–6 months infants are sensitive to cues of social cognition (such as selective attention to face-to-face interactions) in turn-taking (Augusti et al., 2010), which coincides with the development of their ability to use infant-directed speech cues to choose their preferred social partners (Schachner and Hannon, 2011). Before semantics and syntax are shared between conversational partners, cues to turn-taking are manifest in culturally dependent vocal prosody, eye gaze, and body movements (Wilson and Wilson, 2005) but with a universal target of timing in turn-taking (Stivers et al., 2009). A further cognitive skill in interpretation of conversational turn-taking is anticipating transitions in conversation (i.e., when the next speaker will begin to speak), which does not develop until around 3 years of age and might be influenced by language development (von Hofsten et al., 2009). This anticipatory timing skill might also correspond to improvements in planned coordination of actions (Figure 2).

Spatiotemporal imitation appears in the repertoire of young infants – for example in facial (see Meltzoff and Moore, 1997) and manual gestures (Bekkering et al., 2000), as well as affective mirroring, beginning with the first social smiles at just 6 weeks of age (Rochat, 2007) – which are all examples of emergent coordination (Figure 2). The timing of imitation is limited by cognitive–motor maturity but is taught in part by caregivers to young children, often via imitation games aimed at encouraging joint action (Gergely et al., 2002; Sebanz et al., 2006; Papoušek, 2007). In early stages imitation is automatic and relies primarily on motor resonance (Paulus et al., 2011), even more strongly in practiced behaviors such as crawling than in novel behaviors (i.e., walking; van Elk et al., 2008). While many actions of young infants may rely on resonance, according to von Hofsten (2004), even from birth such actions are not mere reflexes but can be motivated, informed, and goal-directed. Evidence for this claim includes infants' interest in tracking and imitating the purpose and the outcomes of observed actions (e.g., von Hofsten and Siddiqui, 1993; Gergely et al., 2002; Gergely and Csibra, 2003). The goal-directed nature of infants' actions, revealing planning, prediction, and motor representations, could enable the progression (with muscular and especially cognitive development) from resonant imitation to more complex forms of joint and complementary joint action.

In the earliest joint actions that are performed by infants with adults, the adult typically helps the infant to achieve her goal, in which case a precise motor representation is not required (Vesper et al., 2010). By around 1 year of age, several cognitive changes facilitate joint action. First, joint attention has emerged (see Rochat, 2007), in which individuals knowingly attend to the same object or event. This skill coincides with monitoring of relative shared attention between the infant and his social partner (Rochat, 2007). The 1-year-old understands intentions, which has been argued (Tomasello et al., 2005) to be the basis for understanding beliefs (i.e., theory of mind) – emerging around 15 months (Onishi and Baillargeon, 2005) and continuing to develop with experience with language and shifting of perspective (Tomasello et al., 2005). At 1 year action-planning is evident, as infants show goal-directed eye movements that reflect motor representations (Falck-Ytter et al., 2006). These motor representations are thought

to be supported by the mirror neuron system: a brain network that is recruited similarly during action perception and action execution (Falck-Ytter et al., 2006), and is thought to be involved in the interpretation of music and dance in adults (Stevens et al., 2001; Zatorre et al., 2007; Overy and Molnar-Szakacs, 2009).

Beyond the first year emerge the abilities to interpret goal-directed actions in a rational manner (e.g., Gergely et al., 2002), suppress imitative motor representations, and eventually perform complementary joint actions (Sebanz et al., 2006). Motivation for these changes stems in part from “shared intentionality,” or the sharing of psychological states in order to reach mutual collaboration (Tomasello and Carpenter, 2007). Planned coordination calls upon the more advanced cognitive–motor skills of precise (and shared) sensory and motor representations, action monitoring, and behavioral modification (cf. Keller, 2008). Presumably, concurrent with the maturation of the above-mentioned cognitive and motor skills (hence less reliance on scaffolding by adults) as well as language development (e.g., von Hofsten et al., 2009), the practice of musical activities during childhood may continue to engage attention and foster coordination skills. For example, prioritized integrative attending, when musical ensemble performers monitor the aggregate sounds (which children’s choirs can do to an extent) presumably builds upon the earlier abilities for joint attention and dividing attention between actors, once a shared goal representation is also established. Between the ages of 2.5 and 3 years, the skills of coordination (timing) in joint action show substantial improvement even if individual performance on a task improves only marginally (Meyer et al., 2010). Complementary roles in joint action appear to be mastered from the age of 3 years (similar to linguistic turn-taking; see von Hofsten et al., 2009), as action-planning and control become refined (Meyer et al., 2010). Improvisation in social contexts may have a role in the building of planned coordination capacities upon emergent coordination capacities. For example, turn-taking has been observed in children’s vocal play, even in improvised and complementary forms (Dissanayake, 2000), and such vocal play is expressed in virtual social contexts, as in imaginary dialogs and play-acting (see Rochat, 2007).

To achieve temporal entrainment in music, it may be necessary to practice the skill of sensorimotor synchronization – that is, the ability to move in time with perceived external events (Repp, 2005). For example, the clapping games and nursery rhymes, songs, and group dances, of school-age children begin to demonstrate the kind of coordination that is required to keep time as a group (e.g., Provasi and Bobin-Begue, 2003), with an external periodic pulse (McAuley et al., 2006), and using multiple levels of metrical hierarchy (Drake et al., 2000). Such music and games, which represent collective social entrainment (Phillips-Silver et al., 2010), may also foster the development of ensemble playing, and play an important role in the improvement of automatic and deliberate adaptive timing skills, attention, and auditory–motor imagery (Keller, 2008).

The ability for precise synchronization seems to mature gradually, probably building on early perceptual abilities for processing the musical beat (Hannon and Trehub, 2005; Phillips-Silver and Trainor, 2005; Winkler et al., 2009). In studies of synchronization of body movement with music or with a musical partner (in

children between the ages of 5 months and 5 years), the children’s motions – or the sounds produced by them – are not tightly synchronized (phase-locked) to the musical beat (Eerola et al., 2006; Kirschner and Tomasello, 2009; Merker et al., 2009; Zentner and Eerola, 2010). This suggests that sensorimotor synchronization in music is not typically developed until sometime later in childhood or near adolescence (Merker et al., 2009; although cases of exceptional children’s musical performances can suggest otherwise, e.g., Merker et al., 2009; Sowinski et al., 2009)¹.

Practice of the affective component of entrainment in joint action is natural, as infants and young children show a predisposition to “groove” to music in social contexts – that is, they are compelled to move to the music and derive pleasure from it (Janata et al., 2012). Infants and toddlers display a variety of dance gestures in response to music (Eerola et al., 2006), and they produce spontaneous dance motions more to music than to speech (Zentner and Eerola, 2010). The social component is clear in that infants’ and toddlers’ dancing is associated with positive affect (Zentner and Eerola, 2010), and young children’s musical drumming is enhanced in a social context (Kirschner and Tomasello, 2009). The development of action simulation in childhood could further facilitate the understanding of others’ affective states (Decety and Grèzes, 2006) and complementary joint action as in musical exchange (Kirschner and Tomasello, 2009), especially when the roles are truly complementary.

DOES ENSEMBLE PRACTICE SHOW MIMESIS OF EARLY COORDINATION?

We have proposed that temporal and affective forms of entrainment together support musical joint action by enabling chorusing, turn-taking, and hybrid modes of interpersonal coordination via dual routes. On a low road, emergent coordination arises through motor and perceptual resonance that underlies imitative behavior, while on a high road, planned coordination is achieved with the assistance of covert internal simulations that facilitate complementary joint action. We then described the relevance of these mechanisms and behaviors to ensemble skills that allow experienced individuals to coordinate their actions during group music making and dance, as well as to the consolidation of social coordination skills in human development. In the present section, we speculate that the development of coordinated social action in infancy and early childhood, and the practice of advanced musical ensemble skills at later stages, are mimetic processes. Specifically, the process of acquiring and refining skills pertaining to temporal adaptation, attention, and anticipation in early human development is repeated in the acquisition and refinement of skills that are required for precise and flexible interpersonal coordination in aesthetic displays of music and dance (**Figure 1**). This may be viewed as a form of mimesis where advanced interpersonal coordination skills serving artistic purposes (aesthetic communication through music and dance) are grounded in early coordination skills that serve basic functions (mother–infant bonding, cognitive–motor

¹In other atypical cases, synchronization ability may not develop adequately to support normal dance behavior (Phillips-Silver et al., 2011). This is referred to as beat deafness, and future research will shed light on the perceptual and action bases of the disorder.

development, and social and cultural learning, cf. Cross, 2001). Indeed, it has been proposed that both the infant–caregiver relationship and the arts including music and dance reveal a propensity in humans to respond to temporally dynamic social stimuli with emotional affiliation and temporal cooperation (Dissanayake, 2000; Miall and Dissanayake, 2003).

We have suggested that musical joint action builds upon early manifestations of affective and temporal entrainment. The affective component of entrainment – which is central to interpersonal synchrony in music – is grounded in emotional resonance and affective mirroring in early infancy (Rochat, 2007). This component is arguably the first and foremost in ontogeny of coordinated action and musical communication. For example, imitative forms of chorusing and turn-taking are mastered more readily than complementary varieties of musical joint action. Imitation may also provide the easiest route to the socio-emotional benefits of joint musical experience. Indeed, synchronous, matched motor activity has been found to foster interpersonal affiliation (McNeill, 1995; Hove and Risen, 2009), cooperation and pro-social behavior (Wiltermuth and Heath, 2009; Kirschner and Tomasello, 2010), and even altruism (Valdesolo and DeSteno, 2011).

The temporal component of entrainment emerges in various forms during childhood: through joint attention activities in infancy, the improvement of adaptive timing, and finally anticipatory processes in school-aged children. Formal music training

plays a role in developing such capacities for attention and executive functioning (see Hannon and Trainor, 2007), as well as refined auditory–motor interactions (see Zatorre et al., 2007). Hannon and Trainor (2007) have suggested that joint task goals, action monitoring, and synchronization can contribute to the benefits of musical training on cognitive and motor processing abilities in musical joint action. As increasing capacities for imitation and complementary joint action are developed, the cognitive and motor demands of chorusing and turn-taking can be met in musical joint action.

Musicians and dancers often strive to attain the height of temporal and affective entrainment, and so we look to the roots of those behaviors to understand the process by which they are embodied. From the earliest musical exchanges between the infant and mother, temporal and affective entrainment serve more than the primary bond: they lay the groundwork for the refinement of skills in joint action and musicianship. In infants and children, as well as in musical experts via a mimetic process, maturation and musical experience result in entrainment that allows for precision and flexibility in timing, a sense of participation and emotional communion, and a musical aesthetic that reveals the complexity and richness of human interaction.

ACKNOWLEDGMENTS

We thank the reviewers for their helpful comments, and Yoel Diaz for permission to use the ensemble photo in **Figure 1**.

REFERENCES

- Augusti, E. M., Melinder, A., and Gredebäck, G. (2010). Look who's talking: pre-verbal infants' perception of face-to face and back-to-back social interactions. *Front. Psychol.* 1:161. doi:10.3389/fpsyg.2010.00161
- Bakker, M., Kochukhova, O., and von Hofsten, C. (2011). Development of social perception: a conversation study of 6-, 12- and 36-month-old children. *Infant Behav. Dev.* 34, 363–370.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H. J., and Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* 30, 917–926.
- Baumann, S., Koeneke, S., Meyer, M., Lutz, K., and Jäncke, L. (2005). A network for sensory-motor integration: what happens in the auditory cortex during piano playing without acoustic feedback? *Ann. N. Y. Acad. Sci.* 1060, 186–188.
- Bekkering, H., de Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R., van Schie, H. T., and Meulenbroek, R. (2009). Joint action: neurocognitive mechanisms supporting human interaction. *Trends Cogn. Sci. (Regul. Ed.)* 1, 340–352.
- Bekkering, H., Wohlschläger, A., and Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Q. J. Exp. Psychol.* 53A, 153–164.
- Brass, M., and Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends Cogn. Sci. (Regul. Ed.)* 9, 489–495.
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249.
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16, 1905–1910.
- Calvo-Merino, B., Jola, C., Glaser, D. E., and Haggard, P. (2008). Towards a sensorimotor aesthetics of performing art. *Conscious. Cogn.* 17, 911–922.
- Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.
- Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Soc.* 7, 507–525.
- Clayton, M., Sager, R., and Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *Eur. Meet. Ethnomusicol.* 11, 3–142.
- Cohn, J. F., and Tronick, E. Z. (1988). Mother-infant face-to-face interaction: influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Dev. Psychol.* 24, 386–392.
- Cross, E. S., Hamilton, A. E., and Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage* 31, 1257–1267.
- Cross, E. S., and Ticini, L. (2011). Neuroaesthetics and beyond: new horizons in applying the science of the brain to the art of dance. *Phenomenol. Cogn. Sci.* 1–12. doi: 10.1007/s11097-010-9190-y
- Cross, I. (2001). “Music, cognition, culture, and evolution,” in *The Biological Foundations of Music*, eds R. J. Zatorre and I. Peretz (New York: The New York Academy of Sciences), 28–42.
- Cross, I. (2003). Music and evolution; consequences and causes. *Contem. Music Rev.* 22, 79–89.
- Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., and Jaffe, J. (2002). The cross-modal coordination of interpersonal timing: six-week-olds infants' gaze with adults' vocal behavior. *J. Psycholinguist. Res.* 31, 1–23.
- Davidson, J. W., and King, E. C. (2004). “Strategies for ensemble practice,” in *Enhancing musical performance*, ed. A. Williamson (Oxford: Oxford University Press), 105–122.
- Davidson, J. W. (2009). “Movement and collaboration in musical performance,” in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 364–376.
- Decety, J., and Grèzes, J. (2006). The power of simulation: imagining one's own and other's behaviour. *Cogn. Brain Res.* 1079, 4–14.
- Dissanayake, E. (2000). “Antecedents of the temporal arts in early mother-infant interaction,” in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 389–410.
- Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 251–288.

- Eckerdal, P., and Merker, B. (2009). "Music' and the 'action song' in infant development: an interpretation," in *Communicative Musicality. Exploring the Basis of Human Companionship*, eds. S. Malloch and C. Trevarthen. (Oxford: Oxford University Press), 241–262.
- Eerola, T., Luck, G., and Toiviainen, P. (2006). "An investigation of preschoolers' corporeal synchronization with music," in *Proceedings of the 9th International Conference on Music Perception and Cognition*, eds. M. Baroni, A. R. Addessi, R. Caterina, and M. Costa. (Bologna: The Society for Music Perception and Cognition and European Society for the Cognitive Sciences of Music), 472–476.
- Eitan, Z., and Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition* 114, 405–422.
- Falck-Ytter, T., Gredebäck, G., and von Hofsten, C. (2006). Infants predict other people's action goals. *Nat. Neurosci.* 9, 878–879.
- Feldman, R. (2007). Parent-infant synchrony: biological foundations and development. *Curr. Dir. Psychol. Sci.* 16, 340–345.
- Feldman, R., Weller, A., Zagoory-Sharon, O., and Levine, A. (2007). Evidence for a neuroendocrinological foundation for human affiliation: plasma oxytocin levels across pregnancy and the postpartum predict maternal-infant bonding. *Psychol. Sci.* 19, 965–970.
- Fernald, A. (1991). Prosody in speech to children: prelinguistic and linguistic functions. *Ann. Child Dev.* 8, 43–80.
- Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci. (Regul. Ed.)* 8, 396–403.
- Gergely, G., Bekkering, H., and Király, I. (2002). Rational imitation of goal-directed actions in 14-month-olds. *Nature* 415, 755.
- Gergely, G., and Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends Cogn. Sci. (Regul. Ed.)* 7, 287–292.
- Ginsborg, J., Chaffin, R., and Nicholson, G. (2006). Shared performance cues in singing and conducting: a content analysis of talk during practice. *Psychol. Music* 34, 167–192.
- Goebel, W., and Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Percept.* 26, 427–438.
- Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442.
- Hannon, E. E., and Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci. (Regul. Ed.)* 11, 466–472.
- Hannon, E. E., and Trehub, S. E. (2005). Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12639–12643.
- Haueisen, J., and Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792.
- Hove, M. J., and Risen, J. L. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–961.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., and Jasnow, M. D. (2001). Rhythms of dialogue in infancy. *Monogr. Soc. Res. Child Dev.* 66, 1–149.
- Janata, P., Tomic, S. T., and Haberman, J. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75.
- Jones, M. R. (2009). "Musical time," in *The Handbook of Music Psychology*, eds. S. Hallam, I. Cross, and M. Thaut (Oxford, NY: Oxford University Press), 81–92.
- Keller, P. (1999). Attending in complex musical interactions: the adaptive dual role of meter. *Aust. J. Psychol.* 51, 166–175.
- Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychol. Music* 29, 20–38.
- Keller, P. E. (2008). "Joint action in music performance," in *Enacting Intersubjectivity: A Cognitive and Social Perspective to the Study of Interactions*, eds. F. Morganti, A. Carassa, and G. Riva (Amsterdam: IOS Press), 205–221.
- Keller, P. E., and Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Percept.* 28, 27–46.
- Keller, P. E., and Burnham, D. K. (2005). Musical meter in attention to multipart rhythm. *Music Percept.* 22, 629–661.
- Keller, P. E., Knoblich, G., and Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Conscious. Cogn.* 16, 102–111.
- Keller, P. E., and Koch, I. (2008). Action planning in sequential skills: relations to music performance. *Q. J. Exp. Psychol.* 61, 275–291.
- Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child. Psychol.* 102, 299–314.
- Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364.
- Knoblich, G., Butterfill, S., and Sebanz, N. (2011). "Psychological research on joint action: theory and data," in *The Psychology of Learning and Motivation*, Vol. 54, ed. B. Ross (Burlington: Academic Press), 59–101.
- Knoblich, G., and Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 2021–2031.
- Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 2220–2230.
- Kornysheva, K., von Cramon, D. Y., Jacobsen, T., and Schubotz, R. I. (2010). Tuning-in to the beat: aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Hum. Brain Mapp.* 31, 48–64.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.
- Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* 27, 308–314.
- Large, E. W. (2008). "Resonating to musical rhythm: theory and experiment," in *The Psychology of Time*, ed. S. Grondin (West Yorkshire: Emerald), 189–232.
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159.
- Large, E. W., and Palmer, C. (2002). Perceiving temporal regularity in music. *Cogn. Sci.* 26, 1–37.
- Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.
- Leman, M., and Naveda, L. (2010). Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in Samba and Charleston. *Music Percept.* 28, 71–91.
- Lidji, P., Kolinsky, R., Lochy, A., and Morais, J. (2007). Spatial associations for musical stimuli: a piano in the head? *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1189–1207.
- London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford: Oxford University Press.
- Lucas, L., Clayton, M., and Leante, L. (2011). Inter-group entrainment in Afro-Brazilian Congado ritual. *Empir. Musicol. Rev.* 6, 75–102.
- Madison, G. (2006). Experiencing groove induced by music: consistency and phenomenology. *Music Percept.* 24, 201–208.
- Malloch, S., and Trevarthen, C. (2009). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford: Oxford University Press.
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H., and Miller, N. S. (2006). The time of our lives: lifespan development of timing and event tracking. *J. Exp. Psychol. Gen.* 135, 348–367.
- McNeill, W. H. (1995). *Keeping Together in Time. Dance and Drill in Human History*. Cambridge, MA: Harvard University Press.
- Meltzoff, A. N., and Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 491–500.
- Meltzoff, A. N., and Moore, M. K. (1997). Explaining facial imitation: a theoretical model. *Early Dev. Parent.* 6, 179–192.
- Merker, B. (2000). "Synchronous chorusing and human origins," in *The Origins of Music*, eds. N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 315–327.
- Merker, B. H., Madison, G. S., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45, 4–17.
- Meyer, M., Bekkering, H., Paulus, M., and Hunnius, S. (2010). Joint action coordination in 2-1/2 – and 3-year-old children. *Front. Hum. Neurosci.* 4:220. doi:10.3389/fnhum.2010.00220
- Miall, D. S. and Dissanayake, E. (2003). The poetics of babytalk. *Human Nature* 14, 337–364.
- Naveda, L., and Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Percept.* 28, 93–111.
- Nketia, J. H. K. (1962). The hocket-technique in African music. *J. Int. Folk Music Council.* 14, 44–52.

- Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240.
- Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258.
- Oullier, O., and Kelso, J. A. S. (2009). “Coordination from the perspective of social coordination dynamics,” in *Encyclopedia of Complexity and Systems Science*, ed. R. Meyers (Berlin: Springer-Verlag), 8198–8212.
- Overy, K., and Molnar-Szakacs, I. (2009). Being together in time: musica experience and the mirror neuron system. *Musik Percept.* 26, 489–504.
- Papoušek, H. (1996a). “Musicality in infancy research: biological and cultural origins of early musicality,” in *Musical Beginnings: Origins and Development of Musical Competence*, eds I. Deliège and J. Sloboda (Oxford, NY: Oxford University Press), 37–55.
- Papoušek, M. (1996b). “Intuitive parenting: a hidden source of musical stimulation in infancy,” in *Musical Beginnings: Origins and Development of Musical Competence*, eds I. Deliège and J. Sloboda (Oxford, NY: Oxford University Press), 88–112.
- Papoušek, M. (2007). Communication in early infancy: an arena of intersubjective learning. *Infant Behav. Dev.* 30, 258–266.
- Paulus, M., Hunnius, S., Vissers, M., and Bekkering, H. (2011). Imitation in infancy: rational or motor resonance? *Child Dev.* 82, 1047–1057.
- Pecenka, N., and Keller, P. E. (2009a). Auditory pitch imagery and its relationship to musical synchronization. *Ann. N. Y. Acad. Sci.* 1169, 282–286.
- Pecenka, N., and Keller, P. E. (2009b). “The relationship between auditory imagery and musical synchronization abilities in musicians,” in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, eds J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, and P.-S. Eerola (Jyväskylä: University of Jyväskylä), 409–414.
- Pecenka, N., and Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Exp. Brain Res.* 211, 505–515.
- Phillips-Silver, J., Aktipis, C. A., and Bryant, G. A. (2010). The ecology of entrainment: foundations of coordinated rhythmic movement. *Musik Percept.* 28, 3–14.
- Phillips-Silver, J., Toiviainen, P., Gosselein, N., Piché, O., Nozaradan, S., Palmer, C., and Peretz, I. (2011). Born to dance but beat-deaf: a new form of congenital amusia. *Neuropsychologia* 49, 961–969.
- Phillips-Silver, J., and Trainor, L. J. (2005). Feeling the beat: movement influences infants’ rhythm perception. *Science* 308, 1430.
- Pressing, J. (2002). Black Atlantic rhythm: its computational and transcultural foundations. *Musik Percept.* 19, 285–310.
- Provasi, J., and Bobin-Begue, A. (2003). Spontaneous motor tempo and rhythmical synchronization in 2-1/2- and 4-year-old children. *Int. J. Behav. Dev.* 27, 220–231.
- Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992.
- Repp, B. H., and Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Hum. Mov. Sci.* 27, 423–456.
- Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi:10.3389/fpsyg.2011.00038
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rochat, P. (2007). Intentional action arises from early reciprocal exchanges. *Acta Psychol. (Amst.)* 124, 8–25.
- Rusconi, E., Kwan, B., Giordano, B., Umiltà, C., and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition* 99, 113–129.
- Schachner, A., and Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-olds. *Dev. Psychol.* 47, 19–25.
- Schmidt, R. C., and Richardson, M. J. (2008). “Dynamics of interpersonal coordination,” in *Coordination: Neural, Behavioural and Social Dynamics*, eds A. Fuchs and V. Jirsa (Heidelberg: Springer-Verlag), 281–308.
- Schütz-Bosbach, S., and Prinz, W. (2007a). Perceptual resonance: action-induced modulation of perception. *Trends Cogn. Sci. (Regul. Ed.)* 11, 349–355.
- Schütz-Bosbach, S., and Prinz, W. (2007b). Prospective coding in event representation. *Cogn. Process.* 8, 93–102.
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint actions: bodies and minds moving together. *Trends Cogn. Sci. (Regul. Ed.)* 10, 70–76.
- Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367.
- Shockley, K., Santana, M. V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332.
- Sowinski, J., Dalla Bella, S., and Laskowska, I. (2009). “Timing abilities in a child drummer prodigy (IF),” in *Poster Presented at ESCOM, Jyväskylä*.
- Stern, D. (2007). *The First Relationship*. Cambridge, MA: Harvard University Press.
- Stevens, K., Malloch, S., and McKechnie, S. (2001). Moving mind: the cognitive psychology of contemporary dance. *Broglia* 15, 55–67.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E., and Levinson, S. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592.
- Toiviainen, P., Luck, G., and Thompson, M. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. *Musik Percept.* 28, 59–70.
- Tomasello, M., and Carpenter, M. (2007). Shared intentionality. *Dev. Sci.* 10, 121–125.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 1–61.
- Trainor, L. J. (2008). The neural roots of music. *Nature* 453, 598–599.
- Trainor, L. J., Austin, C. M., and DeJardins, R. N. (2000). Is infant-directed speech prosody a result of the communication of emotion? *Psychol. Sci.* 11, 188–195.
- Trehub, S. E. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673.
- Trehub, S. E. (in press). “Communication, music and language in infancy,” in *Language, Music, and the Brain: A Mysterious Relationship*, ed. M. Arbib (MIT Press).
- Trevarthen, C., and Aitken, K. J. (2001). Infant intersubjectivity: research, theory, and clinical applications. *J. Child Psychol.* 42, 3–48.
- Valdesolo, P., and DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion* 11, 262–266.
- van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., and Bekkering, H. (2008). You’ll never crawl alone: psychophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage* 43, 808–814.
- van Noorden, L., and Moelants, D. (1999). Resonance in the perception of musical pulse. *J. New Music Res.* 28, 43–66.
- Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003.
- von Hofsten, C. (2004). An action perspective on motor development. *Trends Cogn. Sci.* 8, 266–272.
- von Hofsten, C., Adell, M., Uhlig, H., and Kochukhova, O. (2009). How children with autism look at events. *Res. Autism Spectr. Disord.* 3, 556–569.
- von Hofsten, C., Dahlström, E., and Fredriksson, Y. (2005). 12-month-old infants’ perception of attention direction in static video images. *Infancy* 8, 217–231.
- von Hofsten, C., and Siddiqui, A. (1993). Using the mother’s actions as a reference for object examination in 6- and 12-month old infants. *Br. J. Dev. Psychol.* 11, 61–74.
- Waterman, R. A. (1999). “African influence on the music of the Americas,” in *Write me a few of Your Lines: A Blues Reader*, ed. S. C. Tracy (Amherst: University of Massachusetts Press), 17–27.
- Williams-Jones, P. (1975). Afro-American gospel music: a crystallization of the black aesthetic. *Ethnomusicology* 19, 373–385.
- Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473.
- Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 2, 957–968.
- Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5.
- Winkler, I., Håden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2468–2471.
- Wöllner, C., and Cañal-Bruland, R. (2010). Keeping an eye on the violinist: motor experts show superior timing consistency in a visual perception task. *Psychol. Res.* 74, 579–585.
- Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control

- and social interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 593–602.
- Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci. (Regul. Ed.)* 2, 338–347.
- Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558.
- Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 26 July 2011; accepted: 07 February 2012; published online: 28 February 2012.
- Citation: Phillips-Silver J and Keller PE (2012) Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6:26. doi: 10.3389/fnhum.2012.00026
- Copyright © 2012 Phillips-Silver and Keller. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.