Communication for coordination: gesture kinematics and conventionality affect synchronization success in piano duos

[Laura Bishop](#) & [Werner Goebl](#)

Abstract

Ensemble musicians often exchange visual cues in the form of body gestures (e.g., rhythmic head nods) to help coordinate piece entrances. These cues must communicate beats clearly, especially if the piece requires interperformer synchronization of the first chord. This study aimed to (1) replicate prior findings suggesting that points of peak acceleration in head gestures communicate beat position and (2) identify the kinematic features of head gestures that encourage successful synchronization. It was expected that increased precision of the alignment between leaders' head gestures and first note onsets, increased gesture smoothness, magnitude, and prototypicality, and increased leader ensemble/conducting experience would improve gesture synchronizability. Audio/MIDI and motion capture recordings were made of piano duos performing short musical passages under assigned leader/follower conditions. The leader of each trial listened to a particular tempo over headphones, then

cued their partner in at the given tempo, without speaking. A subset of motion capture recordings were then presented as point-light videos with corresponding audio to a sample of musicians who tapped in synchrony with the beat. Musicians were found to align their first taps with the period of deceleration following acceleration peaks in leaders' head gestures, suggesting that acceleration patterns communicate beat position. Musicians' synchronization with leaders' first onsets improved as cueing gesture smoothness and magnitude increased and prototypicality decreased. Synchronization was also more successful with more experienced leaders' gestures. These results might be applied to interactive systems using gesture recognition or reproduction for music-making tasks (e.g., intelligent accompaniment systems).

Introduction

Interpersonal communication is critical for joint action tasks like playing piano duets, playing team sports, or dancing, which require collaborators to align their intentions and coordinate their actions in time. Communication during these tasks is continuous and interactive, with collaborators constantly adapting their own intentions and actions in response to the signals they receive from each other (Schiavio & Høffding, 2015). The signals exchanged are typically multimodal (e.g., auditory and visual) and multilayered (e.g., involving facial expressions and body movements simultaneously), and can be subtle, comprising only a raised eyebrow or a brief moment of eye contact (Davidson, 2012). Given these complexities, it is often

difficult to determine from a research standpoint exactly what is being communicated and how group members are assimilating incoming information.

Many researchers have used group music-making paradigms to investigate the communication processes underlying interpersonal coordination. Group music-making presents an intriguing context in which to study communication and coordination, since precise coordination must be achieved under inherently ambiguous temporal conditions (even for notated music, timing is only loosely defined by the score). Moreover, the possible means of communication are constrained by the task of performing an instrument, which can limit freedom of movement for much of the body, as well as conventions of public performance, which may prohibit, for example, counting out loud or using a metronome. In recent years, researchers have been applying their knowledge of the communication processes involved in group music-making to computer systems that replicate or react to performer movements (Dahl, 2014; Hoffman & Weinberg, 2011). Such an application requires a detailed understanding of gesture kinematics and how they relate to performers' intentions.

During skilled ensemble performance, most communication through audio and visual channels is nonverbal. Usually, perception of jointly-produced sound gives sufficient information for performers to coordinate, but visual communication can be important too (Bishop & Goebl, 2015). Visual communication is only rarely a matter of one performer giving directions to another; rather, even if there is a designated leader,

collaborating musicians' body movements interrelate (Chang, Livingstone, Bosnyak, & Trainor, 2017; Moran, Hadley, Bader, & Keller, 2015) and can be mutually influential (Badino, D'Ausilio, Glowinski, Camurri, & Fadiga, 2014). Research has shown that musicians move more predictably when performing with others than when performing alone (Glowinski et al., 2013), a finding that parallels observations made elsewhere in the joint action literature (Hart, Noy, Feniger-Schaal, Mayo, & Alon, 2014; Vesper, van der Wel, Knoblich, & Sebanz, 2011). Visual communication is particularly important in more ambiguous temporal contexts (e.g., at abrupt tempo changes or following long pauses), when co-performers' interpretations are less certain to align (Bishop & Goebl, 2015; Kawase, 2013).

The current study investigates the gestures that ensemble musicians use to coordinate piece entrances. Typically, at piece entrances, in the absence of a conductor, one musician in an ensemble acts as the leader and gives the others an audio-visual signal to begin. This visual signal should indicate the timing of the first beat as well as the starting tempo for the piece. Ensemble musicians coordinate piece entrances with varying degrees of success. While professionals typically synchronize their first notes with near-perfect precision (at least in concert), students may require several attempts to begin. Synchronization success can vary depending on a range of factors, including musicians' expertise and familiarity with each other's playing style, the genre and tempo of the music, the number and combination of instruments, and the presence or absence of a conductor. The aim of our study was to

identify factors that contribute to successful coordination at piece onset during piano duo performance. Specifically, we examined how cue gesture kinematics relate to note synchronization.

Kinematics of effective cueing gestures

Successful coordination is partly dependent on the quality of the visual signal given by the leader—particularly at piece onset, where there is no prior audio to indicate when the first notes should fall or at what tempo. Musicians commonly use head gestures to signal piece onsets, regardless of their instrument; head gestures were therefore our focus here, though we acknowledge that much of the upper body, as well as facial expressions and breathing, can be involved. The current study investigated how head movement kinematics communicate beats, and tested four kinematic properties of head gestures that we predicted could help observers detect communicated beats more successfully. This section of the paper develops these hypotheses.

For both conductors and instrumentalists, the kinematics rather than trajectories of cueing gestures have been shown to indicate the position of the beat, or tactus (Luck & Toiviainen, 2006). "Followers" attempting to synchronize with instrumentalists' cueing gestures tend to perceive beats as aligning with major peaks in the leader's head acceleration, rather than with points of direction change in head trajectories (Bishop & Goebl, 2017). This was observed among pianists and violinists in a study of synchronization in duo performance. Performers' head movements were

measured using accelerometers and Kinect sensors as they took turns cueing each other in at the starts of short passages. An aim of the current study was to replicate these findings (H1) using an expanded version of the same procedure and a more precise motion measurement system.

The easiest gestures to synchronize with are presumably those that convey beat position clearly and reliably. If followers aim to align their starting notes with peaks in leaders' head acceleration, then leader/follower synchronization should be most successful when the leader's starting notes align precisely with his or her own head acceleration peaks. The current study tested this hypothesis (H2) by calculating the offset of leaders' first note onsets from major peaks in leaders' head acceleration curves, and relating the magnitude of these offsets to success in note synchronization.

The clarity of a gesture, and how readily others synchronize with it, might also be influenced by its articulation. Wöllner, Parkinson, Deconinck, Hove and keller, (2012) found that observers synchronized finger-taps more successfully with quantitatively averaged conductor gestures, which were low in jerk, than with individual conductor gestures, which were higher in jerk. Jerk, the third derivative of position, indicates the smoothness of acceleration changes. The authors also observed more successful synchronization with marcato gestures, where the differences between acceleration maxima and minima were large, than with legato gestures, where the differences between acceleration maxima and minima were small. Here, we hypothesized

that gestures high in smoothness (low in jerk) would be clearer and synchronized with more successfully than gestures with high jerk (H3).

We also tested the possibility that musicians synchronize more successfully with gestures marked by a larger-magnitude indication of the beat than with gestures marked by a smaller-magnitude indication of the beat (H4). Gesture magnitude was quantified in terms of how far the head travelled along the forwards–backwards axis while indicating the beat. Instrumentalists sometimes exaggerate their gestures at piece entrances and other places where exchanging visual cues benefits synchronization, and a test of how gesture magnitude affects synchronization should indicate whether this is an effective strategy.

Observers' success at synchronizing with instrumentalists' cueing gestures might also relate to the prototypicality of those gestures (Wöllner, Parkinson, Deconinck, Hove, & keller, 2012). People synchronize most successfully with gestures that are similar to those they produce themselves (Keller, Knoblich, & Repp, 2007; Wöllner & Cañal-Bruland, 2010). This effect is generally attributed to the strengthening of action prediction mechanisms that occurs with experience. According to this theoretical perspective, people use their own action planning systems to simulate observed movements—a process that may or may not yield overt motor output (Calvo-Merino, Grezes, Glaser, Passingham, & Harrad, 2006; van der Wel, Sebanz, & Knoblich, 2013). They then predict the course of observed gestures using the same mechanisms

that they use to predict the course of their own gestures. When the observed or performed gesture is similar to gestures a person has performed in the past, action-perception links are stronger and prediction is facilitated. In the current study, we expected that highly prototypical cueing gestures would be more likely than highly idiosyncratic cues to align with followers' own gesture tendencies, and would therefore be easier to predict and synchronize with (H5).

Gesture mimicry to facilitate synchronization

Theories of embodied music cognition posit that we use our own bodies to interpret the musical gestures produced by others (Leman, 2012). In other words, we understand others' motor intentions by overtly or internally mirroring aspects of their actions (Jacob & Jeannerod, 2005; Jeannerod, 2003). Our ability to internally simulate others' gestures is thus central to the concept of embodiment. Simulation mechanisms facilitate the translation of gestures into sound and the translation of sound into expressive gestures (Leman & Maes, 2014).

As stated above, observed actions can be simulated without overt replication, though in some cases the process clearly shapes motor output. For example, an imitation bias is observed among people who are asked to make speeded movements that are either congruent or incongruent (e.g., in terms of magnitude or direction) to irrelevant movements viewed simultaneously on a computer screen (Grosjean, Zwickel, & Prinz, 2009; Schubö, Aschersleben, & Prinz, 2001). Incongruent movements are performed less accurately than

congruent movements, indicating an unintentional assimilation of observed motion parameters into the observer's own performed actions.

At times, people overtly mimic each other's behaviour. This mimicry has cognitive benefits: when people perform gestures that are high in similarity and coordinated in time, their attention is drawn towards each other and their perception and memory for each other's behaviour is facilitated (Macrae, Duffy, Miles, & Lawrence, 2008). Furthermore, moving in rhythmic coordination with others can promote social bonding, improving participants' ratings of partner trust and likeability (Hove & Risen, 2009) and increasing the likelihood of prosocial behaviour (Wiltermuth & Heath, 2009).

The current study included a test of whether coordination of body gestures occurs between duo performers at piece onset. Previously, some correlation in head and upper body sway has been observed within pairs of duo pianists. Goebl and Palmer (2009) found that duo pianists' head movements were more synchronized when they performed under reduced auditory feedback conditions (unable to hear themselves or unable to hear their partner) than when they performed with normal two-way auditory feedback. Despite the heightened synchrony of head movements, however, note synchronization under reduced auditory feedback conditions was poor. Keller and Appel (2010) tracked the upper body movements of piano duos and found that the further the body movements of the primo performer (who usually plays the higher-pitched part)

lagged behind those of the secondo (who usually plays the lower-pitched part), the greater note asynchrony became. Since primo note onsets consistently led secondo note onsets, the authors suggested that congruence between leader/follower relations at the levels of note onsets and body sway may be important for successful ensemble coordination.

Still unclear is whether leader/follower coordination of cueing gestures occurs at piece onset, and to what extent this coordination of body movement relates to note synchronization. In line with theories of embodiment, we hypothesized that coordinating cueing gestures would help performers gauge each other's intended timing and, therefore, facilitate note synchronization (H6). Followers' gestures were expected to mimic the form of leaders' gestures and follow a parallel timecourse.

Current study

This study aimed to assess how kinematic measures affect the synchronizability of ensemble musicians' cueing-in gestures. Motion capture recordings were made of nine piano duos performing short musical passages under alternating leader and follower conditions. The assigned leader for each trial was responsible for cueing in the follower at a particular tempo, with the aim of synchronizing their performance of the passage as precisely as possible. During a subsequent gesture-following task, a subset of the motion capture recordings of leader performances were presented (with audio) to an independent sample of 10 skilled musicians, who tapped in time with the leaders'

performed beats. Using data from this test, a measure of "synchronizability" (i.e., average leader–follower first beat asynchrony) was obtained for each leader gesture.

The alignment between followers' first taps (for gesture-following task participants) or performed beats (for interactive duo performance task participants) and extremes in leaders' head position, velocity, and acceleration curves was examined. Our focus was exclusively on leader–follower synchronization at piece onset, though a similar investigation of how gesture kinematics affect synchronization across the first few beats of a piece could also be made. It was expected that followers' first taps would align with acceleration peaks in leaders' gestures (H1), confirming previous findings (Bishop & Goebl, 2017). It was also expected that the precision of alignment between leaders' first note onsets and their own head acceleration peaks (H2), as well as increased gesture smoothness (H3), magnitude (H4), and prototypicality (H5) would improve the synchronizability of leaders' gestures. Finally, the hypothesis that increased similarity in the movements made by leader–follower pairs at the time of piece onset relates to improved leader–follower note synchronization was tested (H6), using data from the recording sessions.

Methods
Interactive duo performance experiment

Our first experiment investigated pianists' synchronization with their duo partners' cueing-in gestures under interactive conditions, while two-way communication was possible. The hypothesis that

followers synchronize their piece onsets with peaks in leaders' head acceleration was assessed. We also tested for coordination in duos' body movements around piece onsets.

Participants

Eighteen pianists (10 female) recruited from among the students at the University of Music and Performing Arts Vienna completed the experiment. Our sample size was set with the aim of obtaining enough recorded performances to carry out the gesture-following task. Six pianists had minimal experience playing the piano in small ensembles, six had extensive experience, and six were completing a degree in either choral or orchestral conducting. Some pianists (10 of the 18) had completed a similar version of the task for the experiment reported in Bishop and Goebl ([2017](#)). Some of the pianists knew their assigned partner, but none had performed together before. Participants provided informed consent before completing the experiment and received a small travel reimbursement.

"Conductor", "ensemble-experienced", and "ensemble-inexperienced" groups did not differ significantly in terms of age (conductors $M = 28.0$, SD = 7.5; ensemble-experienced $M = 27.2$, SD = 4.7; ensemble-inexperienced $M = 25.2$, SD = 3.4; $F(1, 13) = 0.37$, $p = 0.55$) or years of piano-playing experience (conductors $M = 17.0$, SD = 7.4; ensemble-experienced $M = 22.0$, SD = 5.6; ensemble-inexperienced $M = 17.8$, SD = 3.4; $F(1, 13) = 1.64$, $p = 0.22$). However, the ensemble-experienced group had more experience playing in duos

and other small ensembles (self-rated M = 12.7 out of 15, SD = 1.4; conductors M = 8.5, SD = 2.1; ensemble-inexperienced M = 8.2, SD = 1.3; $F(1, 13) = 18.45$, $p = 0.001$,

$\eta^2 = 0.59$

$\eta^2 = 0.59$

). Only the conductors had prior conducting experience (M = 3 years, SD = 1.7).

Table 1 Musical stimuli are listed with their starting tempi and meters

Stimuli and equipment

Pianists performed 15 passages adapted from the starts of pieces in the Western classical repertoire (Table 1). Some further details on these pieces are given in Bishop and Goebl (2017). A sample piece is shown in Fig. 1. All passages were in duple meter, 2–4 bars in length, multi-voiced (to be played with both hands), and adapted so that the two performers would always start in unison on the first downbeat. Pieces that were likely to be unfamiliar to participants were chosen to encourage communication between duo members and to ensure that they would not have preexisting expectations regarding the tempo. A tempo was selected for each passage based on the original tempo indications in the

score; these ranged from 45 to 220 bpm, with approximate mean interbeat intervals of 111 ms at the slowest tempo and 1000 ms at the fastest tempo.

Pianists performed on two Yamaha CLP470 Clavinovas and faced each other directly, as shown in Fig. 2. Audio and MIDI from the Clavinovas were recorded via a Focusrite Scarlett 18i8 sound card in Ableton Live, along with audio from a standing microphone placed between the two performers (44.1 kHz sampling rate).

Pianists' upper body movements were recorded using an eight-camera (Prime 13) OptiTrack motion capture system. Each pianist wore a jacket and cap, to which 25 spherical markers were affixed (including three on the head). Marker positions were recorded at a rate of 240 frames per second.

To synchronize audio/MIDI and motion capture data, a film clapboard was placed on top of one of the Clavinovas with an OptiTrack marker attached and struck once at the start and end of each block. These claps were clearly discernible in the OptiTrack data and in the audio recorded by the standing microphone, which was recorded in synchrony with the audio and MIDI from the Clavinovas.

Fig. 1

Sample piano duet stimulus: primo (*upper*) and secondo (*lower*) parts for the passage based on Kuhlau's Rondo, Op. 111

Full size image

Fig. 2

Procedure

Pianists were given hard copies of the passage scores at the start of the recording session and had time to practice together. The recording phase began once both performers could play the passages without error.

Recording sessions were divided into two blocks. In each block, the performers played once through each of the 15 passages (in a pseudorandomized order, structured so that passages with a similar tempo were played consecutively). Each performer was instructed to play either the part labelled "A" or the part labelled "B"; these indicated primo and secondo lines and were assigned pseudorandomly, so that each performer played about the same number of primo and secondo parts (7 or 8 of each). Thus, each participant played a total of 30 trials, going once through the 15 passages in each block.

Leader/follower roles were assigned on an alternating basis, so each performer led each passage once. At the start of each trial, the assigned leader was handed a pair of headphones and listened to a metronome beat indicating the tempo for the passage. They then returned the headphones to the experimenter before beginning to play. The leader's task was to coordinate the entrance of the passage without speaking (e.g., counting out loud). Duos were instructed to focus on playing together and to ignore pitch errors as much as possible, but if major timing or pitch errors made it impossible to continue,

they were allowed to redo the trial. Once the recordings were finished, pianists completed a musical background questionnaire.

Analysis

Alignment of audio/MIDI and motion capture data

The experimenter struck a film clapboard at the start and end of each block (see "Stimuli and equipment"). The initial strike was used as a reference "time 0", and the timestamps for all recordings were recoded to indicate elapsed time since this point. To check the precision of this inter-recording alignment, for each recording, the interval between first and second clapboard strikes was calculated, and discrepancies between recording devices in interval lengths were assessed. The mean discrepancy was minimal, less than the duration of one sample of motion capture data (M = 2.9 ms, SD = 2.4).

Gesture position, velocity, acceleration, and jerk

Motion capture data comprised series of x, y, and z axis coordinates for 25 upper body markers, indicating forwards/backwards, left-right, and up/down movement, respectively. Here, we report only on the motion of the front-most head sensor (positioned slightly above the forehead), as motion was also measured from this location in Bishop and Goebl (2017), and the current study was partially designed to validate our earlier findings. For analyses of position and velocity, only forwards–backwards (x axis) data were used. For analyses of acceleration, a 3D measure was computed using the square root of the sum of squares for x, y, and z axes, with gravity added into

the y dimension (gravity was included, again, for the purpose of equating our measures with the earlier work).

Gesture position data were smoothed using functional data analysis (Ramsay & Silverman, 2002; Goebl & Palmer, 2008). Order-7 b-splines were fit to the trajectories with knots every 50 ms, applying a roughness penalty on the fifth derivative (

$$\lambda = 10^{-18} \text{�} = 10^{-18}$$

), which smoothed the third derivative (jerk). The functional data were then converted back for further analysis with samples every 5 ms.

Motion data were segmented into trials, based on visual analysis of the motion capture recordings. A "cue window" was then identified in each trial, comprising the two interbeat intervals prior to the leader's first note onset and the leader's first performed interbeat interval. Interbeat intervals were defined as the duration of a quarter note for pieces written in 4/4 and as the duration of a half note for pieces written in 2/2. Any cueing-in gestures that were given would fall within that

window.

Primo-secondo note asynchronies

MIDI data from the Clavinovas were aligned with the corresponding notation using the performance-score matching system developed by Flossmann, Goebl, Grachten, Niedermayer and Widmer (2010), which pairs MIDI pitches with score notes according to pitch sequence information. Only pitch sequence is considered, so rhythm errors are not penalized. Mismatched pitches resulting from performer error or incorrect interpretation of the pitch sequence by the matching system can be corrected via a graphical user interface. Matched performances thereby include only correctly performed and correctly matched notes. The mean pitch error rate across all completed performances (i.e., excluding false starts, but including all other notes) was 9.5% (SD = 8.8%). Using these matched performances, primo-secondo asynchronies were calculated for notes that should have been synchronized, according to the score. Asynchronies were calculated for the entirety of each performance, but for the analyses presented here, we used the asynchronies achieved on the first chord of each piece as our main dependent variable. Asynchronies were not normally distributed, so non-parametric tests were used.

Gesture-following experiment

A second experiment was carried out with the aim of identifying the kinematic properties that improve cueing gesture synchronizability. Audio-visual recordings of pianist performances, collected during the first experiment, were used as stimuli for a beat-tapping task,

which was completed by a sample of 10 musicians. The average accuracy of these musicians' synchronization was taken as an indicator of gesture synchronizability, serving as a more reliable measure than the accuracy of individual followers' synchronization during the interactive duo performance experiment, due to the larger sample size.

Participants

Ten skilled musicians (2 female) completed the task (age M = 28.1, SD = 3.5). Their experience covered a variety of instruments, including piano (5 musicians), violin (2), saxophone (1), voice (1), and percussion (1). They reported an average 19.8 years of instrument-playing experience (SD = 2.8), and gave an average self-rating of 11.2 out of 15 (SD = 1.8) to describe the extent of their experience playing in duos and other small ensembles. All participants provided written informed consent.

Stimuli and equipment

Musicians were presented with audio–video clips of 108 (of the total 270) leader performances recorded during the main experiment. Only leader performances were used, since they contained the cueing-in gestures. Twelve of the performances with the highest note accuracy were selected at random from each duo. The selected performances represented a wide range of tempi (36–210 bpm; median 82.6 bpm)—nearly the full range used in the interactive duo performance task (listed in Table 1). The videos featured a point-light representation of the leader, shown as black dots connected by straight black lines on a white background

(Fig. 3). Point-light figures excluded hand markers and hip markers, as these had not been visible to followers during the original performances, but included the head (front, top, and right markers) and all other upper body markers.

3D images of each frame were drawn up in R (using the "rgl" package), then combined into videos in VideoMach. Presentation (Neurobehavioral Systems), running on an HP EliteBook, was used to display videos and play the accompanying temporally-aligned audio.

Footnote

1 Only leaders' audio was presented. Videos were displayed in a 1280 × 720 pixel box on a black screen and played at 60 fps. The corresponding musical notation was shown below the video display, in a 1200 × 180 pixel box. During familarization trials (see below), the complete score for each passage (including both primo and secondo parts) was presented in a 1250 × 500 pixel box, and the full (primo and secondo) audio recordings were played.

Fig. 3

Sample display for an experimental trial during the cue quality test

Full size image

Participants' tapping responses were made on the middle C key of one of the Clavinovas used during the recording sessions. (A Clavinova was used so that the motor requirements would be similar to those encountered by the pianists who made the recordings.) The output volume was turned off, so participants heard

only the thud of their finger on the key in addition to the stimulus audio. The PC presenting stimulus clips was placed on top of one of the Clavinovas. So that MIDI response data could be aligned with video timestamps, a photoresistor was taped to the top of the computer screen, at the corner of the video display. The photoresistor registered the change in lighting that occurred the start of each trial; this information was transmitted as audio data and recorded via a Focusrite sound card in Ableton, in sync with the MIDI data from the Clavinova.

Procedure

The experiment was completed in six blocks. The first block constituted a familiarization phase: scores for 8 of the 15 passages were presented along with the corresponding audio tracks (one duo's recording of each passage was selected at random from those with good synchronization and few pitch errors). Participants were instructed to tap along with the beat of each sounded passage, aligning their first taps as closely as possible with the first notes. The purpose of this block was twofold: (1) to familiarize participants with the passages they were about to hear and (2) to obtain a measure of the delay between piece onsets and first taps when participants could only react to the start of the music, not predict it. Participants were free to choose at which hierarchical level of the beat they tapped (e.g., per eighth note, quarter note, or half note, depending on the tempo of the piece).

Two blocks of 25 experimental trials were then completed, starting with five practice trials. Practice

trials were supervised by the experimenter, who reminded participants of the task instructions as necessary. At the start of each trial, the score for the upcoming performance was presented; participants were free to look it over, then pressed a key on the computer keyboard to start the audio–video recording. They were again instructed to tap along with the beat of the music, using the video recordings to help in aligning their first taps with piece onsets. Only recordings of the pieces presented during the first familiarization phase were included in these blocks.

The second half of the experiment followed the same pattern as the first. A second familiarization block was completed, in which scores for the remaining seven passages were presented and participants tapped along with the beat of the corresponding audio. Two blocks of 29 experimental trials followed. At the end of the session, participants completed a musical background questionnaire.

Fig. 4

Head position, velocity, acceleration, and jerk curves for one performance of a duo in the ensemble-inexperienced group. Each plot shows the pianists' movements during the cue window. The *solid lines* indicate the leader's movements and the *dotted lines* indicate the follower's movements. *Vertical lines* have been added to show the position of the performers' first note onsets; *horizontal lines* show the length of their first IBIs. The *bolded* segments of the leader's position, velocity, and acceleration *curves* indicate the peak-trough pair that was identified as corresponding to the main cue gesture

Full size image

Results
Indicators of beat position

A central aim of this study was to test the hypothesis that points of peak acceleration in instrumentalists' cueing gestures communicate beat position (H1). This prediction was addressed with an analysis of how followers' first note onsets aligned with leaders' gestures. More specifically, we assessed the alignment between followers' first onsets and extremes in leaders' head position, velocity, and acceleration curves. If beats were to be communicated via head position, it is logical to expect that beat locations would align with points of path reversal, as these occur in all gestures, regardless of their trajectory. If beats were to be communicated via head velocity or acceleration, they would likely coincide with either maxima or minima in the velocity or acceleration curves.

Peaks and troughs, therefore, were identified in the cue window of each leader's position, velocity, and acceleration curves. Peaks were defined as points preceded by five consecutively increasing observations and followed by five consecutively decreasing observations that were outside the 99% confidence interval for a surrounding window of 300 ms. Troughs were defined as points preceded by consecutively decreasing observations and followed by five consecutively increasing observations that were likewise outside the 99% confidence interval for a surrounding 300 ms window. A constant rather than tempo-adjusted window size was used, as adjusting for tempo would have required a subjective judgement of at what hierarchical level of the beat each performer had gestured (e.g., 2 beats per bar vs. 4 beats per bar).

Cue gestures were assumed to have at least two points of path reversal, so peak-trough pairs separated by no more than one beat were identified. Since we expected the cue gesture to be more prominent than other movements made during the cue window, the peak-trough pair spanning the greatest range in position, velocity, or acceleration values was selected. The time interval between each selected peak and trough and the follower's first note onset was calculated as an indication of the precision of their alignment. Peak-to-onset and trough-to-onset intervals were averaged across trials to produce a mean interval for each follower. Interval durations were divided by performers' average interbeat intervals to achieve normalized values with units of interbeat intervals (IBIs). Sample head position, velocity, acceleration, and jerk curves are given in Fig. 4.

Table 2 Peak-to-onset and trough-to-onset distribution medians and standard deviations (in IBIs) for followers from both experiments

Full size table

If either peaks or troughs in a given dimension indicate beats, then peak- or trough-to-onset intervals could be expected to cluster around two points: interval lengths of approximately 1 IBI would occur if the selected gesture feature preceded the follower's onsets by one IBI (communicating a preparatory beat), while interval lengths of approximately 0 IBIs would occur if the feature and the follower's onsets were synchronized. For our purposes, a clustering of intervals around either value was taken as an indication that the point communicated beat position. To assess the reliability of alignment between the selected peaks and troughs and

followers' first onsets, the proportion of average intervals approximating either 0 IBIs or 1 IBI (±0.2 IBIs) was calculated. Separate analyses were done for followers from the interactive duo performance task and gesture-following task.

Statistics for peak- and trough-to-onset interval distributions are presented in Table 2. To make sense of these data, we have to consider both (1) how reliably followers' onsets aligned with each landmark and (2) around which values each distribution centered. Reliable alignment with a particular landmark (high percentages in columns 3 and 6 of Table 2), plus a median value near 0 or 1 IBI, would be evidence that the landmark communicates beat position.

Interval distributions for the interactive duo performance and gesture-following tasks are shown in Figs. 5 and 6, respectively. For head position, we found that followers in both tasks aligned their first onsets more closely with peaks than troughs. Medians for peak-to-onset distributions were near 1 IBI, while medians for trough-to-onset distributions were not. For head velocity, neither peaks nor troughs seemed to communicate beats, as alignment percentages were low and medians were not reliably near to either 0 or 1 IBI. For head acceleration, the results were a bit more complex: followers aligned their onsets more closely with peaks than troughs in the interactive task, while in the gesture-following task, onsets aligned slightly more closely with troughs than peaks. The period of deceleration between acceleration peak-trough pairs might have communicated beats to participants in the

gesture-following task—a possibility that is considered in the discussion. For both tasks, however, followers aligned their first onsets more reliably with acceleration landmarks than with position or velocity landmarks.

Fig. 5

?

Distributions of peak-to-onset and trough-to-onset intervals for followers in the interactive duo performance task. Intervals within the *vertical dotted lines* were counted as approximately equivalent to 0 or 1 IBI in length. The proportion of intervals that fell within these ranges are given for each distribution, along with the distribution median and standard deviation

Full size image

Fig. 6

?

Distributions of peak-to-onset and trough-to-onset intervals for gesture-following task participants. Intervals within the *vertical dotted lines* were counted as approximately equivalent to 0 or 1 IBI in length. The proportion of intervals that fell within these ranges are given for each distribution, along with the distribution median and standard deviation

Full size image

Across motion parameters, a timing difference was noticeable between the interactive duo performance and gesture-following tasks: gesture-following participants' first taps tended to align with a later point on leaders' gesture curves than did interactive duo followers' first onsets. For example, peaks in position and acceleration preceded interactive duo followers' first onsets by slightly less than one beat and gesture-following participants' first taps by slightly more than one beat. Correspondingly, leader-follower asynchronies were greater for gesture-following participants than for interactive duo participants, $Z = 20.44$, $p < 0.001$, $r = 0.56$ (gesture-following task $M = -0.16$ IBIs, SD = 0.64 IBIs;

interactive duo task $M = -0.01$ IBIs, SD = 0.17).

Footnote
2 This timing difference could reflect better anticipation of the beat among interactive duo followers or a task-dependent difference in how beats were perceived.

The potential effects of leader gesture/follower onset alignment on note synchronization were assessed as an additional test of which gesture parameters were most useful in communicating beat position during the gesture-following task. Only position peak-to-onset, velocity peak-to-onset, and acceleration peak- and trough-to-onset interval distributions were considered, since their medians were close to 1 IBI (Table 2, column 7). For each distribution, trials with intervals approximating 0 or 1 IBI (±0.1 IBIs; "aligned") were compared to trials without intervals approximating 0 or 1 IBI ("not aligned"), using mean absolute asynchronies of first tapped beats as the dependent variable. Significantly improved synchronization (at

$\alpha = 0.01$

�

=

0.01

) was observed when first taps aligned with position peaks, $Z = 2.95$, $p = 0.003$, $r = 0.11$ (aligned $M = 0.33$ IBIs, SD = 0.29; not aligned $M = 0.41$ IBIs, SD = 0.33), acceleration peaks, $Z = 4.83$, $p < 0.001$, $r = 0.15$ (aligned $M = 0.33$ IBIs, SD = 0.32; not aligned $M = 0.41$ IBIs, SD = 0.34), and acceleration troughs, $Z = 7.13$, $p < 0.001$, $r = 0.22$ (aligned $M = 0.27$

IBIs, SD = 0.27; not aligned M = 0.42 IBIs, SD = 0.34). No significant difference was observed for velocity peaks, Z = 2.02, p = 0.04. These findings provide evidence that leaders' head trajectories and acceleration patterns are used as cues to beat position.

Gesture properties that support successful synchronization

In this section, analyses testing the potential effects of gesture kinematics and leader expertise on leader–follower synchronization are presented, using data from the gesture-following task. Asynchronies obtained from the gesture-following task were not normally distributed, so the results of non-parametric tests are reported.

Alignment between leaders' gestures and sounded performance (H2)

Increased precision in the alignment between leaders' first note onsets and their own cueing gestures was expected to facilitate leader–follower note synchronization. To test this hypothesis, the time intervals between leaders' first note onsets and peaks and troughs in their head position, velocity, and acceleration curves were assessed, using the same analysis procedure as described in the previous section. This analysis had the additional effect of clarifying which kinematic landmarks correspond to leaders' intended beats.

Interval distribution statistics are presented in Table 3. As we saw for followers, leaders' first onsets aligned more closely with peaks than troughs in head position—

only the peak-to-onset interval distribution median was close to 1 IBI. Leaders' first onsets did not reliably align with either velocity peaks or troughs. For acceleration, alignment was more precise and reliable with peaks than troughs, as evidenced by the peak-to-onset distribution median near 1 IBI and the relatively high proportion of leaders whose average peak-to-onset intervals approximated 1 IBI in length. Leaders' onsets aligned slightly more reliably with acceleration peaks than with position peaks, as we saw in the previous section for followers' onsets.

We also tested whether note synchronization was more successful in the gesture-following task on trials where leaders' head position or acceleration peaks either aligned with or preceded their own first onsets by 1 IBI (±0.1 IBIs) than on other trials. The difference in mean absolute note asynchronies was significant (at

$\alpha = 0.03$

$\diamond$

$=$

0.03

) for position peaks, $Z = 2.44$, $p = 0.01$, $r = 0.08$ (aligned $M = 0.32$ IBIs, SD = 0.22; not aligned $M = 0.40$, SD = 0.33), but not acceleration peaks, $Z = 1.01$, $p = 0.31$ (aligned $M = 0.36$, SD = 0.27; not aligned $M = 0.41$, SD = 0.35). The alignment of leaders' first onsets with peaks in their own head trajectories, therefore, systematically improved note synchronization.

Table 3 Peak-to-onset and trough-to-onset distribution means and standard deviations (in IBIs) for leaders

Gesture smoothness and magnitude (H3–4)

Better synchronization was expected with gestures that were smooth than with gestures that were high in jerk. Better synchronization was also expected with gestures that provided a large rather than small magnitude indication of the beat. For each trial, an average value of 3D gesture jerk was calculated (using the root sum squared of jerk values in x, y, and z dimensions), and a measure of gesture magnitude (calculated as the spatial distance between the leader's maximum and minimum head positions) was obtained. The degree of correlation between these values and the mean absolute asynchronies achieved by participants in the gesture-following task on their first tap of each trial were assessed. There was a positive correlation between mean gesture jerk and mean absolute asynchronies,

$\tau = 0.22$

�

$=$

0.22

, $z = 3.26$, $p = 0.001$ (significant at

$\alpha = 0.03$

�

$=$

0.03

), suggesting a tendency for asynchrony to increase with

increasing jerk. Gesture magnitude correlated slightly but significantly with mean absolute asynchronies,

$$\tau = -0.19$$

$, z = 2.48, p = 0.01$, indicating that asynchronies decreased as gesture magnitude increased.

Gesture prototypicality (H5)

Gestures that followed prototypical patterns of motion were expected to encourage more successful synchronization than gestures that followed idiosyncratic patterns of motion. To obtain a measure of "gesture prototypicality", we evaluated how similar each gesture was to all other gestures in the stimulus set. Cross-correlations were calculated between all recorded leaders' cue gestures, within and between duos. For each gesture, a mean absolute lag-0 correlation magnitude was then computed. The acceleration curves with the lowest and highest prototypicality (i.e., highest and lowest mean correlation magnitudes, respectively) are shown in Fig. 7.

Correlations were calculated between average lag-0 correlation magnitudes and the mean absolute asynchronies achieved by participants in the gesture-following task, on the first beat of each trial. Positive correlations (at

$\alpha = 0.02$

�

$=$

0.02

) were observed for head position,

$\tau = 0.19$

�

$=$

0.19

, $z = 2.87$, $p = 0.004$, velocity,

$\tau = 0.38$

�

$=$

0.38

, $z = 5.81$, $p < 0.001$, and acceleration,

$\tau = 0.25$

�

$=$

0.25

, $z = 3.89$, $p < 0.001$, indicating that as gesture prototypicality increased, mean asynchronies also increased. Thus, contrary to our hypothesis, followers synchronized less successfully with leaders who gave more prototypical gestures.

Fig. 7

?

Leader acceleration curves with the lowest (*left*) and highest (*right*) measured

prototypicality values. Recording followers' acceleration curves are shown as *dotted lines*. *Vertical lines* indicate the position of leaders' first note onsets; *horizontal lines* mirrored on either side of the onset *line* indicate the duration of leaders' first performed IBI. The curve segment corresponding to the main cue gesture is in *bold*. The mean absolute asynchrony across gesture-following task participants was 0.88 IBIs for the gesture with the lowest prototypicality and 0.35 IBIs for the gesture with the highest prototypicality

Evaluating predictors of synchronization success

The potential value of the gesture attributes discussed above as predictors of followers' synchronization success was evaluated via multiple regression. A (non-hierarchical) model was constructed that included (1) leader experience group (ensemble-inexperienced, ensemble-experienced, conductor-pianists), (2) leaders' note alignment with their own head acceleration peaks, (3) gesture jerk, (4) gesture magnitude, and (5) gesture prototypicality as predictors. Mean absolute asynchronies achieved by participants in the gesture-following task on their first taps served as the dependent variable.

The overall model was significant, $F(5, 720) = 27.32$, $p < 0.001$, (adjusted)

$R^2 = 0.18$

$R^2 = 0.18$

. It accounted for a low proportion of variance, but this is

not surprising given the number of factors involved in synchronizing with visual cues. Significant effects at an adjusted

$\alpha = 0.01$

$\alpha = 0.01$

were observed for gesture magnitude, $t(720) = 3.73$, $p < 0.001$,

$\eta^2 = 0.02$

$\eta^2 = 0.02$

, gesture jerk, $t(720) = 2.96$, $p = 0.003$,

$\eta^2 = 0.01$

$\eta^2 = 0.01$

, and gesture prototypicality, $t(720) = 10.24$, $p < 0.001$,

η

$$\eta^2 = 0.12$$

$$\eta^2 = 0.12$$

. We also found a significant effect of leader experience: synchronization was more successful with ensemble-experienced pianists' gestures than with ensemble-inexperienced pianists' gestures, $t(720) = 4.77$, $p < 0.001$,

$$\eta^2 = 0.03$$

$$\eta^2 = 0.03$$

, and more successful with conductor-pianists' gestures than with ensemble-inexperienced pianists' gestures, $t(720) = 3.35$, $p < 0.001$,

$$\eta^2 = 0.81$$

$$\eta^2 = 0.81$$

. The effect of leader gesture-note

alignment, $t(720) = 1.61$, $p = 0.11$, was not significant. We can conclude, therefore, that increased ensemble and conducting experience, increased gesture smoothness and gesture magnitude, and decreased gesture prototypicality contribute to improved follower synchronization.

Gesture coordination and note synchronization in interactive duo performance task

Similarity in leader–follower gesture patterns (H6)

It was hypothesized that, during the interactive duo performance task, some followers would make gestures that were similar in timing and form to the gestures made by leaders. To assess the similarity in movements made by leaders and followers, cross-correlation functions were calculated between leaders' and followers' head position, velocity, and acceleration curves, for each trial, in intervals of 15 ms, up to a maximum lag of three IBIs.

For each trial, the lag with the strongest absolute correlation was identified. Positive correlations indicated that the leader and follower were moving in-phase; negative correlations indicated that they were moving in anti-phase. Correlation values and corresponding lags are reported in Table 4. Moderate negative correlations were observed between absolute maximum correlation values and their corresponding lags for position,

$$\tau = -0.27$$

�

$=$

$-$

0.27

, $z = 6.56$, $p < 0.001$, velocity,

$\tau = -0.26$

�

$=$

$-$

0.26

, $z = 6.33$, $p < 0.001$, and acceleration curves,

$\tau = -0.25$

�

$=$

$-$

0.25

, $z = 6.28$, $p < 0.001$ (all significant at

$\alpha = 0.02$

�

$=$

0.02

), suggesting that when leader and follower gestures aligned more closely in time, the degree of similarity in their movements also increased.

Table 4 Leader–follower cross-correlations

We also examined whether greater temporal alignment

in performed gestures related to note synchronization. As a measure of temporal alignment between gestures, we used the lag that corresponded to the maximum correlation value. When maximum correlations occurred at lags close to 0, this would indicate high temporal alignment between leader and follower. Mean absolute note asynchronies achieved on trials in which maximum correlations occurred close to lag 0 (±0.3 IBIs) were compared to the asynchronies achieved on all other trials. None of these tests yielded significant results (at

$\alpha=0.02$

$$\diamond$$

$$=$$

0.02

), $Z = 0.71$, $p = 0.48$ (position), $Z = 2.07$, $p = 0.04$ (velocity), $Z = 0.05$, $p = 0.96$ (acceleration), indicating that note synchronization success did not depend on the temporal alignment of leaders' and followers' head gestures.

Fig. 8

Bars indicate the mean asynchronies per piece achieved by conductor-pianist, ensemble-experienced, and ensemble-inexperienced duos during the interactive duo task. *Error bars* indicate standard error

Full size image

Effects of ensemble and conducting experience on note synchrony in duo performance

Leader experience was found to affect the quality of synchronization by participants in the gesture-following task (see above). An ANOVA was run on the absolute mean asynchronies achieved by participants in the

interactive duo performance task, on the first onset of each piece, to test whether the same between-group differences would emerge. It was expected that the experience shared by members of the conductor-pianist and ensemble-experienced duos would enable both better leading and better following than was the case for ensemble-inexperienced duos, resulting in more successful synchronization among conductor-pianist and ensemble-experienced groups. The effect of ensemble experience was not significant, however, $F(1, 265) = 2.93$, $p = 0.09$. Figure 8 shows the mean asynchronies achieved by duos in each group.

General discussion

This study aimed to identify the kinematic landmarks in pianists' cueing-in gestures that indicate beat position and the kinematic parameters that improve gesture synchronizability. Motion capture recordings were made of pianists' upper body movements as they performed short passages under assigned leader/follower conditions. Audiovisual recordings of leaders' performances were then presented to an independent sample of musicians, who tapped in synchrony with the beat of the music. As explained below, communicated beats occurred near points of backwards-to-forwards path reversal (head position peaks), and, with still greater reliability, near acceleration peaks. Gesture synchronizability was influenced by leader experience and gesture jerk, magnitude, and prototypicality.

Kinematic landmarks indicating beat position

Leaders' first onsets aligned more precisely with acceleration peaks than with other kinematic landmarks,

as evidenced by the high percentage of average acceleration peak-to-onset intervals that approximated 1 IBI. There was also some alignment with position peaks, but no evidence of alignment with position troughs, suggesting that head position also makes some contribution to the communication of beats. The first onsets performed by followers during the interactive duo performance task likewise aligned most reliably with peaks in leaders' head acceleration, with a high percentage of average acceleration peak-to-onset intervals approximating 1 IBI. Acceleration peaks preceded both leaders' and interactive duo followers' first onsets by slightly less than 1 IBI. Alignment with position peaks was again more reliable than alignment with position troughs, indicating that beats tend to be communicated when leaders are near points of backwards-to-forwards path reversal.

The first taps performed by followers during the gesture-following task showed greater alignment with position peaks than with position troughs, as was the case for participants in the interactive duo performance task. In contrast to recording followers, however, participants in the gesture-following task showed even more reliable alignment with acceleration troughs than with acceleration peaks (63% of average trough-to-onset intervals and 50% of average peak-to-onset intervals approximated 0 or 1 IBI). Gesture-following task participants' first taps followed acceleration peaks by slightly more than 1 IBI and acceleration troughs by slightly less than 1 IBI, suggesting that perceived beats may have fallen in between these points.

The slight delay in gesture-following task participants' first taps relative to interactive duo followers' first onsets, apparent in the shifted distribution medians, was presumably a result of the difference in task completed by the two groups. The reduced information available to followers during the gesture-following task meant that only movement cues could be used to predict piece onsets. The access followers in the interactive duo performance task had to other cues, such as facial expressions and the sound of breathing, might have influenced their prediction processes, leading to earlier onsets.

The significance of acceleration patterns in communicating beat position may relate to the kinematics of sound-producing gestures. The sound-producing striking gestures used in drumming or piano-playing are similar in form to the head-nodding gestures studied here, as in both cases, changes in gesture trajectory are involved in communicating beats. Research on air drumming has shown that, when people are instructed to mime drumming gestures in synchrony with a sounded rhythm, acceleration peaks in their hand gestures slightly precede audio onsets, while "hits" (points of downwards-to-upwards path reversal) lag slightly (and more variably) behind audio onsets (Dahl 2014). Sharp decelerations from peak acceleration points, therefore, indicated beat position. In piano-playing, peaks in finger acceleration correspond to moments of key impact and, as such, sounded beats (Dalla Bella & Palmer, 2011; Goebl & Palmer, 2008). If peak accelerations typically lead the sound onsets produced by percussive sound-producing gestures, then

intrinsic knowledge of this association could shape our performance and perception of non-sound-producing gestures, including the cueing-in gestures intended to communicate timing information.

Maximizing gesture synchronizability

Our finding that gesture synchronizability improved with increasing gesture smoothness and magnitude is in line with prior research suggesting that synchronization is more successful with averaged conductor gestures that are low in jerk (Wöllner et al., 2012). During joint action tasks, people tend to reduce the variability and increase the magnitude of their gestures, and the positive effects of gesture smoothness and magnitude observed here show that gesture predictability can improve as a result. These effects might have played a particularly strong role in the context of the gesture-following task, given that leaders' upper body movements were the only source of timing information prior to piece onset. Normally, facial expressions are also involved, and likely to help with securing the follower's attention and discriminating the cueing-in gesture from other preparatory gestures. The sound of the leader's breathing also often acts as a cue to piece onset, but was not part of the audio presented to followers during the gesture-following task. These factors, along with the inability of followers to interact with the leader, would account for the much higher variability in synchronization success that gesture-following task participants achieved, relative to recording followers.

Contrary to our hypothesis, gesture prototypicality related to a decline, rather than an improvement in

gesture synchronizability. Our measure of prototypicality (i.e., the average lag-0 cross-correlation coefficient for each gesture) gave preference to flatter, less distinctive curves. Given the relationship between gesture magnitude and synchronizability, it is therefore not surprising that synchronization was less successful with gestures scoring high in prototypicality. We should also note that none of the pianists in this study performed gestures with particularly idiosyncratic trajectories, in contrast to our previous study, in which a few musicians displayed noticeably idiosyncratic movement styles. Future studies might specifically recruit such individuals in order to disentangle the effects of gesture magnitude from individuality in movement style.

The hypothesis that synchronization would be more successful among interactive duo performance task participants when the follower's movements mirrored the leaders' movements was not supported. Both this and our hypothesis that synchronizability would improve with increasing gesture prototypicality derived from the idea that observers use their own motor systems to interpret and predict others' gestures. While there was evidence that some leader–follower pairs performed head movements that were similar in form and closely aligned in time, leader–follower gesture alignment had no effect on the success of note synchronization. Different results might arise when cueing gestures are more directly tied to sound onset. For example, when pianists perform together on a single piano, they often use exaggerated wrist movements to help synchronize chords. Aligning wrist movements

could help pianists regulate their timing and prove beneficial for note synchronization.

During the gesture-following task, synchronization was more successful with gestures performed by pianists who had either conducting experience or substantial experience performing in small ensembles than with gestures performed by pianists who had little ensemble experience. This difference in synchronization success suggests that conducting and ensemble performance experience improve pianists' cueing gestures similarly. During the interactive duo performance task, slightly superior synchronization was observed among ensemble-experienced duos. The absence of similarly enhanced synchronization among conductor duos suggests that pianists whose cueing experience comes from conducting, rather than ensemble-playing, may be skilled at leading but less skilled at following. In the literature, it has been suggested that good duo coordination may depend on at least one member of the pair having good ensemble "following" skills, which include strong anticipation and timing adaptation abilities (Keller, 2008).

The reduced form of the gesture-following task presented musicians with a situation that differed substantially from normal duo performance. Followers were forced into an exclusive follower role, in which their responses had no effect on leaders' behaviour. Normally, people attempting to synchronize their actions will adapt to each other, regardless of their assigned role (Goebl & Palmer, 2009; Konvalinka, Vuust, Roepstorff, & Frith, 2010). Followers were also

forced to rely exclusively on leaders' movement cues to predict piece onsets, while in normal performance conditions, a range of other cues would be available, including facial expressions and the sound of breathing. Pianists who participated in the interactive duo performance task were not told how their performances would be presented during the gesture-following task, so their movements were natural, and not deliberately exaggerated. Future research should investigate how much use musicians make of their co-performers' gestures when other cues are available. It should also be noted that the performance arrangements to which this study is most relevant are those in which performers have a direct view of each other, and in the future, other viewing angles should be considered.

Conclusions

The results of this study show that acceleration patterns communicate beats in skilled musicians' cueing-in gestures. Musically-trained observers aligned their performed beats with the periods of sharp deceleration that followed acceleration peaks. The communicative quality of cueing gestures depended on their smoothness, magnitude, and prototypicality; both ensemble performance and conducting experience improved the quality of cueing gestures given.

These results may have implications for artificial musical systems employing gesture modelling and recognition functions. In recent years, there has been increased interest in the development of systems that people can interact with musically, some of which use gesture recognition to modulate sound output.

Identification of which kinematic landmarks in musicians' gestures indicate beat positions would benefit systems designed to synchronize discrete output with users' rhythmic body movements (Dahl, 2014). Some recently-developed systems, designed to fill the role of an accompanist or duet partner, output music in real time to accompany human performances. Output timing is continuously adjusted in response to fluctuations in the human performer's timing to maintain coordination. There have been some attempts to introduce expressive and receptive visual communication capabilities to such systems in the form of prescribed gestures, given at prescribed times (Maezawa & Yamamoto, 2016), but an improved understanding of how beats are communicated naturally and how to communicate beats clearly would allow for more robust and natural communication between the system and the human performer.

Music as a coevolved system for social bonding
Authors: Patrick E. Savage Faculty of Environment and Information Studies, Keio University, Fujisawa 252-0882, Japan psavage@sfc.keio.ac.jp http://PatrickESavage.com Psyche Loui College of Arts, Media and Design, Northeastern University, Boston, MA 02115 p.loui@northeastern.edu http://www.psycheloui.com Bronwyn Tarr Institute of Cognitive and Evolutionary Anthropology & Department of Experimental Psychology, University of Oxford, Oxford, OX2 6PN, UK bronwyn.tarr@anthro.ox.ac.uk; bronwyntarr01@gmail.com

https://www.anthro.ox.ac.uk/people/dr-bronwyn-tarr Adena Schachner Department of Psychology, University of California San Diego, La Jolla, CA 92093 schachner@ucsd.edu https://madlab.ucsd.edu Luke Glowacki Department of Anthropology, Pennsylvania State University, University Park, PA 16802 glowacki@psu.edu https://www.hsb-lab.org/ Steven Mithen Department of Archaeology, University of Reading, Reading RG6 6AB, UK s.j.mithen@reading.ac.uk http://www.reading.ac.uk/archaeology/about/staff/s-j-mithen.aspx W. Tecumseh Fitch Department of Cognitive Biology, University of Vienna, Vienna 1090, Austria tecumseh.fitch@univie.ac.at https://homepage.univie.ac.at/tecumseh.fitch/
Abstract: Why do humans make music? Theories of the evolution of musicality have focused mainly on the value of music for specific adaptive contexts such as mate selection, parental care, coalition signaling, and group cohesion. Synthesizing and extending previous proposals, we argue that social bonding is an overarching function that unifies all of these theories, and that musicality enabled social bonding at larger scales than grooming and other bonding mechanisms available in ancestral primate societies. We combine cross-disciplinary evidence from archaeology, anthropology, biology, musicology, psychology, and neuroscience into a unified framework that accounts for the biological and cultural evolution of music. We argue that the evolution of musicality involves gene-culture coevolution, through which proto-musical behaviors that initially arose and spread as cultural inventions had feedback effects on biological evolution due to their impact on social bonding. We emphasize the deep links between production, perception, prediction, and social reward arising from repetition, synchronization, and harmonization of rhythms and pitches, and summarize empirical evidence for these links at the levels of brain networks, physiological mechanisms, and behaviors across cultures and across species. Finally, we address potential criticisms and make testable predictions for future research, including neurobiological bases of musicality and
relationships between human music, language, animal song, and other domains. The music and social bonding (MSB) hypothesis provides the most comprehensive theory to date of the biological and cultural evolution of music. Keywords: comparative; cooperation; cultural evolution; harmony; language; music; prediction; reward; synchrony; vocal learning

1. Introduction Darwin famously considered music to be a puzzle for evolutionary theory. Music is universal across human cultures (Brown & Jordania 2013; Mehr et al. 2019; Savage et al. 2015), yet its function seems mysterious, since "neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man in reference to his daily habits of life…" (Darwin 1871, p. 33). Darwin went on to speculate that music first evolved "for the sake of charming the opposite sex", after which language "derived from previously developed musical powers". Since Darwin there has been no shortage of hypotheses about why and how music evolved (cf. Honing et al. 2015; Wallin et al. 2000). The null hypothesis is that music is an evolutionarily "useless" by-product of other evolved capacities, with no adaptive function and involving no direct selection for musical capacities (Pinker 1997, p. 528). Others hypothesize that musicality evolved for specific adaptive purposes, including signaling mate quality (Miller 2000), advertising male coalitions (Hagen & Bryant 2003; Merker 2000), or soothing infants (Dissanayake 2000; Falk 2004; Mehr & Krasnow 2017). Many authors have discussed the evolutionary value of music in facilitating group cohesion (e.g., Benzon 2001; Brown 2000a, 2007; Cross & Morley 2009; Dissanayake 2009; Dunbar

2012a; Freeman 2000; Gioia 2019; Huron 2001; Loersch & Arbuckle 2013; McNeill 1995; Merker et al. 2018; Mithen 2005; Oesch 2019; Patel 2018; Roederer 1984; Schulkin & Raglan 2014; Trainor 2018; Trehub et al. 2018), sometimes suggesting that music may have arisen via group selection (especially Brown 2000a). Although such proposals succeed in explaining some properties (or genres) of music, we argue that no single account succeeds as a general explanatory framework for the evolution of human musicality. Our purpose in the present target article is to synthesize and extend previous proposals into a new, parsimonious framework that can explain and predict many aspects of human music-making. Our argument is that human musicality is a coevolved system for social bonding. Crucially, following Honing (2018) and others, we clearly distinguish between music and musicality. "Music" encompasses the diverse cultural products generated by and for music making: songs, instruments, dance styles, etc. In contrast, "musicality" encompasses the underlying biological capacities that allow us to perceive and produce music. Distinguishing these clearly is crucial because musical systems are diverse, culture-specific products of cultural development, while musicality comprises multiple biological mechanisms, shared across human cultures, that enable musical production, perception and enjoyment1 . Musicality is not a monolithic trait evolved to solve one particular problem (coalition signaling, infant mood regulation, sexual attraction, etc.), but rather a set of capabilities that can be utilized in different ways to support multiple functions, all involving social affiliation, but no one of which is the "primary" or "original" function. The key phrase "social bonding" refers to the formation, strengthening and maintenance of affiliative connections ("bonds") with certain conspecifics (i.e., the set of social processes that engender the bonded relationships that underpin prosocial behaviour). As a group-living primate species, such bonds are psychologically and biologically central to human survival and reproduction (e.g. via enhanced predator protection, cooperative child-rearing, collaborative foraging, expansion and defense of territories; Hrdy 2009, Dunbar & Schultz 2010, Dunbar 2012b, Tomasello & Vaish 2013). For the purpose of this paper we use "social bonding" as an umbrella term to encompass both bonding processes (over short and longer time scales) and their effects. Consequently, we take "social bonding" to encompass a variety of social phenomena including social preferences, coalition formation, identity fusion, situational prosociality, and other phenomena that bring individuals together. The social functions of music share a general social utility: to forge and reinforce affiliative inter-individual relationships, for example by synchronizing and harmonizing the moods, emotions, actions or perspectives of two or more individuals. Crucially, we argue that music achieves this in a variety of situations where language is less effective, and on a scale greater than that achievable by the ancestral bonding mechanisms available to other primates (e.g., grooming). We argue that social bonding promotes, and is the consequence of, interactions not only during music making, but also subsequently via long-lasting changes in affiliative dispositions of group members towards one another, and their 1 For discussion of practical and ethical challenges involved in defining and comparing "music" and "musicality" in cross-culturally valid ways, see Nettl (2015); Savage (2019b); Jacoby, Margulis, et al. (2020); Ewell (2020); Danielle Brown (2020); and Iyer & Born (2020). https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https:// www.cambridge.org/core/terms. associated longer-term prosociality. Because social interactions involve multiple levels of group structure, our conception of social bonding also includes darker phenomena such as out-group exclusion that bring certain individuals closer together by pushing away others (see Section 6.4). The final keyword here is "coevolved": we argue that culturally evolving systems of music (Savage 2019a) have developed in tandem with the human capacity for musicality through a process of gene-culture coevolution. We

build on recent arguments by Patel (2018) and Podlipniak (2017), who suggest that music arose initially as a cultural "invention" that created the context for later selection enhancing human musicality. In much the same way that the use of fire by early hominins provided the preconditions for biological adaptations to cooked food (Wrangham 2009), or the invention of dairy farming in some European and African cultures created selection for lactase persistence (Tishkoff et al. 2007), early instantiations of music provided the selective preconditions for later neurobiological changes underlying human musicality. Notably, both Patel and Podlipniak identified social bonding as a candidate function driving such gene-culture coevolution, with Patel (2018, p. 118) noting the possibility that "musical behavior first arose as a human invention and then had (unanticipated) beneficial effects on social cohesion." We argue that because music had multiple adaptive effects on social bonding, this led to subsequent selection (both genetic and cultural) for the ability and motivation to make particular forms of music -- music that has features that most effectively function to promote social bonding. This combination of cultural and biological selection led to the particular features and ubiquity of modern human music and musicality. Our article closely examines this claim, and provides a framework for understanding the biological and cultural evolution of music, taking this argument as foundational. We provide a detailed cross-disciplinary review of the evidence for specific mechanisms by which music functions to enhance social bonding, and consider how some of the mechanisms underlying musicality may have coevolved with music. Like Patel, we take for granted the large and sophisticated literature on gene-culture coevolution in general, and will not review it here (cf. Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985; Durham 1991; Jablonka & Lamb 2005; Laland et al. 2000, 2010; Richerson et al. 2010; Henrich 2016; Tomlinson 2018). However, we do not see the "invention" of music as a unitary event later followed by genetic adaptation, but rather as an iterated process where different proto-musical components of musicality arose over an extended period as behavioral innovations that, due to initial positive effects, generated new cognitive and social niches for subsequent biological adaptations, themselves yielding new innovations, etc. in a virtuous spiral. We thus posit essentially an iterated Baldwin effect (Baldwin 1896; Bateson 2004; Griffiths 2003; Podlipniak 2017), or more generally, prolonged cognitive "niche construction" (Laland et al. 2000). This mechanism is closely related to many contemporary models of language evolution involving a series of "protolanguages" (Arbib 2005; Fitch 2010, 2017). Although hypotheses about the specific ordering of events involved (e.g., Mithen 2005, Dunbar 2012a) are useful, it is not our purpose here to propose a specific sequence, but rather to advance a new conception of the entire process. Mehr, Krasnow, Bryant and Hagen (in this issue) present a contrasting hypothesis for the origins of music. Their hypothesis synthesizes and extends their previous proposals (Hagen & Brant 2003; Mehr & Krasnow 2017) into a generalized "credible signaling" hypothesis that incorporates signaling of both coalition strength and parental attention. They also present critiques of the social bonding hypothesis and other candidate hypotheses. The BBS editors decided that publishing these two target articles with contrasting hypotheses would stimulate productive commentary beyond that usually possible for only a single target article. Both target articles originated from the same symposium on "The Origins of Music in Human Society"2 , but differ in multiple ways in addition to the focus on social bonding vs. credible signaling. In particular, Mehr et al. take an approach grounded in evolutionary psychology, focused on demonstrating domain-specificity and evidence for adaptation. In contrast, our approach emphasizes cultural evolutionary theory, including in particular gene-culture coevolution and cognitive niche construction (cf. Laland & Brown 2011). We take a pluralistic approach to adaptation and modularity, involve experts from diverse disciplines to synthesize evidence into a single framework, and propose testable predictions for future research. We expand on more detailed contrasts between the two

articles in Section 6. The following sections lay out the details and implications of the music and social bonding (MSB) hypothesis. Section 2 describes the proposed evolutionary functions and coevolutionary process. Section 3 details cross-disciplinary evidence supporting the MSB hypothesis. Section 4 specifies the neurobiological mechanisms proposed to underlie music's social bonding functions. Section 5 describes testable predictions that follow from the MSB hypothesis. Section 6 addresses a number of potential criticisms of our hypothesis, and Section 7 provides a brief conclusion. 2 The symposium program is available at https://www.iast.fr/sites/default/files/IAST/conf/royaumont/ royaumont_program051217.pdf

2. Social bonding as a unifying function in the evolution of musicality The Music and Social Bonding (MSB) hypothesis posits that core biological components of human musicality evolved as mechanisms supporting social bonding. Musicality relies on multiple neurocognitive components, which likely evolved at different times and for different reasons: musicality is more a cognitive toolkit than a single tool (Fitch 2015a). Most of the tools in this musical toolkit function to facilitate social bonding, but some may also be used for non-social purposes such as individual mood regulation (see Section 6.5). We avoid arguing for one specific single adaptive function for music (e.g. coalition advertisement, courtship, or infant mood regulation) because we think it unlikely that a single "main" evolutionary function for complex, multi-component abilities like language or music exists. Imagine asking the parallel question "what is vision for?" and coming up with a hypothesis set including "spotting predators", "judging mate quality", "finding food", and "avoiding obstacles". It seems clear that these are all functions of vision, and all provide potential causal explanations for adaptive improvements in vision during evolution. But the desire to identify ONE function as primary seems misguided. A better approach is mechanistic: we ask "what are lenses for?," and answer in engineering terms: lenses are for focusing an image on the retina, to enable accurate visual perception. Whether the image is of a predator, mate, or food is not critical, since improved visual resolution will aid them all. Turning to music, "social bonding" provides an umbrella explanation analogous to "vision is for seeing". Particular design features of music (singing discrete pitches, generating an isochronous beat, use of repetitive patterns based on small-integer ratios) function mechanistically to enhance predictability, aiding synchronization and harmonization when multiple people sing, dance, and play instruments together. Coherent and harmonious merging of sounds and movements during group activity leads to positive feelings of prediction, fulfillment of expectation, and mutual accomplishment. These, through activation of the dopaminergic reward system and other pathways, have affiliative emotional and rewarding effects immediately and also long after music-making ceases (see Section 4). Crucially, the resulting strengthened social bonds are operative over multiple types and sizes of groups, ranging from dyads (e.g., parent and infant, potential mates) to bands of small coalitions and large groups of unrelated individuals (Fig. 1). Social bonding through music thus produces its ultimate evolutionary dividends in multiple complementary ways, including a larger group of potential allies, increased child rearing success, increased mating success, and better-functioning coalitions. Figure 1. We propose that supposedly competing hypotheses for the evolution of human music, including mate bonding, parent-infant bonding, and group cohesion (within both small coalitions and larger groups), are complementary subcomponents of a broader social bonding function.

2.1 Ancestral bonding mechanisms Why was

social bonding adaptive for our ancestors, and in what ways does music improve or increase social bonding? Group living comes with costs (e.g., increased local competition for food and mates) and benefits (e.g., safety in numbers, cooperative hunting/defense). Animals that live in groups, particularly primates, have evolved mechanisms that help balance these costs and benefits by forging strong affiliative bonds: good quality, persistent, differentiated inter-individual commitments that require investment of time and energy (Dunbar 1991). Strong social bonds enhance individuals' prospects of receiving support through coalitions, which, in certain primate species, influence dominance rank and reproductive performance (Silk 2007). These coalitions form the backbone of successful cooperative hunting, child rearing, and joint defense against predators or competitors (Dunbar & Shultz 2010). Ecological factors typically constrain the size of a group, but larger groups of well-coordinated, strongly bonded humans enabled exploitation of new forms of resources (e.g., larger prey), and more reliable protection from predators (Dunbar 2012b). Ancestral bonding mechanisms (ABMs) in other primates include grooming, play, and - in some species - non-procreational sex. These ABMs are essentially dyadic (or for play, very small groups mostly limited to young animals), and require substantial time commitments even in small groups if all individuals in the group are to invest in all others. Although vocal duets are present in tropical birds and some primates (Farabaugh 1982; Haimoff 1986; Mann et al. 2009; Thorpe 1972), group vocal choruses that are both differentiated and coordinated appear nearly unique to humans (but see Mann et al. 2006 for the fascinating example of the groupchorusing plain-tailed wren). As Dunbar (1993) has argued, the steady increases in group size, complexity, and fluidity that occurred during hominin evolution put increasing strain on ABM-based social bonds. Beyond group sizes of 20 or so, dyadic bonding based on ABMs like grooming became unsustainably time-consuming, so supra-dyadic bonding mechanisms were needed. Dunbar (2012a) suggests that another ABM in great apes and humans was laughter (Davila-Ross et al. 2009), which facilitates social bonds among reasonably large groups. However, there are limits to a bonding mechanism based on laughter: Unlike music, which people can intentionally choose to engage in at any time, large group laughter can be difficult to elicit and to sustain for long periods. Music may have provided our ancestors with a novel system that, like laughter, allowed for simultaneous bonding with a larger group of individuals, but across a broader set of times and contexts, and for longer periods of time than otherwise possible (Dunbar 2012a; Launay et al. 2016). This new system augmented the smaller-scale ABMs that became less robust in larger groups. Specific design features of human musicality - particularly our capacity and proclivity to produce repetitive, synchronized, harmonized music for extended periods – provided a flexible toolkit for bonding, allowing our ancestors to achieve social bonding on a large scale. 2.2 Design features of musicality 2.2.1 Rhythm and dance Most music has two distinctive rhythmic components: an isochronous (equal-timed) beat, and a metric structure (a hierarchical arrangement of sonic events into small groups with differentially accented constituents; Arom 1991; London 2004; Savage et al. 2015). These features together provide a predictable, repetitive structure underlying extended, coordinated and varied group performances, while allowing room for variation and improvisation. Isochronicity and metric structure make the performance predictable, which facilitates planning synchronized and coordinated movements (e.g., dancing). While synchronization solely to the beat (e.g. in marching or unison chanting) allows large groups to integrate, it tends to submerge individual contributions. Meter solves this problem by allowing many individuals to contribute, out of phase, to the same integrated rhythm. Neither of these core design features of musicality appears well-designed for solo performances, but they support the synchronized and coordinated musical sounds and dance movements of groups that are widespread features of human musical systems (Savage et al. 2015). Dancing is another intrinsically rhythmic component of human musicality (cf. Fitch 2015a/b; Laland et al. 2016).

Even newborn infants perceive a musical beat (Winkler et al. 2009), and dance develops early: infants hearing music produce spontaneous rhythmic movements during their first year, although the ability to entrain these movements reliably to a beat takes several years to develop (Kim & Schachner, 2020; McAuley et al. 2006; Merker et al. 2009; Zentner & Eerola 2010). The capacity to perceive and move to a beat is a core component of musicality, rare among vertebrates (Patel 2014; Schachner et al. 2009) but universal across human cultures (Donald Brown 1991). Dance provides an energetic mode of musical participation that is accessible to large numbers of individuals regardless of age, familiarity with the music, or instrumental/singing virtuosity. In addition to its visual effects, dance can also generate an auditory signal, e.g. due to foot stamping or hand clapping, and certain styles of dance (such as tap dancing) create their own sonic accompaniment. These factors suggest that dance is a core part of music-making ("musicking") and not a separate domain (Tarr 2017). Dance thus expands the potential circle of rhythmically coordinated participants in musical interactions. The inclusive aspect of human musicality provided by dance is predicted by the MSB hypothesis, but poses a challenge to hypotheses seeing music primarily as a signal of virtuosity. Hereafter, we consider dance a core component of musical performance. 2.2.2 Melody, harmony, and vocal learning The human capacity for song entails vocal production learning: the ability to imitate and learn vocal patterns beyond our speciestypical repertoire of screams, laughter, etc. By about 2 or 3 years of age (often earlier), children reproduce songs that their caregivers sing to them, with intact pitch range and contours (Trehub 2016). Young children commonly exhibit greater fluency in song than in speech (e.g., singing Twinkle Twinkle Little Star from beginning to end with fractured, word-like sounds). This vocal learning ability is highly developed in humans relative to other primates, and the neurobiological mechanisms of its evolution are relatively wellunderstood, due in part to its convergent evolution in songbirds and other non-human species (Janik & Slater 1999, Jarvis 2019; Fitch 2015a; Syal & Finlay 2011; see Section 4.4 for details). Vocal learning forms a foundation for group participation in singing culturespecific songs. In contrast to the continuously varying pitch of normal speech, the discrete pitches used in song and instrumental music generate predictable sequences that enable frequency matching between individuals during group music production (Merker 2002; Savage et al. 2015). Unison performance in which multiple parts produce the same melodies at either the same frequencies (1:1 frequency ratio) or an octave apart (2:1 ratio) is so widespread among humans it is often not even considered a form of harmonization (although cf. Jacoby et al. 2019 for evidence that octave equivalence is not completely universal). Octave singing in particular represents the most universal form of musical harmony: different pitches performed simultaneously with maximally overlapping acoustic spectra (cf. Bowling & Purves 2015). The common tendency for men and women to sing together in octaves is paralleled by the roughly octave difference in men and women's average vocal pitch, based on vocal anatomy (Titze 1989). This is an unusual feature among primates (and mammals more generally) not observed in chimpanzees (Grawunder et al. 2018) – a potential anatomical adaptation for vocal harmonization. Harmonious overlapping of acoustic spectra also shapes another common design feature: Musical systems around the world restrict pitches to scales containing a limited number of discrete pitch classes (rarely more than 7; Savage et al. 2015). These pitch classes often reflect small-integer frequency relationships which sound consonant together (e.g., the 3:2 frequency ratio underlying musical fifths, 4:3 ratios for fourths, etc.; Bowling et al. 2018; Gill & Purves 2009; Kuroyanagi et al. 2019; McDermott et al. 2010; Terhardt, 1984). By producing pitches that adhere to scales, groups

of singing individuals effectively minimize uncertainty in fundamental frequency, thus maximizing harmony via spectral alignment (Sethares 2004). Coordinating with other individuals musically, by aligning acoustic spectra, can sound pleasing and promote bonding. The specific mechanisms and causal relationships behind this effect remain contested (Bowling & Purves 2015; Bowling et al. 2017, 2018; Harrison & Pearce 2019; Jacoby et al. 2019; Large et al. 2016; McBride & Tlusty 2020; McDermott et al. 2010, 2016; Merker et al. 2018; Pfordresher & Brown 2017). Regardless, scales facilitate harmony, where individual voices/instruments combine consonantly – another design feature supporting group coordination but not solo performance. 2.2.3 Repetitive structure The synchronization of rhythms and harmonization of pitches described above is facilitated and enhanced by the widespread use of repetitive musical structures (Savage et al. 2015). Structural building blocks can range from short rhythmic and/or melodic motives of only a few notes, to entire phrases, to large-scale sections or entire works. The level of repetition in music is one of its most striking differences from language (Fitch 2006; Margulis 2014), and multiple repetitions of a recording of a spoken phrase cause it to sound sung rather than spoken (Deutsch et al. 2011). Repetition enhances memorization and predictability, allowing multiple performers to engage in long periods of coordinated music-making, with all-night music-and-dance rituals common from contemporary Western nightclub culture to ethnographic descriptions of small-scale societies (Merriam 1964; Thornton 1995). In contrast, language and ancestral bonding mechanisms such as laughter are more difficult to sustain for long periods, making them less suitable for the kind of sustained inclusive interactions that promote the strongest social bonds. However, extreme repetition can lead to boredom and to a dearth of memorable distinguishing features, preventing music from serving as a cue of social identity (see below). Both human and bird songs tend to balance repetition and novelty in the form of repetition with variation (Kroodsma 1978; Lomax 1968). 2.2.4 Music and social identity A final potential design feature of culturally-transmitted group music concerns its role in flexibly and hierarchically indicating kinship and group identity (Stokes 1994; Turino 2008). Because songs are variable, complex and memorable, two people knowing the same song likely acquired this knowledge via social learning -- and thus are likely to share a common socio-cultural history. Thus, shared knowledge of musical repertoire provides information about shared socio-cultural background (Soley & Spelke 2016; Schachner et al. 2018). Musicality may have coevolved in support of this social bonding function: Cultural innovations created a wide variety of musical styles and features, and musical knowledge became a cue to social history and cultural group membership. This created selective feedback favoring individuals who tended to perceive music as a cue to group membership, as they would have more accurate ideas about others' social group membership. This hypothesized combination of cultural and biological evolution would lead to an evolved bias to use music as a cue to guide and facilitate social interactions, consistent with findings that shared musical knowledge serves as a social cue from early in childhood through adulthood (see Sections 3.3 & 3.4). Synchronized and harmonized group performances help cement group identity, and eventually allow skilled participation in ritualized performances to serve as a hard-to-fake indicator of group membership. Furthermore, the existence of diverse pieces and sub-styles allows subgroups to express their uniqueness within a broader shared musical repertoire or style. Such expressions of identity at multiple hierarchical levels are useful because human biological and cultural evolution has been characterized by increasing complexity of social structure, as exemplified by the large-scale nation-states characteristic of modern human societies (Turchin et al. 2018). Thus, group musical performance - including

dance - facilitates lasting, culturally evolving indicators of group identity and bonds - akin to passwords or shibboleths (cf. Feekes 1982, Fitch 2004) – that extend beyond individual recognition and memory, aiding intercultural marriage and trade. 2.3 Gene-culture coevolution These specific design features and their interactions – dancing to an isochronous beat with a metrical hierarchy, singing learned melodies based on discrete scales in harmony, using predictable, repetitive musical structures, and using musical performances as cues for social identity – are widespread throughout the world's musical systems (Savage et al. 2015; see Section 3.1). These features have clear functions for group performance, but little or no function in solo performance (hence their rarity in birdsong, whale song, and certain solo human music genres such as lament; Tolbert 1990, Frigyesi 1993). These design features are therefore predicted a priori by the MSB hypothesis, but not by solo signaling hypotheses such as sexual selection for mate attraction (Miller 2000) or maternal singing to infants (Mehr & Krasnow 2017; Mehr et al. This issue). While these features promote coordination in dyadic music (e.g., duets) and memorability/communicative power in solo music (e.g., lullabies; Corbeil et al. 2016; Cirelli & Trehub 2020), their added value in supporting extended, coordinated group performances is most evident for larger groups. MSB posits an extended timeline in which different core mechanisms of musicality arose through a coevolutionary "virtuous spiral." While many of the specific design features above could in principle function independent of the others, and would prove adaptive independently at any proto-musical stage, over evolutionary time we hypothesize that isochronous beats coevolutionarily enabled meter and dance, and that pitched singing enabled scale-based melody and harmony. Each new feature added value in supporting extended, coordinated, harmonious group performance. Each feature may have been initially based on behavioral innovations involving synchronization of the ancestrally individualistic displays seen in other great apes (e.g. chimpanzee pant-hoot displays and fruit tree "carnival" displays, cf. Merker 1999; Merker et al. 2018). However, each innovation opened a new cognitive/musical niche selecting for independent specialization of relevant neural circuitry (see Section 4). Early instantiations of music provided selective preconditions for later cognitive and neurobiological changes underlying human musicality, analogous to the well-documented examples of gene-culture coevolution involving fire and dairy farming. Cultural innovations created a variety of proto-musical behaviors, with musical knowledge becoming a potential cue to social history and cultural (sub-)group membership. For example, this could have created selective feedback favoring individuals who used music as cues to group membership. Together, biological and cultural coevolution created a framework for the coordinated, harmonious, emotional group performances that are evident today throughout the world's musical cultures. The major inter-relationships among these components of human musicality are summarized in Figure 2 (but see Section 6.3 for caveats regarding causality in our proposed coevolutionary mechanisms). Figure 2. Proposed coevolutionary relationships among multiple musical features and mechanisms, indicating their contributions to ultimate functions by facilitating social bonding in multiple ways, their proximate neurobiological underpinnings in prediction and reward systems, and feedback loops among these different levels. 2.4 Benefits of social bonding We hypothesize that musicality increased the number of "simple" relationships (e.g. "friends"), and increased the quality (depth and complexity) of existing relationships. The opportunity for many individuals to participate productively in social interaction through proto-musical behaviors facilitates an efficient bonding mechanism for groups of varying sizes, thereby conferring associated benefits (as outlined in Section 2.1). However, we must consider the nature of the subsidiary relationships and social structures in which they operate. Many vertebrate species

live in large groups (e.g., fish schools, bird flocks, ungulate herds), but do not exhibit strong social bonds with more than a small number of individuals, and/or the relationships are undifferentiated. Indeed, the "number of differentiated relationships" (Bergman and Beehner, 2015) can vary independently from raw group size. For example, a monogamous pair with bi-parental care involves two differentiated relationships (sexual mate, and caregiving partner) or even three (adding joint territory defense), a situation typical in many birds. The social bonding design features we have identified can operate at multiple levels simultaneously, in the same way that a couple dancing at a party can intensify their own relationship, and their relationship with the broader social group. 2.5 Participatory versus presentational music For most of hominin evolution, the only way to experience music was to make it oneself, or to observe others making music in real time. But as music-making technology culturally evolved, opportunities for solo listening increased (e.g., recording technology, personal music-playing devices) and individual virtuosity became increasingly emphasized. Cross-cultural analyses suggest that forms of music-making coevolved in parallel with social structures: larger-scale, more hierarchical societies tend to emphasize "presentational" music made by small numbers of performers for large numbers of passive (or virtual) audiences. Conversely, smallerscale, more egalitarian societies tend to emphasize "participatory" music in which large groups sing, dance, and play instruments together with little or no distinction between performers and audience (Lomax 1968; Turino 2008). Once group size increases substantially, it may not be feasible for all individuals to participate actively in a coordinated manner, but music can facilitate bonding https://doi.org/10.1017/S0140525X20000333

via passive (including digital) participation. This enables music (e.g., national anthems) to help construct social identities even among massive "imagined communities" (Anderson 1991) whose members may never physically interact with one another. The participatory mode of musical performance is hypothesized to be the ancestral one that operated over long time scales. It is imperative to avoid conflating pervasive technology-driven aspects of contemporary musical practice (e.g. static audiences, solo listening, control by global corporations) with the conditions under which humans experienced music during most of our evolutionary history. As a result, testing predictions of the MSB hypothesis should favor contexts such as drumming circles, campfire singalongs and folk dances over solo-listening via headphones, or collective, static listening at a Mozart performance. Even in societies dominated by presentational music, participatory contexts retain their social and emotional potency, as highlighted by the collective singing of Italians from their balconies during the coronavirus lockdown (Grahn et al. 2020; Horowitz 2020; Kornhaber 2020). 2.6 Summary Summarizing, the MSB hypothesis argues that music is a derived bonding mechanism, akin to but augmenting previous ABMs like grooming and laughter. This augmentation occurs via the provision of a shared framework for individual participants to establish and maintain strong bonds with more than one individual (or a small group of individuals) at a time, thus bridging the "bonding gap" problem posed during human evolution by increasing group size and complexity (Dunbar 1993, 2012b). Proto-musical features may initially have arisen as behavioral innovations that later initiated a process of gene-culture coevolution. Crucially, the design features of music discussed above make music better suited than ABMs or language for coordinating behavior and facilitating social bonding in larger and more complex groups. 3. Cross-disciplinary evidence Evidence in support of the MSB hypothesis comes from cross-cultural, historical/archaeological, developmental and social psychological research. 3.1. Cross-cultural evidence One line of evidence for the MSB hypothesis comes from the study of cross-cultural musical universals (Brown & Jordania 2013; Lomax 1968; Mehr et al. 2019; Nettl 2015; Savage & Brown 2013;

Savage 2018, 2019b; Stevens & Byron 2016; Trehub et al. 2018). Music, like language, is a human universal found in all known cultures (Donald Brown 1991; Mehr et al. 2019). Few if any specific musical features are found in all known musics, just as few specific linguistic features are found in all known languages (Evans & Levinson 2009). However, researchers have identified dozens of "statistical universals" that predominate throughout diverse samples of the world's music, relating both to functional context and to musical structure (Savage et al. 2015; Mehr et al. 2019; Table 1). These cross-cultural similarities suggest selection by biological and/or cultural evolution. Table 1: Cross-culturally widespread musical structures and functions. Functional contexts were found by Mehr et al. (2019) to be associated with singing in ethnographic descriptions of the 60 societies from the Human Relations Area Files Probability Sample (Lagacé, 1979). Musical structures were found by Savage et al. (2015) to predominate (items 1-18) or to co-occur (item 19) consistently in each of nine world regions across a sample of 304 audio recordings from the Garland Encyclopedia of World Music (Nettl et al. 1998-2002). Nested relationships are indicated with indented italics; *indicates associations that were only significant using one of the two methods reported by Mehr et al. (2019)3 . 3 Mehr et al. (2019) used two methods to examine universal associations with singing: "topic annotations from the Outline of Cultural Materials ['OCM identifiers'] and automatic detection of related keywords". The second method was needed "because some hypotheses correspond only loosely to the OCM identifiers (e.g., 'love songs' is only a partial fit to ARRANGING A MARRIAGE [the OCM identifier used] and not an exact fit to any other identifier)". Similarly, "group bonding" is only a partial fit to the OCM identifier "SOCIAL RELATIONSHIPS AND GROUPS", which covers a broader range of social behaviors than simply "group bonding". After adjusting for ethnographer bias and multiple comparisons, Mehr et al. found "support from both methods for 14 of the 20 hypothesized associations between music and a behavioral context, and support from one method for the remaining six". See Mehr et al. (2019) for further details. https://doi.org/10.1017/ S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Crucial to our hypothesis, music performs similar social bonding functions across cultures. All of the 20 widespread functional contexts supported by at least one analysis in Mehr et al. (2019) summarized in Table 1 relate to social bonding, particularly through the ubiquitous use of music in communal ceremonies and rituals (e.g., healing, procession, mourning, storytelling, greeting visitors, praise/religion, weddings). Even the secular use of music as art or entertainment is itself often a form of communal ritual. For example, aspects of Western art music concert attendance function to cement social bonds between participants and exclude nonparticipants in similar ways to other elite rituals throughout history (Small 1998; Nooshin 2011). Other non-ritual contexts have social bonding functions in bringing together parents and infants (lullabies and play songs), mates (love songs), or coordinating activities among multiple individuals (work songs, dance music). Finally, regulation of moods/emotions is one of the key components of our definition of social bonding ("...synchronizing and harmonizing the moods, emotions, actions or perspectives of two or more individuals"). Even mood regulation via solo music can support social functions or evoke social contexts. For example, people may ease the pain of separation from loved ones by listening to or playing music that evokes shared memories (Kornhaber 2020), or use music to prepare their mood for an effective social interaction, allowing them to regulate their behavior and behave in the sociallyexpected manner (Erber et al. 1996; Greenwood & Long 2009). Likewise, most of the widespread structural aspects of music support coordinated music-making. Throughout the world, humans tend to sing, play percussion instruments, and dance to simple, repetitive music in groups, and this is facilitated by the widespread use of simple-integer pitch and rhythm ratios, scales based on a limited

number of discrete pitches (<=7), and isochronous beats grouped in multiples of two (Bowling et al., 2015; Ravignani et al., 2016; Jacoby & McDermott, 2017; Kuroyanagi et al., 2019; Jacoby et al., in prep.; Savage et al. 2015). The widespread use of simple, discrete meters and scales also enables multiple people to memorize and coordinate their performances. These widespread musical properties have few direct parallels in language. Group coordination provides a common purpose that unifies the cross-cultural structural regularities of human music (Savage et al. 2015). 3.2 Fossil and archaeological evidence While music itself leaves no fossil record, inferences can be drawn from evidence about the evolution of musicality, the role this played in early human society, and its relationship to other evolutionary developments such as brain size, language, group size, and sociality (Mithen, 2005; Morley, 2013). The fossil record for human evolution indicates that capacities for sophisticated and diverse vocalizations and body language, including dancing, were present before there is credible evidence for compositional language (as reviewed in Mithen 2005). Archaeological evidence from the Paleolithic indicates increasing group size and long-distance contacts (Gamble 2010; Read and van der Leeuw 2015), suggesting that ABMs had become insufficient by at least two million years ago. The earliest surviving musical instruments - bone flutes - have been dated to over 35,000 years ago and are speculated to have functioned to support larger social networks (Conard et al. 2009). Prehistoric rock art often appears to be positioned with regard to the acoustic https://doi.org/10.1017/ S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. properties of either the cave or cliff face on which it is located (e.g. Fazenda, F. et al. 2017; Rainio et al. 2018), suggesting that music played a role in the social-bonding rituals associated with that art. Similarly, prehistoric and early historic architecture used for socialbonding ceremonies often appears to have been designed with regard to its acoustic properties and to facilitate music making (e.g. Göbekli Tepe, Notroff et al. 2016; Stonehenge and other Neolithic monuments in Britain, Watson & Keating 1999; Banfied 2009; Ancient Mayan temples, Sanchez 2007). 3.3. Developmental evidence Extensive evidence demonstrating spontaneous and early development of social functions of music also supports the MSB hypothesis. Adults around the world produce infant-directed songs, such as lullabies, with similar, cross-culturally recognizable acoustic features (Trehub et al. 1993; Mehr et al. 2018). Song is highly effective at emotional modulation in infants – reliably more effective than speech, with infants exhibiting longer visual fixations and greater reductions in stress and body movement to maternal singing than to speaking (Cirelli & Trehub 2020; Corbeil et al. 2016; Ghazban 2013; Nakata & Trehub 2000; Trehub 2016). Infants also respond differently to songs sung in different styles (e.g. lullaby, vs. playsong; Cirelli et al. 2019; Rock, Trainor & Addison, 1999). Singing to infants thus appears to serve a communicative function, allowing parents to communicate specific emotional messages to infants before they can understand the semantic content of language (Rock et al. 1999; Trainor et al.1997; Trehub et al. 1997). Singing and musical interactions also directly improve parent-infant social bonds: Interventions promoting singing and musical interaction between parents and infants strengthen parents' attachment to their infants, more so than nonmusical play (Vlismas et al. 2013). Music thus facilitates both parent-infant communication and parent-infant bonding from early in life, before extensive experience or opportunity for learning. Beyond infancy, musical activities continue to promote bonding: Across a range of tasks, group musical involvement increases children's prosocial behavior. Thus, young children act more prosocially (in terms of sharing and fairness) after a musical game than a similar non-musical game (Kirschner & Tomasello 2010); after group singing than group art or competitive games (Good & Russo 2000); and after joint synchronized, rhythmic movement than non-synchronized movement (Rabinowitch & Meltzoff 2017).

Children (like adults) choose to affiliate with members of their own social group (Bigler et al. 1997). From early infancy, music serves as a marker of social group membership, allowing for the identification of preferred social partners (Cirelli et al. 2018). Shared knowledge of specific songs serves as a particularly informative signal of common group membership: due to the wide range of forms a song can take, knowledge of a particular song implies common social or cultural background (Soley & Spelke 2016). Infants accordingly treat shared musical knowledge as socially meaningful from early in life: five-month-old infants prefer to look at people who sing melodies previously sung by a parent, over people who sing melodies previously sung by an unfamiliar adult (Mehr et al. 2016). These early preferences appear to form the foundation for selective social affiliations based on music: at preschool age, children use knowledge of a familiar song as a social cue to select friends (Soley & Spelke, 2016), and by 14 months exhibit more prosocial behavior (helping) toward an unfamiliar woman who sings a familiar song (previously sung by a parent) than an unfamiliar song (Cirelli & Trehub 2018). Together, these results suggest that musical knowledge shapes the formation of children's social bonds, and that the link between shared musical knowledge and social connection is rooted in early infancy. 3.4. Social psychological evidence Behavioral experiments from social psychology support the MSB hypothesis, suggesting that musical behavior is not only associated with, but may causally support, social bonding. In particular, music provides a foundation for synchronized behavior in large groups (as argued above), and a number of experiments and meta-analyses show that rhythmic synchronization with other individuals promotes increased prosocial behavior (i.e., actions that increase others' well-being; Mogan et al. 2017; Rennung & Göritz 2016). Synchrony has been empirically linked to cooperation in economic games (Lang et al. 2017; Launay et al. 2013; Reddish et al. 2014; Wiltermuth & Heath 2009), entitativity (feelings of being on the same team; Reddish et al. 2013; Lakens & Stel 2011), rapport and interpersonal liking (Hove & Risen 2009; Miles et al. 2009; Valdesolo & Desteno 2011), and helping behavior (Cirelli et al. 2014; Valdesolo & Desteno 2011; Kokal et al. 2011). Similarly, dancing in synchrony increases participants' feelings of connectedness to the group with which they are dancing, as well as their liking and assessment of similarity with co-dancers (Tarr et al. 2016; Tarr et al. 2015). These prosocial effects of synchrony are robust in different contexts (Mogan et al. 2017). Although demand characteristics have been suggested as possible confounds underlying these effects (Rennung & Göritz 2016; Atwood et al. 2020), significant prosocial effects of synchrony remain after potential confounds of suggestion, competence and shared intention are eliminated (e.g. in https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https:// www.cambridge.org/core/terms. a virtual reality setting; Tarr et al. 2018). However, meta-analyses implied inconclusive results regarding the precise roles of "music" and of synchrony to an isochronous beat, as opposed to more generally synchronized or coordinated non-musical behaviors such as gaze synchrony, affect synchrony, and motor synchrony (Mogan et al. 2017; Rennung & Göritz 2016). In section 5, we propose clearer predictions and tests of specific mechanisms by which music promotes social bonding. More broadly, behavioral studies indicate varied social bonding effects associated with music-based activities, even those that do not explicitly involve constant synchrony. Young children randomly assigned to activities incorporating music exhibit elevated levels of empathy compared to non-musical controls in longitudinal studies (Rabinowitch et al. 2013), and adults singing in regular group sessions develop feelings of social closeness towards co-participants more quickly than people engaged in other (non-musical) group activities (Pearce et al. 2015). Feelings of inclusion, connectivity and positive affect emerge in small and large singing groups, with participants in large choirs (>80 participants) reporting greater changes in these measures

compared to smaller choirs (Weinstein et al. 2016). These findings highlight the relevance of music-based activities for large-scale social bonding. 4. Neurobiological mechanisms The MSB hypothesis proposes that social bonding is the ultimate, functional explanation of the evolution of musicality. We now propose specific hypotheses about underlying neurobiological proximate mechanisms underpinning music's social effects (Fig. 3). In brief, music involves predictable combinations of rhythms and pitches, activating neural mechanisms for perception that are tightly coupled with the motor system. Learning to form predictions about these features activates the dopaminergic reward system, which synchronizes its activity with distal regions within the brain. Crucially, predictability also supports synchronization of homologous regions in other individuals' brains. This "neural resonance" (synchronous brain activity across individuals) facilitates social bonding through shared experience, joint intentionality, and "self-other merging". Through the production of oxytocin and endogenous opioids, neural resonance also facilitates prosociality by associating the rewarding musical experience with specific co-experiencers. Furthermore, because these prosocial experiences are themselves rewarding, we seek them out by attending to and learning more musical features/experiences, updating our predictions (e.g. through statistical learning, by performing and/or experiencing new music), and closing the mechanistic cycle. This proposed mechanistic cycle is detailed below. a) b) Figure 3. a) Proposed neurobiological mechanisms underlying music's social bonding functions, showing intra- and interindividual levels. We propose that the dopaminergic reward system interacts with the Endogenous Opioid System (EOS) and the release of oxytocin, ultimately providing opportunities for individuals to synchronize their moods, emotions, actions and/or perspectives through musical engagement. (Dashed arrow indicates need for more evidence to confirm that the perception/production of music stimulates this pathway). b) Key neuroanatomical regions in the human brain underlying the MSB hypothesis. ST: superior temporal lobe structures important for auditory perception including Heschl's gyrus, planum temporale, superior temporal gyrus, superior temporal sulcus, middle temporal gyrus. Motor: frontal lobe structures crucial for action planning and execution including premotor and supplementary motor areas as well as primary motor cortex. BG: https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Basal ganglia and related structures, including amygdala, striatum, ventral tegmental area / substantia nigra, caudate, putamen, globus pallidus, nucleus accumbens. vmPFC: ventromedial prefrontal cortex. 4.1 Perception-action coupling Perception-action coupling refers to anatomical and/or functional connectivity between brain regions involved in sensory perception (e.g., of pitch or rhythm) and those that are involved in movement (e.g., vocalization, dance). Specifically, auditory-motor coupling is a key neural mechanism that underlies social bonding through music because it enables individuals to synchronize and/or harmonize their own music and actions with others, which is crucial for coordinated group music making. Even during the perception of solo music, the tight coupling between perceptual and motor regions leads to spontaneous and obligatory activity in premotor and supplementary motor areas, classic motor areas that are also part of the action observation network that drives physical and observational learning (Cross et al. 2008). Rhythm and beat consistently activate the premotor area, supplementary motor area, and basal ganglia, regions commonly thought to belong to the motor system (Grahn & Brett 2007). Furthermore, the auditory system is strongly coupled with areas in the motor system during rhythm perception (Grahn & Rowe 2009), and rhythmic oscillatory activity in both the auditory and motor systems tracks the rhythm of music (Fujioka et al. 2015). Some observations show that neural phase-locking activity is even higher in music than in speech (Vanden Bosch der Nederlanden et al. 2020). This process of "neuronal entrainment" (neural

activity changing its frequency, amplitude, and/or phase in response to external stimulation) is a proposed mechanism through which rhythm in sensory stimuli affects the brain by coordinating activity between separate neuronal populations, such as between the auditory and motor systems (Jones 2018; Morillon & Baillet 2017). This neuronal entrainment enables selective attention to specific points in time (Lakatos et al. 2008; Large & Jones 1999). In particular, auditory-motor coupling is strongest when perceiving "high groove" music that elicits the pleasurable drive towards action such as in dance (Janata et al. 2012). Groovy music elicits the urge to dance by increasing the auditory cortex's sensitivity and its coupling with the motor cortex (Stupacher et al. 2013), which is particularly evident with medium levels of rhythmic complexity and expectation violation (Witek et al. 2014; Koelsch et al. 2018). In this respect, dance - or any movement to music - is inextricably linked to musical experiences. Note, however, that like many of the mechanisms proposed here, coding of value in sensory cortices (i.e., a stronger sensory response to more important or rewarding stimuli) is not unique to the auditory domain but is also evident in other sensory domains such as vision (Koelsch et al. 2018). An important pathway underlying perception-action coupling is the arcuate fasciculus, a bundle of axonal connections between frontal lobe (including motor areas) and superior temporal lobe (including auditory areas). Abundant neuroimaging evidence supports the role of the arcuate fasciculus in music making, specifically in auditory perception-action coupling (Loui et al. 2009, 2011; Halwani et al. 2011; Moore et al. 2017; Sammler et al. 2015). This same pathway also plays a role in social functions: more emotionally empathic people have higher microstructural integrity within the arcuate fasciculus (Parkinson & Wheatley 2014). In contrast, people on the autism spectrum, who have known impairments in social bonding, have less connectivity in the arcuate fasciculus (Fletcher et al. 2010; Wan et al. 2010). By enabling perception-action coupling, the arcuate fasciculus thus provides one possible shared neural mechanism between music and social bonding. 4.2 Prediction and the dopaminergic reward system Musical perception-action coupling sets up repeated cycles of prediction, expectation violation, and resolution (Huron 2006). In these hierarchical perception-action trajectories, the predictive context surrounding pitch and rhythm are established, violated, and then resolved (Fitch et al. 2009; Clark, 2013). Successful predictions become rewarding to the brain by activating neurons of the dopaminergic system and its related areas (caudate, nucleus accumbens, amygdala, and ventromedial prefrontal cortex) that code for fundamental evolutionary rewards such as food and sex, and also learned rewards such as money (Friston 2010; Schultz et al. 1997; Knutson et al. 2000). The same dopaminergic reward system is also active during the anticipation and perception of pleasurable music (Blood & Zatorre, 2001; Blood et al. 1999; Salimpoor et al. 2011, 2015; Zatorre & Salimpoor 2013; Zatorre 2018; Cheung et al, 2019), supported by the functional coupling between auditory areas in the superior temporal lobe and reward-sensitive areas such as the nucleus accumbens (Salimpoor et al. 2013). Manipulating expectations for pitch-related musical features, such as consonance and dissonance, can modulate activity in the nucleus accumbens and amygdala. Thus, music can provide its own reward prediction error and motivate learning (Gold et al. 2019; Cheung et al, 2019). Additionally, people who frequently experience chills when listening to music show high white matter connectivity between auditory, social, and reward-processing areas (Sachs et al. 2016). Chills from music are also related specifically to increased binding to dopamine receptor D2 (Salimpoor et al., 2011). In contrast, people with musical anhedonia, who find music unrewarding, have decreased functional connectivity and altered structural connectivity between auditory and reward-related areas (Loui et al. 2017; Martínez-Molina et al. 2016; Mas-Herrero et al. 2014).

Because humans are social animals, the predictions we make and the rewards we receive are often tied to social stimuli. Thus, the brain has to learn from social cues by associating social stimuli with reward predictions (Atzil et al. 2018). Indeed, the same areas in the dopaminergic reward system – the caudate, nucleus accumbens, and ventromedial prefrontal cortex – are causally linked to cooperative behavior as well as prediction and reward. The reward system is activated when we share information with others about ourselves (Tamir et al, 2012), when we view loved ones (Bartels and Zeki, 2004), and when mothers bond with their infants (Atzil et al. 2017). Prosocial behaviors commonly engage the reward system (Zaki & Mitchell 2013); these include cooperating (Decety et al. 2004), perspective taking (Mitchell et al. 2005), and empathizing with others (Beadle et al. 2018). Together, these results suggest that the dopaminergic reward system is involved causally in the link between music and social bonding through the mechanism of prediction. 4.3 Oxytocin and the Endogenous Opioid System (EOS) We propose that opioids released in the EOS, and oxytocin, are also part of the mechanistic underpinnings linking prediction, reward, and social bonding (Chanda & Levitin 2013; Launay et al. 2016; Tarr et al. 2014). The nucleus accumbens and ventral tegmental area are key regions that overlap between the dopaminergic reward system and the EOS (Le Merrer et al. 2009; Dölen et al. 2013), and dopamine is thought to be a salience processing mechanism regulated by oxytocin (Love 2014; Shamay-Tsoory & Abu-Akel 2016). The EOS likely plays a mechanistic role in music-related prosociality. This system has been implicated in the maintenance of social bonds in primate social networks (Keverne et al. 1989; Maestripieri 2010; Ragen et al. 2013; Schino & Troisi 1992). Intervention studies in humans indicate that, compared to a placebo, naltrexone (an opiod blocker) can reduce feelings of social connections with others (e.g. Inagaki 2018, Inagaki et al. 2016), and lower affiliative behaviour and desire for interpersonal closeness (Tchalova & MacDonald 2019). Listening to music influences mu-opiate receptor expression in the EOS (Stefano et al. 2004) and can reduce the need for pain medication4 (e.g. Lepage et al. 2001; Bernatzky et al. 2011). Elevated pain thresholds are experienced after singing (Pearce et al. 2015; Weinstein et al. 2016) and synchronized dancing (Tarr et al. 2015; Tarr et al. 2016), but not after administration of naltrexone (Tarr et al. 2017), suggesting that pain threshold is an appropriate proxy-measure of endorphin uptake in these experiments. There is some evidence of endorphin-mediated synchrony effects on cooperation (e.g. when dancing; Lang et al. 2017), further demonstrating links between music, the EOS, and social bonding. Although more empirical work is needed, there is evidence that oxytocin levels are elevated after taking part in a singing class (Grape et al. 2003), or following a group jam session of improvised singing (Keeler et al. 2015). Elevated oxytocin levels have been correlated with increased generosity (Zak et al. 2007, Fujii et al. 2016), empathy (Domes et al. 2007, Hurlemann et al. 2010), and possibly trust (Kosfeld et al. 2005; Zak et al. 2005, but see Nave et al. 2015 and Declerck et al. 2020). Furthermore, intranasal administration of oxytocin promotes in-group cooperation (e.g., De Dreu & Kret 2016) and increases synchrony in dancing (Josef et al. 2019) and finger-tapping behavior (Gebauer et al. 2016), suggesting a reciprocal feedback loop between music-based activity and social cohesion. Although evidence linking oxytocin specifically with music remains limited, and the strength of oxytocin's relationship with cooperation more generally is debated (particularly studies based on administering intranasal oxytocin; e.g. Walum et al. 2016), current evidence suggests that music engages the oxytocin and EOS systems in ways that facilitate social bonding, as predicted by the MSB hypothesis. Combined with the reward system, these pathways offer a positive-feedback loop following music engagement, enabling groups of individuals to synchronize their moods, emotions, actions and/or perspectives, and providing motivation to continue engaging with others in social and musical contexts. 4.4 Learning and vocal imitation The capacity to learn and reproduce complex motor movements, including

vocalizations (songs), is central to the cultural transmission of music. Although humans are the only primates capable of learning complex, novel vocalizations, this ability has evolved 4 Endogenous opioids (e.g. beta-endorphins) likely interact with other systems in analgesic effects (e.g., Welch & Eads 1999). Nevertheless, elevated pain threshold is a common proxy-measure of elevated beta-endorphin levels, due to the EOS's role in our pain-pleasure circuitry (Mueller et al. 2010), and the fact that direct measures are invasive and expensive.

independently at least seven times in evolutionary history (Fitch & Jarvis 2013; Nowicki & Searcy 2014; Syal & Finlay 2011), allowing us to make inferences about how and why it evolved. Some vocal learning clades (seals, baleen whales, and some songbirds) show a strong male bias in vocal learning abilities consistent with sexual selection. However, such a bias is absent in most other vocal learners (parrots, elephants, toothed whales, many tropical bird species, and humans), suggesting that sexual selection cannot be the only factor driving the evolution of vocal learning (Fitch 2006). Instead, learned animal songs (solo or duet) appear to serve multiple evolutionary functions within the umbrella of social bonding, including mate attraction, cementing and affirming social bonds within pairs or groups, and territorial functions including advertising the bonded group's ability to repel outsiders (Haimoff 1986; Wickler 1980; Geissmann 1999). In vocal learning species, vocal imitation and song production are likely based on similar neurobiological mechanisms (Mercado et al. 2014). Learning to reproduce pitches and rhythms accurately engages reward mechanisms, as shown by evidence that dopamine neurons encode performance error in songbirds (Gadagkar et al. 2016). Furthermore, the presence of a conspecific (of the opposite sex in this case) leads the male zebra finch to decrease variability of sung syllables; this syllabic structure is attributed to perception-action circuits analogous to the human superior temporal and motor structures (Sakata & Brainard 2008; Fitch & Jarvis 2013). Once individuals learn to produce musical features, they not only reproduce learned patterns of features, but also deviate from predicted combinations of features, for example by inventing new melodies (Wiggins et al. 2018). 5. Predictions for future research The MSB hypothesis predicts that core design features of music make it particularly well-suited to facilitate social bonding, and particularly effective in the bonding of large, complex groups. This leads to the following testable predictions: 5.1. Cross-domain predictions (music, language, ritual, etc.) The MSB hypothesis predicts that music (including dance) is better-suited to social bonding of large, complex groups than ABMs (grooming, laughter), language, or other non-acoustic bonding mechanisms like shared decorations or non-musical ritual behaviors (e.g., praying together without music). Music should be more effective and/or efficient relative to other methods as group size and complexity increase, such that while making music in pairs might only produce a small increase in dyadic bonding relative to conversation, making music in larger, more complex groups of people (dozens or hundreds organized into differentiated sub-groups) should be more effective for collective bonding than language, laughter, grooming, etc. In a social species like humans, many activities can develop and enhance social bonding, but we predict that bonding via non-musical methods like language, ritual, or sports should be enhanced by the addition of musical components (e.g., religious services with group singing will result in stronger bonding than those that only involve group prayer). Different musical components are predicted to have synergistic effects such that - all things being equal - including more of these components (e.g., synchronized, harmonized singing and dancing in groups) will tend to increase bonding more than activities that only use one or a few (e.g., conversations or recitation in pairs)5 . We also predict that participatory musical performances will have significantly stronger effects than either non-participatory (e.g., performance for a static

audience) or solo musical experiences (e.g., listening alone to recordings). Group size and complexity should have independent effects (e.g., singing in large choirs should produce greater bonding than singing in small choirs). These predictions can be tested in controlled experiments and/or field studies along the lines of those discussed in section 3. Designing studies that control for specific similarities and differences between closely related domains such as music, language, and dance is challenging but not impossible. For example, to control for the fact that languages have their own (non-isochronous) rhythms, Savage et al. (2020) had groups of participants simultaneously recite the lyrics to "Twinkle, Twinkle, Little Star" to an isochronous beat or in non-isochronous free rhythm. Savage et al. (2020) also propose additional manipulations that would allow this paradigm to test other specific predictions of the MSB hypothesis regarding the social bonding effects of melody, harmony, and dance (cf. Fig. 3 in Savage et al. 2020). 5 However, other factors (e.g., ceiling effects, optimal degrees of complexity, rhythm-melody interactions [Prince et al. 2009]) may limit social bonding effects, leading to non-linear interactions when combining multiple musical components.

5.2. Cross-cultural predictions The MSB hypothesis predicts that music's social bonding functions should be distributed widely in space and time. Hence, the kinds of predictions described in section 5.1 regarding music's superior social bonding power in large groups should apply consistently across cultures and throughout history. Furthermore, it predicts that musical contexts and structures that promote social bonding (e.g., coordinated, participatory group performances) will be more common across cultures than music produced by and for individuals. At the same time, the relative importance of participatory vs. presentational music-making is predicted to vary cross-culturally as a function of social structure (due to limitations on simultaneous coordinated performance discussed in section 2.5). Smaller-scale, more egalitarian cultures should thus perform and value participatory music more than larger-scale, hierarchical cultures where presentational music should be more common and valued. Participatory vs. presentational distinctions are analogous to those found in "imagistic" (high-intensity, small-scale) vs. "doctrinal" (low-intensity, large-scale) religious rituals, respectively (Whitehouse 2004), and are predicted to covary cross-culturally with these modes of religiosity. Even in cultures where music is often consumed passively by individuals (e.g. in Western culture, over headphones on personal listening devices), MSB predicts that music will be more effective than non-musical alternatives for social bonding purposes (cf. Rentfrow & Gosling 2006). These predictions about crosscultural use of music for social bonding could be tested in cross-cultural behavioral experiments (cf. Henrich et al. 2006; Polak et al. 2018; Jacoby et al. in prep) or analysis of cross-cultural databases of recordings, artefacts, ethnographies, or questionnaires (cf. Lomax 1968; Mehr et al. 2019; Savage et al. 2015; Savage 2019c; Whitehouse et al. 2019; Wood et al. In prep). 5.3. Cross-species predictions The MSB hypothesis proposes that human musicality has been shaped by biological and cultural selection, and that the features of music are particularly well-suited for social bonding functions because they support extended, coordinated group performances on a large scale. The MSB account does not claim that music's social bonding function is a unique biological adaptation specific to human musicality. Instead, it argues that music-like behaviors should enhance existing bonding mechanisms in other species as well. Thus, it predicts that, rather than an all-or-nothing divide between human and non-human "music", species will vary continuously in the degree to which they share specific features of human musicality. The social bonding functions associated with different components of musicality should operate similarly across species, depending on the specific subcomponent, its suitability for group coodination, and the importance of social bonding to their species. Thus,

melodic, learned song among songbirds, whales, or other vocal learners are predicted to enhance social bonding in these species in a manner analogous to song in humans. These effects may be limited in many non-human species by their lack of ability and/or interest in performing in coordinated groups (e.g., some primates appear motivated to conduct group displays but are unable to synchronize to a beat, while some birds appear able to move to a beat but are unmotivated to do so in groups in the wild; Hoeschele et al. 2018). However, such effects should be pronounced in species that perform duets (e.g. many birds, and duetting primates such as gibbons or titi monkeys; Hall 2004; Haimoff 1986). Conversely, social primates that do not typically perform in coordinated groups may nonetheless experience social bonding effects of "group" music when exposed to versions of their own vocalizations that have been artificially manipulated to be in synchrony/harmony. Such production/perception dissociations and other nuances of musicality could be tested in controlled cross-species experiments (cf. Hoeschele et al. 2018; Merchant et al. 2018). The MSB hypothesis posits that music and musicality provided a major means by which humans could coordinate behavior on a larger scale than dyads or small groups, allowing for the formation of larger socio-cultural groups. If true, and if different species share components of musicality to differing degrees, then across species, production or proficiency in "musical" behaviors should predict both the number and complexity of social bonds. For example, gelada baboons live in unusually large and complex groups for primates, and they also exhibit rhythmic and melodic vocal features that are unique among primates (Richman 1978, 1987; Gustison et al. 2012; Bergman 2013). Like geladas, many parrot species live in large fission-fusion social groups, and members of the parrot clade show vocal imitation, call convergence, duetting, and the capacity for rhythmic synchronization (Bradbury 2001; Balsby & Scarl 2008; Scarl & Bradbury 2009; Schachner et al. 2009). In both of these clades, pairs or mating "harems" form stronger bonds than those they share with the larger groups in which they are embedded (cf. Wanker et al. 2005; Balsby & Scarl 2008). Other species that live in complex fission-fusion groups and could provide evidence of specific design features are elephants and some odontocetes (e.g., orcas, bottlenose dolphins). Such species live in large, complex fission-fusion groups, and are documented vocal learners, but their possession of other design features of music (e.g., synchronization) have not been tested rigorously. For many species, evidence for design features of musicality would count as evidence against our hypotheses. Examples include solitary species (e.g. many reptiles), for whom groups consist only of mothers and dependent young (e.g. many carnivores), or group living species that do not have differentiated social bonds with other group members (e.g. schooling fish, larger herds, swarming insects). The MSB hypothesis further predicts that if a species does not follow this pattern (e.g. by having a larger social group size than predicted by their features of musicality), then that species will have evolved other non-musical but effective means of coordinating behavior that likely do not appear in human behavior (e.g., reproductive suppression in naked mole rats or pheromonal queen control in eusocial insects; Alaux et al. 2010; Dengler-Crish & Catania 2007). Thus, although the social bonding design features seen in human musicality are not the only way to achieve large, well-bonded groups, they are effective enough that we predict them to evolve convergently (cf. Fitch 2006). 5.4. Neurobiological predictions The MSB hypothesis predicts that each of the mechanistic factors proposed above (Fig. 3) contributes to the effects of music on social bonding. Alterations of these mechanistic pathways should therefore produce specific, quantifiable results on bonding. For example, music's perceived social bonding functions should correlate with oxytocin/EOS production, and disrupting the oxytocin/EOS pathway via blocking

oxytocin or opioid receptors should disrupt its social bonding effects. Furthermore, since the dopaminergic reward system is at the center of prediction for musical features, populations with deficient dopaminergic activity may have impaired predictions, which could affect their ability to synchronize or harmonize with others. On the other hand, drugs that restore dopaminergic functions are hypothesized to restore these abilities, and due to the reciprocal nature of these interactions, activities that enhance predictions (such as dancing and harmonizing) may in turn restore dopaminergic functions. These predictions are being tested in the case of Parkinson's Disease, which is a special population with deficient dopaminergic activity (Grahn et al. 2009; Cameron et al. 2016). Another prediction is that special populations with high sociability may respond well to musical features especially when coupled with social stimuli, as in the case of children with Williams Syndrome (Lense et al. 2014; Jarvinen-Pasley et al. 2010). At a neural level, music's social bonding function should correlate with the degree of neural connectivity between the perception-action and predictionreward networks, and disruptions to this network (e.g., lesions or genetic syndromes) should accordingly disrupt music's social bonding effect. For example, people with musical anhedonia, who have disrupted connectivity between auditory prediction and reward networks (Belfi & Loui 2019), are predicted to have weaker social bonds, and genetic differences (e.g. in DRD2) may predict variation in bonding experienced through musical activities. Although some of these predictions may be difficult to test ethically in humans through controlled experiments, many can be tested using neuroimaging combined with neuropsychological testing in special populations, as well as correlational, longitudinal, or intervention (including brain-stimulation) studies, genome-wide association studies, and/or animal models that share specific neurobiological endophenotypes (Finlay et al. 2001; Fitch & Javis 2013; Gingras et al. 2018; Hoeschele et al. 2018; Niarchou et al. 2019)6 . 6. Potential criticisms Having detailed our social bonding hypothesis and its predictions, we wish to preempt several potential criticisms. 6.1 Music, language, and domain-specificity The key criticism that we anticipate regards the degree to which the evolution of musicality and social bonding are uniquely and causally linked. Few would deny that music can facilitate social bonding via neurobiological mechanisms that are evolutionarily adaptive. However, whether music is a domain-specific evolutionary adaptation for social bonding, as opposed to a byproduct of the evolution of other adaptations, is open to debate. Language, in particular, has been proposed as an evolutionary adaptation that led to musicality as a byproduct (Pinker 1997; Patel 2008)7 . Importantly, many researchers have noted that, while there are clear differences 6 We have refrained from making detailed predictions about genetic bases of musicality because our current state of knowledge is limited (Gingras et al. 2018). However, we are hopeful that new findings from initiatives such as the 2019 symposium on "Deciphering the biology of human musicality through state-of-the-art genomics" (http:// www.mcg.uva.nl/musicality2019) will enable researchers to investigate genetic dimensions of the geneculture coevolutionary mechanisms we describe. 7 Others view language evolution as being driven largely by cultural evolution (e.g. Jablonka et al. 2012; Kirby 2017). https://doi.org/10.1017/S0140525X20000333 Downloaded from https:// www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. in the structure and processing of music and language, there is extensive overlap ranging from structural content (e.g., "musilinguistic continua" between speech and song including intermediate forms like poetry and chant) to neurobiological substrates (e.g., similar neural substrates for processing of pitch, rhythm, and syntax; Brown 2000b, 2017; Fitch 2006; Patel 2008; Peretz & Coltheart 2003; Peretz et al. 2018; Savage et al. 2012). Indeed, many have proposed that the evolution of musicality may have paved the way for the evolution of language (Darwin 1871; Brown 2000b, 2017; Fitch 2010; Mithen 2005; Shilton et al. 2020).

We accept that our present level of understanding is insufficient to demonstrate conclusively that music coevolved uniquely with social bonding independent from language or other social behaviors. Accordingly, in Section 5 we proposed future investigations of such relationships. However, the fact that music and language are both found universally in all known societies (Donald Brown, 1991; Mehr et al. 2019) suggests that both music and language independently fulfill more fundamental adaptive functions than technologies or cultural artifacts that are not cross-culturally universal. We make no claim that the mechanisms discussed here are entirely specific to music, or that "musicality" is modular in either the cognitive or neuroscientific senses of this term. For example, prediction and predictive coding are ubiquitous features of vertebrate brains (Schultz & Dickinson 2000; Clark 2013), by no means specific to musicality. However, music affords a uniquely effective scaffolding framework, including rhythm and harmony, within which neural prediction (and occasional expectation violations) can unfold (Hanslick 1858; Huron 2006; Fitch et al. 2009; Koelsch et al. 2019). Similarly, synchrony is widespread in human sociality (including phenomena such as gaze synchrony, affect synchrony, the chameleon effect, and others), but the isochrony of musical rhythm provides an unusually effective affordance for synchronization. Furthermore, phenomena like "groove" seem to be mainly evoked by musical stimuli, and therefore are relatively domain-specific. Thus, musicality encompasses multiple mechanisms that vary in their domain-specificity, but combines them into a uniquely effective package. 6.2 Group selection Most previous social bonding theories of music evolution have relied on an evolutionary mechanism incorporating some form of group selection, in which genetic variants are selected for due to their effects on the reproductive success of entire groups (e.g., Brown 2000a, Wiltermuth & Heath 2009). Group selection was largely dismissed for decades (Williams, 1966), and while it is re-emerging in the form of multi-level selection (Wilson & Wilson 2007; Traulsen & Nowak 2006) and cultural group selection (Richerson et al. 2016), it remains controversial (Pinker, 2012; see also commentary accompanying Richerson et al. 2016). The MSB hypothesis does NOT require group selection (any more than grooming, play, or laughter do): fitness advantages accrue to individuals who are able to bond more effectively with larger numbers of individuals. Although there are often advantages to wellbonded groups for various activities (e.g. group hunting or foraging, jointly repelling enemies), even for such activities the key fitness advantages accrue to individuals. 6.3 Gene-culture coevolution and causality Some evolutionary psychologists have been critical of social bonding theories of music evolution because they consider them circular arguments that fail to explain the ultimate causal mechanism by which music could have evolved as a biological adaptation: Perhaps singing lullabies soothes babies; perhaps dancing relieves tension; perhaps shared stories bond the community. The question is, why would anyone have predicted, a priori, that people would be constituted in such a way that these things would happen? (Pinker 2007, pp. 170-171) Several have posited an adaptive function for music in enhancing "cohesion" or "bonding".… But this reasoning is circular: it takes as a given the fact that music performance and listening produces reliable effects… and then argues that one or more parts of the music faculty evolved in order to produce these effects. But why should music produce these effects and not others?....accounts invoking cohesion and/or bonding as an adaptive target provide neither a specific account of the ultimate functional mechanism by which music should increase cohesion, nor an account of how that cohesion would produce fitness advantages. And if cohesion is indeed fitness enhancing, why should individuals wait for music-making to produce that cohesion? Why not just be cohesive without music? (Mehr & Krasnow 2017, p.676) https://doi.org/10.1017/S0140525X20000333

Music does not directly cause social cohesion: rather, it

signals existing social cohesion that was obtained by other means (Mehr, Krasnow, Bryant, and Hagen This issue, p. ?? [emphasis in original]; paraphrasing Hagen & Brant 2003, p. 30) Our preceding account provides a priori arguments detailing why and how specific design features of human musicality have social bonding effects, the mechanisms underlying these effects, and how and why these may have evolved. In particular, we provided specific reasons that behaviors with the design features of music would have social bonding effects: because such behaviors allow people to predict, synchronize, share goals, distinguish individual contributions, experience shared positive emotions, and make social decisions more than other human behaviors (ABMs or language). This explains why music should produce "[social bonding] effects and not others": behaviors that allow us to align in time and frequency, coordinate behaviors in large groups while distinguishing individual contributions, share emotions and goals, and choose appropriate social partners have tangible and predictable social bonding effects. Music is a particularly effective cognitive "technology" (Patel 2008, 2018) that fulfills these design criteria, making musicality an effective toolkit for social bonding functions, shaped by both biological and cultural evolution. Our hypothesis differs from most traditional social bonding theories because we do not argue that musicality necessarily originated as a biological adaptation. Instead, components of musicality may have arisen initially as cultural inventions and/or byproducts of other adaptations, later exapted and modified through gene-culture coevolution for their social bonding functions in a musical context (e.g., beat synchronization initially as a byproduct of the evolution of vocal learning, as argued by Patel et al. 2009 and Schachner et al. 2009, although cf. Merker et al. 2018 for an alternative interpretation). The initial social cohesion functions may not have begun as genetic adaptations. In this sense, we largely agree with Mehr et al., who write: We also agree with proponents of the social bonding hypothesis that musical abilities evolved because musical performances played an important role in cooperative sociality. But given the issues described above, we find it more likely that music evolved to credibly signal decisions to cooperate that were already reached by other means, not to determine them.. (Mehr et al. This issue, p. ??) But in a social environment in which social bonding already enhanced individual reproductive fitness, the subsequent cultural evolution of musical behaviors would lead to biological selection on musicality (e.g. to promote motivation to engage in/ attend to musical behaviors), because of the adaptive consequences of musicality for social bonding. In this way, just as social bonding is crucial in most primates, generating selection on the mechanisms that achieve it, social bonding functions of "proto-musical" mechanisms may have played important roles in hominin evolution long before today's full-blown musicality evolved. We emphasize that past adaptive function, while important, should not be the sole criterion by which to judge theories of the evolution of musicality. As previously argued at length (e.g., Fitch 2006, 2015b; Honing et al. 2015), Tinbergen's (1963) multi-factorial perspective, which seeks understanding of traits at the four interlinked explanatory levels of mechanism, ontogeny, phylogeny and adaptive function, is a fruitful method for understanding the evolution of musicality. We may never know with certainty the precise ancestral adaptive conditions or specific genetic mutations involved in the evolution of musicality. Even so, the comparative method provides a key tool for empirically testing evolutionary hypotheses (Fitch 2015b). Section 5 lists a variety of testable empirical predictions of the MSB hypothesis. 6.4. Parochial altruism and outgroup exclusion Enhanced social bonding between some individuals inevitably means a relative decrease between others. In-group social bonding has a dark side of increasing hostility towards outgroups (Whitehouse 2018; Gelfand et al. 2020), as exemplified in the use of music in warfare by the Nazis and other groups throughout history (Brown & Volgsten 2006). The traditional Maori haka "Ka Mate" is famously used by New Zealand's national rugby team to simultaneously bind team-mates together and

intimidate the opposing team through coordinated dancing and vocalization (Jackson & Hokowhitu 2002). The ability of music to exclude out-group members might appear to be an argument against its function in bonding in-group members, but outgroup exclusion is entirely consistent with the social bonding hypothesis. Because the creation or strengthening of a social bond between some (participating) individuals by definition excludes others, the observation that particular forms of music can cause emotional dissonance or fear in others is compatible with a social bonding function. Earlier expositions of the social bonding hypothesis (Brown 2000a; Freeman 2000) noted that "bonding is always exclusionary" and "individuals who do not 'belong' become enemies... The process is similar to sexual jealousy, which manifests the exclusionary nature of the pair bond" (Freeman 2000:421-2). This observation is mirrored in the recent literature on oxytocin which, far from being an indiscriminate "love drug," simultaneously exerts affiliative effects among in-group members and exclusionary effects towards out group individuals (cf. Beery 2015; Shamay-Tsoory & Abu-Akel 2016). The use of music to exclude others is no argument against its social bonding origins. 6.5. Solo music, sexual selection, and individual signaling While coordinated group performances predominate throughout the world, various widespread musical genres are not necessarily performed in coordinated groups. In particular, lullabies and love songs are found throughout the world and are often performed by a lone singer (Trehub et al. 1993; Mehr et al. 2018). This is perfectly consistent with the MSB hypothesis, as lullabies and love songs are often dyadic: sung by a soloist to bond with another person (by soothing an infant or wooing a potential mate). More generally, some may wonder if social bonding is so important to the evolution of musicality, why do people enjoy playing or listening to music alone? We emphasize that even solo music listening can support social bonding goals (Trehub et al. 2018). A young adult meeting a new person in an online chat discusses music preferences more often than other topics, and based on music preferences alone, people draw social inferences about others (Rentfrow & Gosling 2006). Thus, music preferences developed during solo listening can be used as social cues, displayed and evaluated when establishing new social bonds. Solo listening may serve other, non-social functions (e.g. mood regulation, staying awake while driving; DeNora 2000; North et al. 2004; Sloboda et al. 2001). We do not argue that social bonding is the only possible function of music. By analogy, language's primary function may be to communicate information between people, but it is also useful in private thought, or to allow one to preserve thoughts for the future (particularly after the invention of writing). Likewise, the same auditory-motor-reward connections that make music so socially powerful also allow people to enjoy playing or listening to music alone. Often, solo music was experienced previously in a social context, which is re-evoked by solo listening/playing. Related to the idea of virtuosic solo music-making is the distinction between social bonding and theories such as sexual selection or honest signaling that emphasize music as a signal of individual fitness. The MSB hypothesis does not reject such theories. Instead, it emphasizes that individual signaling theories are insufficient to explain all of the broader social functions of music, while social bonding provides more explanatory power (although we concede that the MSB hypothesis cannot explain all possible functions of music; Oesch 2019). For example, in contemporary Western night clubs and traditional non-Western societies, all-night music and dance rituals function both to bond participants and as opportunities to find potential mates (Merriam 1964; Thornton 1995). In such contexts, dancing, singing, and/or playing instruments can function to bond with same and opposite-sex partners and to advertise evolutionary fitness to potential mates. Bonding and signaling hypotheses are not mutually exclusive, but rather

complementary. The complementarity of the MSB and alternative hypotheses makes it challenging to falsify the MSB hypothesis. However, we have provided a number of specific predictions, each of which is potentially falsifiable and would count as evidence against the MSB hypothesis, particularly if alternative hypotheses better predict the data. For example, our hypothesis and Hagen & Bryant's (2003) coalitional signaling hypothesis make predictions regarding synchrony: we argue that synchrony should enhance social bonding, while Hagen & Bryant argue that synchrony should enhance perceived coalitional quality. To differentiate between these and other competing hypotheses, our predictions regarding the effects of synchrony (or other aspects of musicality) on social bonding could be compared directly against perceived coalition quality or other competing predictions (e.g., attractiveness; Miller 2000, parental investment; Mehr & Krasnow 2017; Mehr et al. this issue) in future research. If synchrony increases perceived bonding relative to perceived coalition quality, attractiveness, or parental investment, it would constitute evidence favoring the MSB hypothesis over competing alternatives. Another example of predictions that differentiate among alternative hypotheses is the MSB predictions that social bonding functions will be common cross-culturally but the relative frequencies of specific genres and sub-functions (e.g., lullabies vs. love songs vs. group dancing) will vary across societies. In contrast, theories that focus on infant-directed song or sexual selection predict instead that these categories should be more common and consistent cross-culturally than the other categories of social bonding. Furthermore, phylogenetic or other cross-species analyses (e.g., Shultz et al. 2011; Hoeschele et al. 2018; Schruth et al. In press) could allow us to quantify the relative effects of group size, sexual competition, parental investment strategies, or other factors on the evolution of vocal learning, beat perception, or other aspects of musicality. We encourage tests of MSB predictions against those of competing hypotheses. 7. Conclusion Social bonding has long been acknowledged as an important function of contemporary music, but its causal role in the evolution of music has often been dismissed as a naïve application of group selection theory. Recent advances in gene-culture coevolution theory allow us to provide a more nuanced model of music evolution that does not rely on group selection. Our argument has focused on social bonding as the primary factor shaping the evolution of human musicality. This MSB hypothesis provides a framework for understanding the past evolution of musicality, and a starting point for the future cultural evolution of new forms of music that harness the social power of music to bring people together. Music may not be a "universal language" (Longfellow 1835; Savage 2019b), but music's universal power to bring people together across barriers of language, age, gender, and culture sheds light on its biological and cultural origins, and provides humanity with a set of tools to create a more harmonious future - both literally and figuratively.

References: Alaux, C., Maisonnasse, A., & Le Conte, Y. (2010). Pheromones in a superorganism: From gene to social regulation. Vitamins & Hormones 83: 401–23. Anderson, B. R. O. (1991). Imagined communities: Reflections on the origin and spread of nationalism. New York: Verso. Arbib, M. A. (2005). "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics," Behavioral and Brain Sciences 28: 105–67. Arom, S. (1991). African polyphony and polyrhythm: Musical structure and methodology. Cambridge: Cambridge University Press. Atwood, S., Schachner, A., & Mehr, S. A. (2020). Expectancy effects threaten the inferential validity of synchrony-prosociality research. PsyArXiv preprint. https://doi.org/10.31234/osf.io/zjy8u Atzil, S., Touroutoglou, A., Rudy, T., Salcedo, S., Feldman, R., Hooker, J. M., ... & Barrett, L. F. (2017). Dopamine in the medial amygdala network mediates human bonding. Proceedings of the National Academy of Sciences USA 114(9): 2361-66. Atzil, S., Gao, W., Fradkin, I., & Barrett, L. F. (2018). Growing a social brain. Nature Human Behaviour 2(9): 624–36. Baldwin, J. M. (1896). A new factor in evolution. The American Naturalist 30(354): 441–51. Balsby, T. J. S., & Scarl, J. C. (2008). Sex-specific responses to vocal convergence and divergence of contact calls in orange-fronted conures (Aratinga canicularis). Proceedings of the Royal Society B: Biological Sciences 275: 2147-54. Bateson, P. (2004). The active role of behaviour in evolution. Biology and Philosophy 19: 283–98. Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. Neuroimage 21(3): 1155- 66. Beadle, J. N., Paradiso, S., & Tranel, D. (2018). Ventromedial prefrontal cortex is critical for helping others who are suffering. Frontiers in Neurology 9: article 288. http://doi.org/10.3389/fneur.2018.00288 Beery, A. K. (2015). Antisocial oxytocin: Complex effects on social behavior. Current Opinion in Behavioral Sciences, 6: 174–182. https://doi.org/10.1016/j.cobeha.2015.11.006 Belfi, A. M., & Loui, P. (2020). Musical anhedonia and rewards of music listening: current advances and a proposed model. Annals of the New York Academy of Sciences, 1464: 99–114. Benzon, W. (2001). Beethoven's anvil, music in mind and culture. Berme: Basic Books. Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. Current Biology 23: 268-9. Bergman, T. J., & Beehner, J. C. (2015). Measuring social

complexity. Animal Behaviour 103: 203-9. Bernatzky, G., Presch, M., Anderson, M., & Panksepp, J. (2011). Emotional foundations of music as a nonpharmacological pain management tool in modern medicine. Neuroscience & Biobehavioral Reviews 35(9): 1989-99. Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. Child Development, 68(3): 530–43. Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences USA 98(20): 11818–23. Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nature Neuroscience 2(4): 382– 87. Bloom, P. (2002). How children learn the meanings of words. MIT press. Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. Proceedings of the National Academy of Sciences USA 112(36): 11155–60. https://doi.org/10.1017/S0140525X20000333

Bowling, D. L., Purves, D., & Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. Proceedings of the National Academy of Sciences USA 115(1): 216–21. Bowling, D. L., Hoeschele, M., Gill, K. Z., & Fitch, W. T. (2017). The nature and nurture of musical consonance. Music Perception, 35(1), 118–121. Boyd, R., & Richerson, P. J. (1985). Culture and the evolutionary process. Chicago: University of Chicago Press. Bradbury, J. (2001). Vocal communication of wild parrots. Journal of the Acoustical Society of America. 115: 2373. Brown, Danielle (2020, June 12). An open letter on racism in music studies: Especially ethnomusicology and music education. My People Tell Stories. https://www.mypeopletellstories.com/blog/open-letter Brown, Donald E. (1991). Human universals. Philadelphia: Temple University Press. Brown, S. (2000a). Evolutionary models of music: From sexual selection to group selection. In F. Tonneau & N. S. Thompson (Eds.), Perspectives in ethology, vol. 13: Evolution, culture, and behavior (Vol. 13, pp. 231–281). New York: Plenum Publishers. Brown, S. (2000b). The "musilanguage" model of musical evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), The origins of music (pp. 271–300). Cambridge: MIT Press. Brown, S. (2007). Contagious heterophony: A new theory about the origins of music. Musicae Scientiae, 11(1): 3–26. Brown, S., & Jordania, J. (2013). Universals in the world's musics. Psychology of Music 41(2): 229–48. Brown, S., & Volgsten, U. (2006). Music and manipulation: On the social uses and social control of music. New York: Berghahn Books. Cavalli-Sforza, L. L., & Feldman, M. W. (1981). Cultural transmission and evolution: A quantitative approach. Princeton: Princeton University Press. Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. Trends in Cognitive Sciences 17(4): 179–93. Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. Current Biology, 29(23): 4084–92. Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. Developmental Science 17(6): 1003–11. Cirelli, L. K., & Trehub, S. E. (2018). Infants help singers of familiar songs. Music & Science, 1: 205920431876162. https://doi.org/10.1177/2059204318761622 Cirelli, L. K., Trehub, S. E., & Trainor, L. J. (2018). Rhythm and melody as social signals for infants. Annals of the New York Academy of Sciences, 1423(1), 66-72. Cirelli, L. K., & Trehub, S. E. (2020). Familiar songs reduce infant distress. Developmental Psychology, 56(5), 861–868. Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2019). Effects of maternal singing style on mother–infant arousal and behavior. Journal of Cognitive Neuroscience. 32(19): 1213–1220 Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences,

36(3): 181–253. Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. Nature 460(7256): 737–40. Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. Child Development 61(5): 1584-95. Corbeil, M., Trehub, S. E., & Peretz, I. (2016). Singing delays the onset of infant distress. Infancy 21(3): 373-91. 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: Exploring the basis of human companionship (pp. 61–82). Oxford University Press. https://doi.org/10.1017/S0140525X20000333
Cross, E. S., Kraemer, D. J., Hamilton, A. F. D. C., Kelley, W. M., & Grafton, S. T. (2008). Sensitivity of the action observation network to physical and observational learning. Cerebral Cortex 19(2): 315-26. Darwin, C. (1871). The descent of man, and selection in relation to sex. London: John Murray. Davila Ross, M., Owren, M. J., & Zimmerman, E. (2009). Reconstructing the evolution of laughter in great apes and humans. Current Biology 19: 1106-11. Declerck, C. H., Boone, C., Pauwels, L., Vogt, B., & Fehr, E. (2020). A registered replication study on oxytocin and trust. Nature Human Behaviour. https://doi.org/10.1038/s41562-020-0878-x Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. Neuroimage 23(2): 744-51. De Dreu, C.K.W., & Kret, M.E. (2016) Oxytocin conditions intergroup relations through upregulated ingroup empath, cooperation, conformity, and defense. Biological Psychiatry. 79(3):165–73. Dengler-Crish, C. M., and Catania, K. C. (2007). Phenotypic plasticity in female naked mole-rats after removal from reproductive suppression. Journal of Experimental Biology 210: 4351-58. DeNora, T. (2000). Music in everyday life. Cambridge University Press. Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. Journal of the Acoustical Society of America 129: 2245-52. Dissanayake, E. (2000). Antecedents of the temporal arts in early mother–infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), The origins of music (pp. 389–410). Cambridge, MA: MIT Press. Dissanayake, E. (2009). Root, leaf, blossom, or bole: Concerning the origin and adaptive function of music. In S. Malloch, & C. Trevarthen (Eds.), Communicative musicality: Exploring the basis of human companionship (pp. 17–30). Oxford: Oxford University Press Dölen, G., Darvishzadeh, A., Huang, K. W., & Malenka, R. C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. Nature 501(7466): 179–84 . Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mindreading" in humans. Biological Psychiatry 61(6): 731–33. Dunbar, R. I. M. (1991). Functional significance of social grooming in primates. Folia Primatologica 57: 121-31. Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences 16: 681-735. Dunbar, R. I. M. (2012a). On the evolutionary function of song and dance. In N. Bannan (Ed.), Music, language, and human evolution (pp. 201–214). Oxford: Oxford University Press. Dunbar, R. I. M., (2012b) Bridging the bonding gap: The transition from primates to humans. Philosophical Transactions of the Royal Society B: Biological Science 367(1597): 1837-46. Dunbar, R. I. M., & Shultz, S. (2010). Bondedness and sociality. Behaviour, 147(7): 775–803. Durham, W. H. (1991). Coevolution: Genes, culture, and human diversity. Stanford, CA: University of California Press. Erber, R., Wegner, D. M., & Therriault, N. (1996). On being cool and collected: Mood regulation in anticipation of social interaction. Journal of Personality and Social Psychology 70(4): 757-66. Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences 32(5): 429–92. Ewell, P. A. (2020). Music theory and the white racial frame. Music Theory Online, 26(2). https://doi.org/

10.30535/mto.26.2.4 Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? Behavioral and Brain Sciences 27(4): 491–541. Farabaugh, S. M. (1982). The ecological and social significance of duetting. In D. S. Kroodsma & E. H. Miller (Eds.), Acoustic Communication in Birds, Volume 2 (pp. 85-124). New York, NY: Academic Press. Fazenda, B., Scarre, C., Till, R., Pasalodos, R.J., Guerra, M.R. Tejedor, C., Peredo, R.O., Watson, A. Wyatt, S., Benito, C.G., Drinkall, H., & Foulds, F. (2017). Cave acoustics in prehistory: exploring the https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. association of Palaeolithic visual motifs and acoustic response. The Journal of the Acoustic Society of America 142, 1332-49. Feekes, F. (1982). Song mimesis within colonies of Cacicus c. cela (Icteridae: Aves): A colonial password? Zeitschrift Tierpsychology 58: 119-52. Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? Child development 60(6): 1497-510. Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture (pp. 391–428). Oxford University Press. Finlay, B. L., Darlington, R. B., & Nicastro, N. (2001). Developmental structure in brain evolution. Behavioral and Brain Sciences 24: 263–308. Fitch, W. T. (2004). Evolving honest communication systems: Kin selection and "mother tongues". In D. K. Oller & U. Griebel (Eds.), Evolution of Communication Systems: A Comparative Approach (pp. 275–296). Cambridge, MA: MIT Press. Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. Cognition 100(1): 173- 215. Fitch, W. T. (2010). The evolution of language. Cambridge: Cambridge University Press. Fitch, W. T. (2015a). Four principles of bio-musicology. Philosophical Transactions of the Royal Society B: Biological Sciences 370(1664): e2014.0091. http://doi.org/10.1098/rstb.2014.0091 Fitch, W. T. (2015b). The biology and evolution of musical rhythm: An update. In I. Toivonen, P. Csúri, & E. van der Zee (Eds.), Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff (pp. 293-324). Cambridge, Massachusetts: MIT Press. Fitch, W. T. (2017). Empirical approaches to the study of language evolution. Psychonomic Bulletin & Review 24: 3-33. Fitch, W. T., von Graevenitz, A., and Nicolas, E. (2009). Bio-aesthetics and the aesthetic trajectory: A dynamic cognitive and cultural perspective. In M. Skov & O. Vartanian (Eds.), Neuroaesthetics, (pp. 59-102). Amityville, NY: Baywood. Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In M. A. Arbib (Ed.), Language, Music, and the Brain: A Mysterious Relationship (pp. 499- 539). Cambridge, Massachusetts: MIT Press. Fletcher, P. T., Whitaker, R. T., Tao, R., DuBray, M. B., Froehlich, A., Ravichandran, C., … Lainhart, J. E. (2010). Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. NeuroImage 51(3): 1117–25. Freeman, W. (2000). A neurobiological role for music in social bonding. In N. L. Wallin, B. Merker, & S. Brown (Eds.), The Origins of Music (pp. 411-24). Cambridge, MA: MIT Press. Frigyesi, J. (1993). Preliminary thoughts toward the study of music without clear beat: The example of "Flowing Rhythm" in Jewish "Nusah". Asian Music 24: 59-88. Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience 11: 127– 37. Fujii, T., Schug, J., Nishina, K., Takahashi, T., & Okada, H. (2016). Relationship between salivary oxytocin levels and generosity in preschoolers. Scientific Reports 6(38662):1–7. Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. Journal of Neuroscience 35(45): 15187–98. Gadagkar, V., Puzerey, P. A., Chen, R., Baird-Daniel, E., Farhang, A. R., & Goldberg, J. H. (2016). Dopamine neurons encode performance error in singing birds. Science 354(6317): 1278-82. Gamble. C. (2010).

Technologies of separation and the evolution of social extension. In Social Brain, Distributed Mind (Eds, Dunbar, R,., Gamble, C. & Gowlett, J), pp. 17-42. Oxford: Oxford University Press. https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Gebauer, L., Witek, M. A. G., Hansen, N. C., Thomas, J., Konvalinka, I., & Vuust, P. (2016). Oxytocin improves synchronisation in leader-follower interaction. Scientific Reports, 6, 38416. https://doi.org/10.1038/srep38416 Geissmann, T. (1999). Duet songs of the siamang, Hylobates syndactylus: II. Testing the pair-bonding hypothesis during a partner exchange. Behaviour 136(8): 1005-39. Gelfand, M. J., Caluori, N., Jackson, J. C., & Taylor, M. K. (2020). The cultural evolutionary trade-off of ritualistic synchrony. Philosophical Transactions of the Royal Society B: Biological Sciences, 375: 20190432. https://doi.org/10.1098/rstb.2019.0432 Geschwind, N. (1965). Disconnexion syndromes in animal and man. Brain 88: 585–644. Ghazban, N. (2013). Emotion regulation in infants using maternal singing and speech. PhD dissertation, Ryerson University, Toronto. Gil, D., & Gahr, M. (2002). The honesty of bird song: Multiple constraints for multiple traits. Trends in Ecology & Evolution 17(3): 133–41. Gill, K. Z., & Purves, D. (2009). A biological rationale for musical scales. PLOS ONE 4(12): e8144. https://doi.org/10.1371/journal.pone.0008144 Gingras, B., Honing, H., Peretz, I., Trainor, L. J., & Fisher, S. E. (2018). Defining the biological bases of individual differences in musicality. In H. Honing (Ed.), The origins of musicality (pp. 221–250). Cambridge, MA: MIT Press. Gioia, T. (2019). Music: A subversive history. New York: Basic Books. Good, A., & Russo, F. A. (2016). Singing promotes cooperation in a diverse group of children. Social Psychology, 47(6): 340–344. Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., & Zatorre, R. J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. Proceedings of the National Academy of Sciences USA 116(8): 3310–15. Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. Journal of Cognitive Neuroscience 19(5): 893–906. Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. Journal of Neuroscience 29(23): 7540–48. Grahn, J., Bauer, A.-K. R., & Zamm, A. (2020, April 28). Music-making brings us together during the coronavirus pandemic. The Conversation. https://theconversation.com/music-making-brings-ustogether-during-the-coronavirus-pandemic-137147 Grape, C., Sandgren, M., Hansson, L.-O., Ericson, M., & Theorell, T. (2003). Does singing promote wellbeing?: An empirical study of professional and amateur singers during a singing lesson. Integrative Physiological & Behavioral Science 38(1): 65–74. Grawunder, S., Crockford, C., Clay, Z., Kalan, A. K., Stevens, J. M. G., Stoessel, A., & Hohmann, G. (2018). Higher fundamental frequency in bonobos is explained by larynx morphology. Current Biology 28(20): R1188–89. Greenwood, D. N., & Long, C. R. (2009). Mood specific media use and emotion regulation: Patterns and individual differences. Personality and Individual Differences 46(5-6): 616-621. Griffiths, P. E. (2003). Beyond the Baldwin effect: James Mark Baldwin's "social heredity", epigenetic inheritance, and niche construction. In B. Weber & D. J. Depew (Eds.), Learning, Meaning and Emergence: Possible Baldwinian Mechanisms in the Co-Evolution of Mind and Language (pp. 193– 215). Cambridge, MA: MIT Press. Gustison, M. L., Aliza, L. R., and Bergman, T. (2012). Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 1847-59. Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. Human Nature 14(1): 21–51. https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.

Haimoff, E. H. (1986). Convergence in the duetting of monogamous Old World primates. Journal of Human Evolution 15: 51-59. Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. Behavioral Ecology and Sociobiology 55: 415-430. Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. Frontiers in Psychology 2: article 156. http://doi.org/10.3389/fpsyg.2011.00156 Hanslick, E. (1858). Vom Musikalisch-Schönen. Leipzig: Ruoldf Weigel. Harrison, P. M. C., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition. Psychological Review. 127(2): 216-244. Henrich, J. (2016). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter. Princeton, NJ: Princeton University Press. Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., & ten Cate, C. (2018). Searching for the origins of musicality across species. In H. Honing (Ed.), The origins of musicality (pp. 149–170). Cambridge, MA: MIT Press. Honing, H., Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. Philosophical Transactions of the Royal Society B: Biological Sciences 370: 20140088. http://doi.org/10.1098/rstb.2014.0088 Honing, H. (Ed.). (2018). The origins of musicality. Cambridge, MA: MIT Press. Horowitz, J. (2020, March 14). Italians find 'a moment of joy in this moment of anxiety'. New York Times. https://www.nytimes.com/2020/03/14/world/europe/italians-find-a-moment-of-joy-in-this-moment-ofanxiety.html Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. Social Cognition 27(6): 949–60. Hrdy SB. (2009) Mothers and others: The evolutionary origins of mutual understanding. Cambridge, MA: Harvard University Press. Hurlemann, R., Patin, A., Onur, O.A., Cohen, M.X., Baumgartner, T., Metzler, S.,…, Kendrick, K.M. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. Journal of Neuroscience 30:4999–5007. Huron, D. (2001). Is music an evolutionary adaptation? Annals of the New York Academy of Sciences 930: 43–61. Huron, D. (2006). Sweet anticipation: Music and the psychology of expectation. MIT Press. Inagaki, T.K. (2018). Opioids and social connection. Current Directions in Psychological Science, 27: 85– 90. Inagaki, T.K., Ray, L.A., Irwin, M.R., Way, B.M., & Eisenberger, N.I. (2016). Opioids and social bonding: naltrexone reduces feelings of social connection. Social Cognitive and Affective Neuroscience 11: 728– 35. Iyer, V., & Born, G. (2020, February 18). Of musicalities and musical experience: Vijay Iyer and Georgina Born in conversation. Wigmore Hall Podcasts. https://wigmore-hall.org.uk/podcasts/of-musicalitiesand-musical-experience-vijay-iyer-and-georgina-born-in-conversation Jablonka, E., Ginsburg, S., & Dor, D. (2012). The co-evolution of language and emotions. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1599): 2152–2159. Jablonka, E., & Lamb, M. J. (2005). Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life. MIT Press. Jackson, S. J., & Hokowhitu, B. (2002). Sport, tribes, and technology: The New Zealand All Blacks haka and the politics of identity. Journal of Sport & Social Issues, 26(2): 125–39. Jacoby, N., Margulis, E., Clayton, M., Hannon, E., Honing, H., Iversen, J.,…, Wald-Fuhrmann, M. (2020). Cross-cultural work in music cognition: Methodologies, pitfalls, and practices. Music Perception, 37(3): 185–195. https://doi.org/10.1017/S0140525X20000333
Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. Current Biology 27: 359–70. Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. Current Biology, 29, 3229–43. Jacoby, N., et al. (In prep.). A global survey of rhythm

representations. Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. Journal of Experimental Psychology 141(1): 54–75. Janik, V. M., & Slater, P. J. B. (1999). Vocal learning in mammals. Advances in the Study of Behavior 26: 59–99. Järvinen-Pasley, A., Vines, B. W., Hill, K. J., Yam, A., Grichanik, M., Mills, D.,…, Bellugi, U. (2010). Cross-modal influences of affect across social and non-social domains in individuals with Williams syndrome. Neuropsychologia, 48(2): 456–466. Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. Science, 366: 50–54. Jones, M. R. (2018). Time will tell: A theory of dynamic attending. Oxford University Press. Josef, L., Goldstein, P., Mayseless, N., Ayalon, L., & Shamay-tsoory, S.G. (2019). The oxytocinergic system mediates synchronized interpersonal movement during dance. Scientific Reports 9(1894):1–8. Keeler, J. R., Roth, E. A., Neuser, B. L., Spitsbergen, J. M., Waters, D. J. M., & Vianney, J.-M. (2015). The neurochemistry and social flow of singing: Bonding and oxytocin. Frontiers in Human Neuroscience 9: article 518. https://doi.org/10.3389/fnhum.2015.00518 Keverne, E. B., Martensz, N. D. & Tuite, B. (1989). Beta-endorphin concentrations in cerebrospinal fluid of monkeys are influenced by grooming relationships. Psychoneuroendocrinology 14: 155—61. Kim, M., & Schachner, A. (2020). The origins of dance: Characterizing the development of infants' earliest dance behavior. Poster presented at the Brain Cognition Emotion Music Conference. https://osf.io/meetings/BCEM/. Kirby, S. (2017). Culture and biology in the origins of linguistic structure. Psychonomic Bulletin & Review, 24: 118–137. Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. Evolution and Human Behavior 31(5): 354–64. Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). fMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12(1): 20-27. Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. Trends in Cognitive Sciences, 23(1), 63–77. Kokal, I., Engel, A., Kirschner, S., & Keysers, C. (2011). Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment - if the rhythm comes easily. PLOS ONE 6(11): e27272. http://doi.org/10.1371/journal.pone.0027272 Kornhaber, S. (2020, April 9). The coronavirus has forced a repurposing of music. The Atlantic. https://www.theatlantic.com/culture/archive/2020/04/coronavirus-has-forced-repurposingmusic/609601/ Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. Nature 435(7042): 673–76. Kroodsma, D. E. (1978). Continuity and versatility in bird song: Support for the monotony-threshold hypothesis. Nature, 274(5672), 681–3. Kuroyanagi, J., Sato, S., Ho, M.-J., Chiba, G., Six, J., Pfordresher, P., … Savage, P. E. (2019). Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. In Proceedings of the 9th International Workshop on Folk Music Analysis (FMA2019), pp. 35-40. http://doi.org/10.31234/osf.io/zpv5w Lagacé, R. O. (1979). The HRAF Probability Sample: Retrospect and prospect. Cross-Cultural Research, 14(3), 211–29.
Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. Social Cognition 29(1): 1–14. Laland, K. N., & Brown, G. R. (2011). Sense and nonsense: Evolutionary perspectives on human behaviour (2nd ed.). Oxford University Press. Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. Behavioral and Brain Sciences 23: 131–75. Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: Bringing genetics and the human sciences together. Nature Reviews Genetics 11: 137–48. Laland, K., Wilkins, C., & Clayton, N. (2016). The evolution of dance. Current Biology 26(1): R5–9. Lang, M., Bahna, V., Shaver, J. H., Reddish, P., & Xygalatas, D. (2017). Sync to link:

Endorphin-mediated synchrony effects on cooperation. Biological Psychology, 127: 191–197. Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. Music Perception, 33(3), 319–31. Launay, J., Dean, R. T., & Bailes, F. (2013). Synchronization can influence trust following virtual interaction. Experimental Psychology 60(1): 53–63. Launay, J., Tarr, B., & Dunbar, R. I. M. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. Ethology 122: 779–89. Le Merrer, J., Becker, J. A., Befort, K., & Kieffer, B. L. (2009). Reward processing by the opioid system in the brain. Physiological reviews 89(4): 1379-1412. Lense, M. D., Gordon, R. L., Key, A. P. F., & Dykens, E. M. (2014). Neural correlates of cross-modal affective priming by music in Williams syndrome. Social Cognitive and Affective Neuroscience, 9(4), 529–537. Lepage, C., Drolet, P., Girard, M., Grenier, Y., & DeGagné, R. (2001). Music decreases sedative requirements during spinal anesthesia. Anesthesia & Analgesia 93(4): 912-16. Lomax, A. (Ed.). (1968). Folk song style and culture. Washington, DC: American Association for the Advancement of Science. London, J. (2004). Hearing in time: Psychological aspects of musical meter. Oxford: Oxford University Press. Longfellow, H. W. (1835). A pilgrimage beyond the sea. New York: Harper & Bros. Loersch, C., & Arbuckle, N. L. (2013). Unraveling the mystery of music: Music as an evolved group process. Journal of Personality and Social Psychology 105(5): 777–98. Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? The Journal of Neuroscience 29(33): 10215–20. Loui, P., Li, H. C., & Schlaug, G. (2011). White matter integrity in right hemisphere predicts pitch-related grammar learning. NeuroImage 55(2): 500–07. Loui, P., Patterson, S., Sachs, M. E., Leung, Y., Zeng, T., & Przysinda, E. (2017). White matter correlates of musical anhedonia: Implications for evolution of music. Frontiers in Psychology 8: article 1664. http://doi.org/10.3389/fpsyg.2017.01664 Love, T. M. (2014). Oxytocin, motivation and the role of dopamine. Pharmacology Biochemistry and Behavior 119: 49-60. Maestripieri, D. (2010). Neurobiology of social behavior. In M. Platt & A. Ghazanfar (Eds.), Primate Neuroethology (pp. 359–384). Oxford: Oxford University Press. Mann, N. I., Dingess, K. A., Barker, K., Graves, J. A., & Slater, P. J. B. (2009). A comparative study of song form and duetting in neotropical Thryothorus wrens. Behaviour 146: 1-43. Mann, N. I., Dingess, K. A., and Slater, P. J. B. (2006). Antiphonal four-part synchronized chorusing in a Neotropical wren. Biology Letters 2: 1-4. ManyBabies Consortium (2020). Quantifying sources of variability in infancy research using the infantdirected speech preference. Advances in Methods and Practices in Psychological Science, 3(1): 24–52. Margulis, E. H. (2014). On repeat: How music plays the mind. New York: Oxford University Press. https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2016). Neural correlates of specific musical anhedonia. Proceedings of the National Academy of Sciences USA 113(46): E7337–45. Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., & Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. Current Biology 24(6): 699– 704. McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. Journal of Experimental Psychology: General, 135(3): 348–367. McBride, J. M., & Tlusty, T. (2020). Cross-cultural data shows musical scales evolved to maximise imperfect fifths. ArXiv Preprint. http://arxiv.org/abs/1906.06171 McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. Current Biology 20(11): 1035–1041. McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in

music perception. Nature, 535: 547–550. McNeil, W. H. (1995). Keeping together in time: Dance and drill in human history. Cambridge, MA: Harvard University Press. Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. Evolution and Human Behavior 38: 674–84. Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (This issue). Origins of music in credible signaling. Behavioral and Brain Sciences. Preprint: https://doi.org/10.31234/osf.io/nrqb3 Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. Current Biology 28: 356–368. Mehr, S. A., Singh, M., Knox, D., Lucas, C., Ketter, D. M., Pickens-Jones, D., … Glowacki, L. (2019). Universality and diversity in human song. Science, 366: eaax0868. https://doi.org/10.1126/science.aax0868 Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-month-old infants, melodies are social. Psychological Science 27(4): 486-501. Mercado, E., Mantell, J. T., & Pfordresher, P. Q. (2014). Imitating sounds: A cognitive approach to understanding vocal imitation. Comparative Cognition & Behavior Reviews, 9: 17–74. Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: A neural perspective across humans and non-human primates. Philosophical Transactions of the Royal Society B: Biological Sciences 370: 20140093. http://doi.org/10.1098/rstb.2014.0093 Merker, B. (1999). Synchronous chorusing and the origins of music. Musicae Scientiae 3: 59-73. Merker, B. (2000). Synchronous chorusing and human origins. In The origins of music (pp. 315–328). Cambridge, MA: MIT Press. Merker, B. (2002). Music: The missing Humboldt system. Musicae Scientiae, 6(1), 3–21. Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. Cortex, 45(1), 4–17. Merker, B., Morley, I., & Zuidema, W. (2018). Five fundamental constraints on theories of the origins of music. In H. Honing (Ed.), The origins of musicality (pp. 70–104). MIT Press. Merriam, A. P. (1964). The anthropology of music. Northwestern University Press. Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. Journal of Experimental Social Psychology 45(3): 585–89. Miller, G. F. (2000). Evolution of human music through sexual selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), The origins of music (pp. 329–360). Cambridge, MA: MIT Press. Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. Journal of Cognitive Neuroscience 17(8): 1306-15. https://doi.org/10.1017/S0140525X20000333
Mithen, S. J. (2005). The singing Neanderthals: The origins of music, language, mind, and body. London: Weldenfeld & Nicholson. Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. Journal of Experimental Social Psychology 72: 13–20. Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., & Overy, K. (2017). Diffusion tensor MRI tractography reveals increased fractional anisotropy (FA) in arcuate fasciculus following music-cued motor training. Brain and Cognition 116: 40–46. Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences USA 114(42): E8913–E8921. Morley, I. (2013). The prehistory of music: Human evolution, archaeology, and the origins of musicality. Oxford: Oxford University Press. Mueller, C., Klega, A., Buchholz, H.-G., Rolke, R., Magerl, W., Schirrmacher, R., … Schreckenberger, M. (2010). Basal opioid receptor binding is associated with differences in sensory perception in healthy human subjects: a [18F]diprenorphine PET study. NeuroImage 49(1): 731–37. Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. Infant Behavior and Development 27(4): 455–64. Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. Perspectives on

Psychological Science, 10(6), 772–89. Nettl, B. (2015). The study of ethnomusicology: Thirty-three discussions (3rd ed.). Champaign: University of Illinois Press. Nettl, B., Stone, R., Porter, J., & Rice, T. (Eds.). (1998-2002). The Garland encyclopedia of world music. New York: Garland Pub. Niarchou, M., Sathirapongsasuti, J. F., Jacoby, N., Bell, E., McArthur, E., Straub, P., … Gordon, R. L. (2019). Unravelling the genetic architecture of rhythm. bioRxiv preprint. https://doi.org/10.1101/836197 Nooshin, L. (2011). Introduction to the special issue: The ethnomusicology of Western art music. Ethnomusicology Forum 20(3): 285–300. North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. Music Perception, 22(1): 41–77. Notroff, J. Dietrich, O. & Schmidt, K. (2015). Gathering of the dead? The early Neolithic sanctuaries of Gobekli Tepe, southerneastern Turkey. In Death Rituals, Social Order and the Archaeology of Immortality in the Ancient World (eds. Renfrew, Sc., Boyd, M.J, and Morely, I), pp. 65-18. Cambridge: Cambridge University Press. Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. Current Opinion in Neurobiology 28: 48-53. Oesch, N. (2019). Music and language in social interaction: Synchrony, antiphony, and functional origins. Frontiers in Psychology, 10, 1514. https://doi.org/10.3389/fpsyg.2019.01514 Parkinson, C., & Wheatley, T. (2014). Relating anatomical and social connectivity: White matter microstructure predicts emotional empathy. Cerebral Cortex 24(3): 614–25. Patel, A. D. (2008). Music, language and the brain. Oxford: Oxford University Press. Patel, A. D. (2014). The evolutionary biology of musical rhythm: Was Darwin wrong? PLoS Biology 12(3): e1001821. http://doi.org/10.1371/journal.pbio.1001821 Patel, A. D. (2018). Music as a transformative technology of the mind: An update. In H. Honing (Ed.), The origins of musicality (pp. 113–26). Cambridge, MA: MIT Press. 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: 827–30. Pearce, E., Launay, J., & Dunbar, R. I. M. (2015). The ice-breaker effect: Singing mediates fast social bonding. Royal Society Open Science 2: 150221. http://doi.org/10.1098/rsos.150221 https://doi.org/10.1017/S0140525X20000333
Peretz, I., & Coltheart, M. (2003). Modularity of music processing. Nature Neuroscience, 6(7): 688–691. Peretz, I., Vuvan, D. T., Lagrois, M.-É., & Armony, J. (2018). Neural overlap in processing music and speech. In H. Honing (Ed.), The origins of musicality (pp. 205–220). Cambridge, MA: MIT Press. Pfordresher, P. Q., & Brown, S. (2017). Vocal mistuning reveals the origin of musical scales. Journal of Cognitive Psychology, 29(1): 35–52. Pinker, S. (1997). How the mind works. New York: Norton. Pinker, S. (2007). Toward a consilient study of literature. Philosophy and Literature 31(1): 162–78. Pinker, S. (2012, June 18). The false allure of group selection. Edge. Available at https://www.edge.org/conversation/steven_pinker-the-false-allure-of-group-selection Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. Frontiers in Neuroscience, 11(542). https://doi.org/10.3389/fnins.2017.00542 Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: A comparative study of tapping synchronization. Music Perception 36(1): 1–23. Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? Journal of Experimental Psychology: Human Perception and Performance, 35(5): 1598–1617. Rabinowitch, T.-C., Cross, I., & Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. Psychology of Music, 41(4): 484–498. Rabinowitch, T. C., & Meltzoff, A. N. (2017). Synchronized movement experience enhances peer cooperation in preschool children. Journal of Experimental Child Psychology 160: 21–32. Ragen, B. J., Maninger, N., Mendoza, S. P., Jarcho, M. R., & Bales, K. L. (2013). Presence of a pair-mate regulates the behavioral and

physiological effects of opioid manipulation in the monogamous titi monkey (Callicebus cupreus). Psychoneuroendocrinology 38(11): 2448–61. Rainio, R., Lahelma, A., Aikas, T., Lassfolk, K. & Okkonen, J. (2018). Acoustic measurements and digital image processing suggest a link between sound rituals and sacred sites in northern Finland. Journal of Archaeological Method and Theory 25, 453-74. Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. Nature Human Behaviour 1: article 7. http://doi.org/10.1038/s41562-016-0007 Read, D. & Van der Feeuw, S. (2015). The extension of social relations in time and space during the Palaeolithic and beyond. In Settlement, Society and Cognition in Human Evolution (eds, Coward, F., Hosfield, R., Pope, M. & Wenban-Smith, F), pp. 31-53. Cambridge: Cambridge University Press Reddish, P., Bulbulia, J., & Fischer, R. (2014). Does synchrony promote generalized prosociality? Religion, Brain & Behavior 4(1): 3–19. Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's dance together: Synchrony, shared intentionality and cooperation. PLOS ONE 8(8): e71182. http://doi.org/10.1371/journal.pone.0071182 Rennung, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony: A meta-analysis. Zeitschrift Für Psychologie 224(3): 168–89. Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. Psychological Science 17(3): 236–242. Richerson, P., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., … Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. Behavioral and Brain Sciences 39: e30. http://doi.org/10.1017/S0140525X1400106X Richerson, P. J., Boyd, R., & Henrich, J. (2010). Gene-culture coevolution in the age of genomics. Proceedings of the National Academy of Sciences USA 107: 8985–92. Gustison, M. L., Aliza, L. R., and Bergman, T. (2012). "Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates," Philosophical Transactions of the Royal Society of London B, 367: 1847-1859. Richman, B. (1978). The synchronisation of voices by gelada monkeys. Primates 19: 569-81. Richman, B. (1987). Rhythm and melody in gelada vocal exchanges. Primates 28: 199-223. https://doi.org/10.1017/S0140525X20000333
Rock, A. M. L., Trainor, L. J. D., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. Developmental Psychology, 35(2): 527–34. Roederer, J. G. (1984). The search for a survival value of music. Music Perception 1(3): 350–356. Sachs, C. (1943). The rise of music in the ancient world: East and West. New York: Norton. Sachs, M. E., Ellis, R. J., Schlaug, G., & Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. Social Cognitive and Affective Neuroscience 11(6): 884–91. Sakata, J. T., & Brainard, M. S. (2008). Online contributions of auditory feedback to neural activity in avian song control circuitry. Journal of Neuroscience 28(44): 11378-90. Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nature Neuroscience 14(2): 257–264. Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. Science 340(6129): 216–219. Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. Trends in Cognitive Sciences 19(2): 86-91. Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. Current Biology 25(23): 3079–85. Sanchez, J.L.J. 2007. Procession and performance: recreating ritual soundscapes among the Ancient Maya. The World of Music 49, 35-44. Savage, P. E. (2018). Alan Lomax's Cantometrics Project: A comprehensive review. Music & Science 1. http://doi.org/10.1177/2059204318786084

Savage, P. E. (2019a). Cultural evolution of music. Palgrave Communications 5: article 16. http://doi.org/10.1057/s41599-019-0221-1 Savage, P. E. (2019b). Universals. In J. L. Sturman (Ed.), The SAGE International Encyclopedia of Music and Culture (p. 2282–2285). Thousand Oaks: SAGE Publications. http://doi.org/10.4135/9781483317731.n759 Savage, P. E. (2019c). An overview of cross-cultural music corpus studies. In D. Shanahan, A. Burgoyne, & I. Quinn (Eds.), Oxford Handbook of Music and Corpus Studies. New York: Oxford University Press. http://doi.org/10.31235/osf.io/nxtbg Savage, P. E., & Brown, S. (2013). Toward a new comparative musicology. Analytical Approaches to World Music 2(2): 148–97. 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 USA 112(29): 8987–92. Savage, P. E., Merritt, E., Rzeszutek, T., & Brown, S. (2012). CantoCore: A new cross-cultural song classification scheme. Analytical Approaches to World Music, 2(1): 87–137. Savage, P. E., Yamauchi, M., Hamaguchi, M., Tarr, B., Kitayama, Y., & Fujii, S. (2020). Rhythm, synchrony, and cooperation [Stage 1 Registered Report]. PsyArXiv preprint. https://doi.org/10.31234/osf.io/46bd9 Scarl, J. C., & Bradbury, J. W. (2009). Rapid vocal convergence in an Australian cockatoo, the galah. Animal Behaviour 77: 1019-26. 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: 831–36. Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. Developmental Psychology 47(1): 19–25. Schachner, A., Brady, T. F., Oro, K., & Lee, M. (Preprint). Intuitive archeology: Detecting social transmission in the design of artifacts. PsyArXiv preprint: https://doi.org/10.31234/osf.io/g6qxh Schino, G., & Troisi, A. (1992). Opiate receptor blockade in juvenile macaques: Effect on affiliative interactions with their mothers and group companions. Brain Research 576: 125–30.
Schulkin, J., & Raglan, G. B. (2014). The evolution of music and human social capability. Frontiers in Neuroscience 8: article 292. https://doi.org/10.3389/fnins.2014.00292 Schruth, D., Templeton, C. N., & Holman, D. J. (In press). On reappearance and complexity in musical calling. PLOS ONE. Preprint: https://doi.org/10.1101/649459 Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. Annual Review of Neuroscience 23: 473-500. Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science 275(5306): 1593-99. Sethares, W. (2004). Tuning, timbre, spectrum, scale. Springer-Verlag. Shamay-Tsoory, S. G., & Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. Biological Psychiatry 79(3): 194-202. Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human social evolution: Self-domestication or selfcontrol? Frontiers in Psychology, 11(134). https://doi.org/10.3389/fpsyg.2020.00134 Shultz, S., & Dunbar, R. I. M. (2007). The evolution of the social brain: Anthropoid primates contrast with other vertebrates. Proceedings of the Royal Society B: Biological Sciences 274: 2429-36. Shultz, S., Opie, C., & Atkinson, Q. D. (2011). Stepwise evolution of stable sociality in primates. Nature, 479(7372): 219–222. Silk, J. B. (2007). Social components of fitness in primate groups. Science 317: 1347–1351. Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the experience sampling method. Musicae Scientiae, 5(1): 9–32. Small, C. (1998). Musicking: The meanings of performing and listening. Hanover: University Press of New England. Soley, G., & Spelke, E. S. (2016). Shared cultural knowledge: Effects of music on young children's social preferences. Cognition 148: 106-116. Stokes, M. (Ed.). (1994). Ethnicity, identity, and music: The musical construction of place. Oxford: Berg. Stefano, G. B., Zhu, W., Cadet, P., Salamon, E., & Mantione, K. J. (2004). Music alters constitutively expressed opiate and

cytokine processes in listeners. Medical Science Monitor 10(6): MS18-27. Stevens, C., & Byron, T. (2016). Universals in music processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), Oxford Handbook of Music Psychology (pp. 19–31). Oxford: Oxford University Press. Stupacher, J., Hove, M. J., Novembre, G., Schutz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: A TMS investigation. Brain and Cognition 82(2): 127-36. Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: Social motivation in the evolution and development of language. Developmental Science 14(2), 417–430. Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. Proceedings of the National Academy of Sciences USA 109(21): 8038-43. Tarr, B. (2017). Social bonding through dance and "Musiking." In N. L. Enfield & P. Kockelman (Eds.), Distributed agency (pp. 151–158). Oxford University Press. Tarr, B., Launay, J., & Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. Frontiers in Psychology 5: article 1096. https://doi.org/10.3389/fpsyg.2014.01096 Tarr, B., Launay, J., Benson, C., & Dunbar, R. I. M. (2017). Naltrexone blocks endorphins released when dancing in synchrony. Adaptive Human Behavior and Physiology 3(3): 241–54. Tarr, B., Launay, J., Cohen, E., & Dunbar, R. I. M. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. Biology Letters 11: 20150767. http://doi.org/http://dx.doi.org/10.1098/rsbl.2015.0767 Tarr, B., Launay, J., & Dunbar, R. I. M. (2016). Silent disco: Dancing in synchrony leads to elevated pain thresholds and social closeness. Evolution and Human Behavior 37(5): 343–49. Tarr, B., Slater, M., & Cohen, E. (2018). Synchrony and social connection in immersive Virtual Reality. Scientific Reports 8: article 3693. http://doi.org/10.1038/s41598-018-21765-4 https://doi.org/10.1017/S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Tchalova, K., & Macdonald, G. (2020) Opioid receptor blockade inhibits self-disclosure during a closenessbuilding social interaction. Psychoneuroendocrinology 113:104559. Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. Music Perception 1(3): 276-95. Thornton, S. (1995). Club cultures: Music, media, and subcultural capital. Wesleyan University Press. Thorpe, W. H. (1972). Duetting and antiphonal song in birds: Its extent and significance. Behaviour. Supplement 18: 1-197. Tinbergen, N. (1963). On aims and methods of ethology. Zeitschrift Für Tierpsychologie, 20: 410–433. Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., … Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genetics 39(1): 31–40. Titze, I.R. (1989). Physiologic and acoustic differences between male and female voices. Journal of the Acoustical Society of America 85(4): 1699-1707. Tolbert, E. (1990). Women cry with words: Symbolization of affect in the Karelian lament. Yearbook for Traditional Music 22: 80-105. Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. Annual Review of Psychology, 64(1): 231–255. Tomlinson, G. (2018). Culture and the course of human evolution. University of Chicago Press. Trainor, L. J. (2018). The origins of music: Auditory scene analysis, evolution, and culture in music creation. In H. Honing (Ed.), The origins of musicality (pp. 81–112). Cambridge, MA: MIT press. Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science, 11(3), 188–195. Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infantdirected singing. Infant Behavior and Development 20(3): 383-96. Traulsen, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. Proceedings of the National Academy of Sciences USA 103(29): 10952–55. Trehub, S. E. (2016). Infant musicality. In S. Hallam, I. Cross, & M. Thaut (Eds.), The Oxford handbook of

music psychology, 2nd ed. (pp. 387–397). Oxford: Oxford University Press. Trehub, S. E., Becker, J., & Morley, I. (2018). Cross-cultural perspectives on music and musicality. In H. Honing (Ed.), The origins of musicality (pp. 129–148). Cambridge, MA: MIT Press. Trehub, S. E., & Nakata, T. (2001). Emotion and music in infancy. Musicae Scientiae 5(1 Supplement): 37- 61. Trehub, S. S. E., Unyk, A. A., Kamenetsky, S., Hill, D., Trainor, L. J. D., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. Developmental Psychology 33(3): 500–7. Trehub, S.E., Unyk, A.M., & Trainor, L.J. (1993) Adults identify infant-directed music across cultures. Infant Behavior and Development 16: 193-211. Turchin, P., Currie, T. E., Whitehouse, H., François, P., Feeney, K., Mullins, D., … Spencer, C. (2018). Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. Proceedings of the National Academy of Sciences USA 115(2): E144–51. Turino, T. (2008). Music as social life: The politics of participation. Chicago: University of Chicago Press. Valdesolo, P., & Desteno, D. (2011). Synchrony and the social tuning of compassion. Emotion 11(2): 262–6. Vanden Bosch der Nederlanden, C. M., Joanisse, M. F., & Grahn, J. A. (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. NeuroImage, 214, 116767. https://doi.org/10.1016/j.neuroimage.2020.116767 Vlismas, W., Malloch, S., & Burnham, D. (2013). The effects of music and movement on mother-infant interactions. Early Child Development and Care, 183(11): 1669–88. https://doi.org/10.1017/ S0140525X20000333 Downloaded from https://www.cambridge.org/core. The University of Reading, on 22 Jan 2021 at 08:48:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. Wallin, N. L., Merker, B., & Brown, S. (Eds.). (2000). The origins of music. Cambridge: MIT Press.Walum, H., Waldman, I.D., & Young, L.J. (2016). Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. Biological Psychiatry. 79(3):251–7. Wan, C. Y., Demaine, K., Zipse, L., Norton, A., & Schlaug, G. (2010). From music making to speaking: Engaging the mirror neuron system in autism. Brain Research Bulletin 82(3–4): 161–68. Wanker, R., Sugama, Y., & Prinage, S. (2005). Vocal labelling of family members in spectacled parrotlets, Forpus conspicillatus. Animal Behaviour 70: 111-18. Watson, A. & Keating, D. (1999). Architecture and sound: an acoustic analysis of megalithic monuments in prehistoric Britain. Antiquity 73, 325-36. Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I. M., & Stewart, L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. Evolution and Human Behavior 37(2): 152–58. Welch, S.P., & Eads, M. (1999). Synergistic interactions of endogenous opioids and cannabinoid systems. Brain Res. 848:183–90. Whitehouse, H. (2004). Modes of religiosity: A cognitive theory of religious transmission. AltaMira Press. Whitehouse, H. (2018). Dying for the group: Towards a general theory of extreme self-sacrifice. Behavioral and Brain Sciences 41: e192. http:// doi.org/10.1017/S0140525X18000249 Whitehouse, H., François, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., … Turchin, P. (2019). Complex societies precede moralizing gods throughout world history. Nature 568: 226–29. Wickler, W. (1980). Vocal Duetting and the pair bond: 1. Coyness and partner commitment: A hypothesis. Zeitschrift für Tierpsychologie 52: 201-209. Wiggins, G. A., Tyack, P., Scharff, C., & Rohrmeier, M. (2018). The evolutionary roots of creativity: Mechanisms and motivations. In H. Honing (Ed.), The origins of musicality (pp. 286–308). Cambridge, MA: MIT Press. Williams, G. C. (1966). Adaptation and natural selection: A critique of some current evolutionary thought. Princeton: Princeton University Press. Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. The Quarterly Review of Biology 82(4): 327-348. Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. Psychological Science 20(1): 1–5. Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. Proceedings of the National Academy of Sciences USA

106(7): 2468-71. Wood et al.(In prep.). The Global Jukebox: An interactive database of the expressive arts and culture. Wrangham, R. (2009). Catching fire: How cooking made us human. New York: Basic Books. Zak, P., Kurzban, R., & Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. Hormones and Behavior 48(5): 522–27. Zak, P., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. PLOS ONE 2(11): e1128. http://doi.org/10.1371/journal.pone.0001128 Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. Current Directions in Psychological Science 22(6): 466-470. Zatorre, R. J. (2018). Why do we love music? Cerebrum. http://www.dana.org/Cerebrum/2018/ Why_Do_We_Love_Music/ Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. Proceedings of the National Academy of Sciences USA 110(Supplement 2): 10430–37. Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. Proceedings of the National Academy of Sciences USA 107(13): 5768–73. https://doi.org/10.1017/S0140525X20000333

Some of the most intricate temporal sequences that humans produce are the sounds they use to communicate, including speech and music. People precisely coordinate their movements with sound; speakers prepare for their turn to talk, and musicians time their tones to synchronize or align with sound. Even young toddlers spontaneously drum or sway in response to auditory rhythms, which suggests that humans are predisposed to synchronize movement with sound. Because movement preparation is often slower (100– 250 ms) than the rate of musical events (up to 8–10 tones/ s), individuals must anticipate future musical events in order to prepare their movements. A striking feature of musical synchronization is the tendency people show to produce musical events prior to the sound with which they intend to synchronize; musicians produce tones about 30 to 50 ms sooner than a regular auditory beat, and nonmusicians anticipate even sooner (50–80 ms before the beat; Repp & Siu, 2015). This behavior is referred to as anticipatory synchronization, a cornerstone of human synchrony. Several key variables influence individuals' ability to synchronize actions with musical sound. One is musical training, which reduces the anticipatory synchrony, defined as negative mean asynchrony (when individuals' produced actions precede stimulus onsets, such as when they clap to music and their claps precede the onsets of the musical tones). Another factor is sensory feedback; the more feedback available from selfgenerated and external auditory outcomes, the smaller the asynchrony (Repp & Siu, 2015). A third factor is the predictability of auditory sequences; the more regular a sequence, the smaller the observed variability in asynchrony. Measures of the mean and variance of asynchrony are used by synchronization models in different ways, as we discuss below. Two prominent theories offer relevant mechanisms for musical synchronization. One is predictive coding (PC), and the other is dynamical systems (DS). We compare these theories as applied to anticipatory synchronization and discuss their assumptions, computations, and limitations (see Table 1 for a summary of key definitions). Predictive Coding Origins Early origins of the concept of predictive coding have been traced back to Helmholtz, whose account of learning was based on hierarchical layers of representation (as described in Friston, 2009). More recent precursors to predictive coding include internal models of motor 1053635 CDPXXX10.1177/09637214211053635Palmer, DemosExplaining Anticipatory Synchrony research-article2022

Corresponding Author: Caroline Palmer, Department of Psychology, McGill University Email:

Caroline.palmer@mcgill.ca

Are We in Time? How Predictive Coding and Dynamical Systems Explain Musical Synchrony

Caroline Palmer1 and Alexander P. Demos2
1 Department of Psychology, McGill University, and 2 Department of Psychology, University of Illinois at Chicago

Abstract Humans tend to anticipate events when they synchronize their actions with sound (such as when they clap to music), which has puzzled scientists for decades. What accounts for this anticipation? We review two theoretical mechanisms for synchrony: predictive coding and dynamical systems. Both theories are grounded in neural activation patterns, but there are important distinctions. We contrast their assumptions, their computations, and their musical applications to anticipatory synchronization. Keywords anticipatory synchronization, coupling, dynamical systems, internal models, predictive coding

148 Palmer, Demos

commands (Wolpert et al., 1995) and mental simulations of partners' joint actions (Sebanz & Knoblich, 2009), according to which internal modeling of future events is based on representations of the mapping between one's perceptions and actions (Clark, 2013). Internal models of motor control predict the sensory consequences of actions by simulating behavior and compensating for time delays in sensorimotor systems (Wolpert & Flanagan, 2001). Internal models have been applied to poor-pitch singing through comparison of emulations of vocal-fold tension with auditory feedback from resulting pitches (Pfordresher & Mantell, 2014). Extensions to interpersonal, or joint, action propose that a partner's actions are simulated through the use of internal models to generate predictions for upcoming events (Sebanz & Knoblich, 2009). A musical application of joint action combines simulation of actions for oneself and for one's partners, as well as imagining sounds produced by partners, with error correction (van der Steen & Keller, 2013). Recent PC models compute the difference between an internal model and perceived events with a goal to minimize the entropy (the inverse of predictability) of predictions relative to future observed outcomes (Friston, 2009, 2018). Influenced by neurophysiological evidence of cascading processes among hierarchical

Table 1. Key Definitions

Predictive coding	Dynamical systems
Oscillation: Waves of excitation and inhibition arising at different hierarchical levels and at different frequencies among brain networks (e.g., prediction errors propagated via gamma-frequency oscillations, predictions propagated via beta-frequency oscillations; Heilbron & Chait, 2018).	Oscillation: A periodic, recurring time series with an associated amplitude and frequency (Pikovsky et al., 2001), such as the movement of a swing at a playground. Often self-sustaining (i.e., continuing in the absence of input), oscillations are modeled with mathematical equations.

Predictive timing: The use of past information to generate an internal model capable of anticipating the timing of future events (Ravignani et al., 2014). Predictive timing is often contrasted with reactive timing. Anticipatory synchronization: Behavior of an oscillator that maintains a stable negative (anticipatory) phase relationship relative to another oscillator or stimulus. This oscillatory behavior demonstrates anticipatory synchronization in the absence of internal models (Stepp & Turvey, 2010). Free energy: Energy available for a system to perform work or cause change. Based on thermodynamics principles, free energy is an information-theory measure that constrains surprise arising from model-based prediction (Friston, 2009). Prediction error: The divergence of the sensory input from the expectation (i.e., the prior). Prediction error is used to update the prior probabilities (Clark, 2013). Energy expenditure: The amount of energy used to implement change. States of least energy expenditure define steady (stable) states to which a system returns (Strogatz, 1994), such as runners' preferred rates that minimize oxygen consumption. Phase: Position within one cycle of a periodic oscillation. The phase relationship between two oscillations is often measured as a proportion of (divided by) the period (1/frequency) of the oscillations. Phase can be measured in radians or in degrees. Prior (probability): In Bayesian inference, the expected probability of a hypothesized outcome before it is known. In common variations of the term, the word "prior" is followed by

"probability," "distribution" (or "statistical information"), "belief," "knowledge," "assumption," or "expectancy." The word often refers to internal a priori "knowledge accumulated through experience" (Vuust & Witek, 2014, p. 3). Coupling: A parameter that defines how oscillators influence each other (share information) or how a stimulus influences an oscillator. Coupling can cause oscillators to match in their phase, their frequency, or both (Pikovsky et al., 2001). Posterior (probability): In Bayesian inference, the updated probability after the outcome is known. The posterior probability is calculated as the likelihood that the hypothesis predicts the input multiplied by the prior probability of the hypothesis (Denham & Winkler, 2020, p. 1157). Common variants of the term parallel the common variants of "prior" (e.g., "posterior" may be followed by "probability," "distribution," "belief," "knowledge," "assumption," or "expectancy"). Delay coupling: An oscillator model that contains a coupling term combined with a time delay and that is implemented in differential equations (Voss, 2000). The time delay can refer to different functional properties (transmission delay, delay of sensory feedback) of the oscillator. Explaining Anticipatory Synchrony 149 levels in the visual system (Rao & Ballard, 1999), recent PC models rely on Bayesian inference to update internal models and have been applied to explain binocular rivalry (i.e., alternating perception of different images presented to the two eyes) and bistable visual percepts (i.e., percepts with two stable states; see Clark, 2013, for a review). Applications of PC models to auditory perception have recently been assessed (Denham & Winkler, 2020; Heilbron & Chait, 2018). Assumptions One key assumption of this approach is that an internal model is used to predict future behavior. Comparison of perceptual input with the prediction generates a difference (prediction error) that is used to adjust the internal model and then discarded. There is no need to store the original perceptual input or the prediction error, and thus the required representation is parsimonious. A second assumption is a hierarchical organization of brain networks in which lower areas receive sensory input that is projected to higher cortical areas. A hierarchy is formed by forward excitatory connections from sensory areas to association areas and by backward inhibitory connections downward, as well as by lateral interactions between units within layers. The hierarchical system permits the computation of prediction error at one level while the internal model is updated at another level; the computations occur in a cascaded fashion. Although the temporal properties of this cascade of forward, backward, and lateral connections are not well specified, the spatial projections are proposed to include the auditory network from brain stem to cortex (Koelsch et al., 2019). Computations To create accurate predictions of future states, PC models minimize prediction error over iterations, on the basis of Bayesian inference. The bottom layer of the hierarchy contains error units that sense the incoming information and compute the prediction error, which is sent to higher layers. The top layer encodes the prior probabilities, that is, the likelihoods that given sensory inputs correspond to predicted states. This is the internal model. State units in the top layer (actively) generate the prediction signal and send it downward to the error units. In addition, units within a level are connected laterally. Error units and state units have bidirectional signals: Forward connections convey prediction error to state units, and reciprocal backward connections send the predictions to error units. These bottom-up and top-down messages serve to minimize the use of free energy (Friston, 2009). Minimizing free energy corresponds to maximizing the probability that the sensory input matches the predicted outcome, given the internal model. Over successive iterations, the state units' prior probabilities (existing beliefs before new evidence is introduced) become closer to the error units' posterior probabilities (i.e., updated beliefs after new evidence is introduced). Eventually, when prior probabilities match posterior probabilities, the internal model is stable. Applications to music Vuust and Witek (2014) proposed a PC model to explain perception of musical syncopation. Syncopation, the occurrence of musical tones in unexpected metrical locations, is typical of groove music and

often causes listeners to move, sway, or dance. Musical meter, a hierarchy of regular pulses that form patterns of strong and weak beats, provides the prior experience that tones will occur more often on strong beats. A listener's internal model of meter generates predictions that include the movements needed to produce the syncopated rhythm, while suppressing the listener's tendency toward overt action. This suppressed tendency to move arises from predictions of when the beat is expected to occur (see Koelsch et al., 2019, for related neurophysiological evidence). Prediction errors have two important parameters: a mean value and a variance associated with the mean. Vuust et al. (2018) observed that high amounts of syncopation are associated with high variance in prediction error (low precision). The more precise the prediction error, the more impact that error is expected to have. Only prediction errors associated with small variances cause the higher-level predictions to be adjusted; prediction errors with large variances are ignored (Vuust et al., 2018), a claim that is testable in synchronization tasks. Two studies have tested PC models of anticipatory synchronization. The first study (Heggli, Konvalinka, et al., 2019) tested dyadic partners' internal models as they synchronized with each other while they tapped identical duration patterns. Each partner heard over headphones a metrical context that was either the same one their partner heard (which created similar shared internal models with matched top-down predictions) or a different one (which created dissimilar internal models with unmatched top-down predictions). These matched and mismatched metrical contexts manipulated the top-down prior beliefs for when the partners' taps should occur. Some of the contexts created polyrhythms (i.e., when participants tapped three times to every 4 context beats) that could be perceived as ambiguous. Small mean negative asynchronies were observed on average. Asynchronies became more variable when 150 Palmer, Demos the two partners' metrical contexts differed (although partners recovered quickly within trials), which was interpreted as evidence for prediction errors resulting from the discrepancy in the partners' different internal models. The second study that tested PC models of anticipatory synchronization (Elliott et al., 2014) involved individual participants who were presented with two auditory cues that created ambiguous percepts with which they tried to synchronize. The two stimulus cues were altered in their asynchrony (relative phase) and temporal regularity across trials. Participants' large mean negative asynchronies decreased (i.e., moved toward 0) as the two cues' relative phase increased (i.e., diverged from 0). Participants used different strategies depending on the cues' relative phase: They synchronized to the integrated (combined) cues when the cues' phase difference was small, but they synchronized with one cue or the other when the phase difference was large. Models fitted to the asynchrony variabilities provided support for a Bayesian causal inference model with four free parameters that captured prior probabilities associated with the auditory cues; prediction errors indicated that participants attempted to minimize the variance in asynchrony and extract a temporally regular beat. Remaining issues PC applications to musical synchrony to date have targeted ambiguous percepts in order to demonstrate the internal model's impact. One open question is whether PC models can predict anticipatory synchrony (captured by mean behavior). Another is whether an internal model can change only in the presence of prediction error (when contextual information is available), or if it will also change in the absence of novel information. A final question is, what are the time costs of implementing a hierarchy that relies on several steps between layers for prediction error to be computed and sent upward and for prediction signals to be adjusted and sent downward? Dynamical Systems Origins DS theories, which explain how systems change over time, have their origins in analysis of physical synchrony among pendulum clocks (Huygens, 1665, described in Pikovsky et al., 2001) as well as biological synchrony in circadian rhythms, cardiac rhythms, and motor coordination (Haken et al., 1985; Winfree, 1967). Synchronization arises when oscillators (alternating waveforms that repeat) that are self-sustaining (continue in the absence of

external input) share information via coupling, which causes them to adapt to each other or to a stimulus (see Table 1). Their self-sustaining nature is supported by neural (magnetoencephalographic) representations of musical pulse that develop at a beat frequency despite the absence of stimulus energy at that frequency (Tal et al., 2017). When an oscillator is momentarily disturbed by input, it soon returns to its original frequency, its stable state of minimum energy expenditure. An oscillator will resonate (respond with increased amplitude) to a stimulus when its natural frequency is close to the stimulus frequency or when the coupling between oscillator and stimulus is high. Linear oscillators synchronize 1:1 with stimulus events; nonlinear oscillators additionally synchronize at higher resonances (e.g., 1:2; 1:3). The perception of hierarchical meter thus arises from higher-order resonances; internal models are not required (Large, 2008). Anticipatory synchronization arises in delay-coupled systems when a driven oscillator (i.e., an oscillator affected by external input) compares its own timedelayed memory of a previous state with the current input (Voss, 2000). Delay coupling refers to coupling between two or more oscillators that is modulated by time delay from at least one oscillator. Stepp and Turvey (2010) described delay coupling as "strong" anticipation: Anticipated future states are based on present and past information already in the system without the need for internal models. Machado and Matias (2020) demonstrated the biological plausibility of delay-coupled models by simulating delay coupling in spike neuronal populations that lead to bistable visual percepts. Time delays in DS models have been implemented to account for intrabrain synchronization dynamics (Deco et al., 2009, 2011) as well as interpersonal synchronization (Varlet et al., 2012). Assumptions DS theories of human synchrony assume that behavior arises in a system composed of coupled subsystems of oscillators whose emergent processes explain perception and cognition. Individual oscillations arise from the joint activity between coupled excitatory and inhibitory neurons and are modeled either at the biophysical level or, more commonly, by using oscillator models with simplifying assumptions. Oscillatory time series arise from neuronal interactions, as well as from neurons interacting with external quasiperiodic stimuli, such as musical sequences. Most DS theories assume continuous change over time, but they can also be modeled discretely, as in the case of musical pulse (Large, 2008). Time delay in DS models, often implemented as a constant for simplicity, is assumed to Explaining Anticipatory Synchrony 151 represent the synaptic transmission rate in a neural system (Machado & Matias, 2020). Computations Delay-coupled DS models have been proposed to account for anticipatory synchronization. Based on selfsustaining oscillators, the coupling term in these models is often adapted from the Kuramoto model (Strogatz, 2000) which assumes that each oscillator has an intrinsic frequency that can be tuned during development via Hebbian learning (strengthened connections between simultaneously firing neurons; Tichko et al., 2021). Oscillators adapt their phase on the basis of how much they differ from incoming stimulus onsets, modulated by the coupling strength; higher coupling means more phase correction and faster synchronization. Time delay compares the oscillator's past (at a constant delay) with the current input (Voss, 2000), thus providing a type of self-feedback. Most critical is that the oscillator's past incorporates the past of the stimulus; thus, the oscillator (passively) reacts to present events as a function of present and past states of both stimulus and oscillator. Applications to music Roman et al. (2019) tested synchronization patterns of individuals who tapped to a metronome, using a single Hopf oscillator (an oscillator whose intrinsic frequency can adapt to a stimulus frequency) with time delay to represent unidirectional coupling with the metronome. The model, designed to mimic neural oscillations and adapt its frequency, simulated the neural delays necessary to account for the observed tappers' negative mean asynchrony. The model successfully predicted the difference in the degree of anticipatory synchronization between musicians and nonmusicians who tapped at a range of metronome rates. Demos et al. (2019) applied a bidirectional

delaycoupled model, adapted from the Kuramoto model, to the more complex case of asymmetric anticipatory synchronization between partners playing musical duets. Each partner was modeled as a simple oscillator with parameters for time delay, coupling, and intrinsic frequency. Each oscillator could be driven by or could drive the other oscillator, depending on whether one partner heard feedback, both partners heard feedback, or neither partner heard feedback. The model accounted for anticipatory synchrony under experimental conditions that manipulated auditory feedback to shift the oscillators from bidirectional to unidirectional information transmission. The use of delay-coupling terms successfully predicted the driven partner's mean anticipatory synchronization (e.g., the driven partner performed earlier than the driver partner) when the driver partner could not hear the driven partner; this was the first delay-coupling implementation to extend beyond single-person models of musical synchrony to interpersonal interaction. Heggli, Cabral, et al. (2019) examined how partners alter their synchronization in a tapping task, using bidirectional coupling but with no time delay. Two oscillators modeled each individual's perception and action as coupled; in addition, bidirectional coupling linked each person's action to the partner's perception. The model captured different patterns of synchronization variability, such as when partners showed mutual adaptation (increased coupling) or when both partners tried to lead at the same time (reduced coupling). Similar coupled oscillator networks without time delay can produce anticipatory synchronization (Pyragiene. & Pyragas, 2015). Remaining issues The delay-coupling term in DS models accounts for both individual and dyadic anticipatory synchronization. One question is how time delays map to neural features, such as synaptic transmission (Machado & Matias, 2020), and to functional properties of neural networks (Deco et al., 2011). Another question is whether the delay is constrained by the individual, task, or social context. Finally, how can these parameters be interpreted and connected back to behavior given the complexity of their nonlinear interactions? Model Comparisons Although assumptions and computations differ between PC and DS models, they have similarities, too. Both classes of models rely on interconnections between excitatory and inhibitory neurons. The primary differences concern the role of time and the organizational architecture: The hierarchical representations in PC models include several layers, more types of nodes than DS models (only PC models include modulatory nodes), and more connections than the bottom-up DS oscillations, which are based solely on excitatory-inhibitory interactions. These computational distinctions suggest that the models diverge in parsimony. These distinctions parallel differences in the synchrony behaviors accounted for: The PC models' Bayesian properties account for variability in synchrony, whereas the DS models' delay-coupled differential equations account for mean (directional) synchrony. Although these theories are not mutually exclusive, important paradigmatic and architectural differences prevent an easy merger of them: PC models, evolved from information theory and 152 Palmer, Demos systems neuroscience, make top-down/bottom-up distinctions (such as in Heggli, Konvalinka et al.'s, 2019, mismatched internal models of musical partners). In contrast, DS models, evolved from physics and mathematical biology, include interactions across multiple spatiotemporal scales (such as unidirectional and bidirectional coupling within and across ensemble musicians). In short, predictive coding explains synchrony by generating predictions based on prior learning; dynamical systems explain synchrony by coupling already existing oscillations without a need to generate predictions. Conclusions PC and DS models of musical synchronization share similarities, including the goal to minimize energy; reliance on differences between expected and observed outcomes; and grounding in neurophysiological models of excitatory and inhibitory activation. The models have important differences in how much they rely on prior knowledge about the resulting output, how previous adaptations to stimuli are retained, and whether they are intended to account for the mean or variability of synchrony. To date, only DS theories have successfully modeled

anticipatory synchronization. Future research directions include modeling noisy conditions; explaining roles of contextual learning, musical pleasure, and reward; and scaling up to larger groups. Advances are likely to be assisted by machine learning and other mathematical tools for capturing musical synchrony.

Recommended Reading Balanov, A., Janson, N., Postnov, D., & Sosnovtweva, O. (2009). Synchronization: From simple to complex. Springer. An easy-to-read introduction to synchronization theory from a dynamical-systems perspective. Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. Frontiers in Human Neuroscience, 4, Article 25. https://doi.org/10.3389/ fnhum.2010.00025. A comprehensive, accessible overview of predictive coding. Keller, P. E., Novembre, G., & Loehr, J. (2016). Musical ensemble performance: Representing self, other, and joint action outcomes. In S. S. Obhi & E. S. Cross (Eds.), Shared representations: Sensorimotor foundations of social life (pp. 280–310). Cambridge University Press. A recent synthesis of predictive coding in musical joint action. Kelso, S. A. J. (1995). Dynamic patterns: The self-organization of brain and behavior. MIT Press. A foundational overview of dynamic systems in human behavior. Washburn, A., Kallen, R. W., Lamb, M., Stepp, N., Shockley, K., & Richardson, M. J. (2019). Feedback delays can enhance anticipatory synchronization in human-machine interaction. PLOS ONE, 14(8), Article e0221275. https://doi.org/10 .1371/ journal.pone.0221275. A recent extension of dynamical systems to human-machine interaction.

References Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral & Brain Sciences, 36(3), 181–204. https://doi.org/10.1017/ S0140525X12000477 Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O., & Kotter, R. (2009). Key role of coupling, delay, and noise in resting brain fluctuations. Proceedings of the National Academy of Sciences, USA, 106(25), 10302–10307. https://doi.org/ 10.1073/ pnas.0901831106 Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. Nature Reviews Neuroscience, 12(1), 43–56. https://doi.org/10.1038/nrn2961 Demos, A. P., Layeghi, H., Wanderley, M. M., & Palmer, C. (2019). Staying together: A bidirectional delay–coupled approach to joint action. Cognitive Science, 43(8), Article e12766. https://doi.org/10.1111/ cogs.12766 Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: Challenges and unresolved questions. European Journal of Neuroscience, 51(5), 1151–1160. https://doi.org/10.1111/ejn.13802 Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. Proceedings of the Royal Society B: Biological Sciences, 281(1786), Article 20140751. https://doi.org/10.1098/rspb.2014.0751 Friston, K. (2009). The free-energy principle: A rough guide to the brain? Trends in Cognitive Sciences, 13(7), 293–301. https:// doi.org/10.1016/j.tics.2009.04.005 Explaining Anticipatory Synchrony 153 Friston, K. (2018). Does predictive coding have a future? Nature Neuroscience, 21(8), 1019–1021. https://doi.org/ 10.1038/s41593-018-0200-7 Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51(5), 347–356. https:// doi.org/10.1007/BF00336922 Heggli, O. A., Cabral, J., Konvalinka, I., Vuust, P., & Kringelbach, M. L. (2019). A Kuramoto model of selfother integration across interpersonal synchronization strategies. PLOS Computational Biology, 15(10), Article

e1007422. https://doi.org/10.1371/journal.pcbi.1007422 Heggli, O. A., Konvalinka, I., Kringelbach, M. L., & Vuust, P. (2019). Musical interaction is influenced by underlying predictive models and musical expertise. Scientific Reports, 9, Article 11048. https://doi.org/ 10.1038/s41598- 019-47471-3 Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? Neuroscience, 389, 54–73. https://doi.org/ 10.1016/j.neu roscience.2017.07.061 Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. Trends in Cognitive Sciences, 23(1), 63–77. https:// doi.org/10.1016/j.tics.2018 .10.006 Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), The psychology of time (pp. 189–232). Emerald. Machado, J. N., & Matias, F. S. (2020). Phase bistability between anticipated and delayed synchronization in neuronal populations. Physical Review E, 102(3), Article 032412. https://doi.org/10.1103/PhysRevE.102.032412 Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor pitch singing. Cognitive Psychology, 70, 31–57. https:// doi.org/10.1016/j.cogpsych.2013.12.005 Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). Synchronization: A universal concept in nonlinear sciences. Cambridge University Press. Pyragiene. , T., & Pyragas, K. (2015). Anticipating synchronization in a chain of chaotic oscillators with switching parameters. Physics Letters A, 379(47–48), 3084–3088. https://doi.org/10.1016/j.physleta.2015.10.030 Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extraclassical receptive-field effects. Nature Neuroscience, 2(1), 79–87. https://doi.org/ 10.1038/4580 Ravignani, A., Bowling, D. L., & Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. Frontiers in Psychology, 5, Article 1118. https:// doi.org/ 10.3389/fpsyg.2014.01118 Repp, B. H., & Siu, Y.-H. (2015). Sensorimotor synchronization: A review of recent research (2006–2012). Psychonomic Bulletin & Review, 20(3), 403–452. https:// doi.org/10.3758/s13423-012-0371-2 Roman, I. R., Washburn, A., Large, E. W., Chafe, C., & Fujioka, T. (2019). Delayed feedback embedded in perception-action coordination cycles results in anticipation behavior during synchronized rhythmic action: A dynamical systems approach. PLOS Computational Biology, 15(10), Article e1007371. https://doi.org/10.1371/journal.pcbi.1007371 Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. Topics in Cognitive Science, 1(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009 .01024.x Stepp, N., & Turvey, M. T. (2010). On strong anticipation. Cognitive Systems Research, 11(2), 148–164. https://doi .org/ 10.1016/j.cogsys.2009.03.003 Strogatz, S. H. (1994). Nonlinear dynamics and chaos. Addison-Wesley. Strogatz, S. H. (2000). From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. Physica D: Nonlinear Phenomena, 143(1–4), 1–20. https://doi.org/10.1016/S0167-2789(00)00094-4 Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. Journal of Neuroscience, 37(26), 6331–6341. https:// doi.org/10.1523/JNEUROSCI.2500-16.2017 Tichko, P., Kim, J. C., & Large, E. W. (2021). Bouncing the network: A dynamical systems model of auditory–vestibular interactions underlying infants' perception of musical rhythm. Developmental Science, 24(5), Article e13103. https://doi.org/10.1111/desc.13103 van der Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. Frontiers in Human Neuroscience, 7, Article 253. https://doi.org/10.3389/ fnhum.2013.00253 Varlet, M., Marin, L., Raffard, S., Schmidt, R. C., Capdevielle, D., Boulenger, J.-P., Del-Monte, J., & Bardy, B. G. (2012). Impairments of social motor coordination in schizophrenia. PLOS ONE, 7(1), Article e29772. https://doi.org/10 .1371/ journal.pone.0029772 Voss, H. U. (2000). Anticipating chaotic synchronization. Physical Review E, 61(5), 5115–5119. https://doi.org/ 10.1103/PhysRevE.61.5115 Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model

for understanding rhythmic incongruity. Annals of the New York Academy of Sciences, 1423(1), 19–29. https://doi .org/10.1111/nyas.13622 Vuust, P., & Witek, M. A. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythms and meter perception in music. Frontiers in Psychology, 5, Article 1111. https://doi.org/10.3389/fpsyg.2014.01111 Winfree, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. Journal of Theoretical Biology, 16(1), 15–42. https://doi.org/10.1016/0022-5193 (67)90051-3 Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. Current Biology, 11(18), R729–R732. https://doi .org/10.1016/S0960-9822(01)00432-8 Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880–1882. https://doi.org/10.1126/science .7569931

Body sway reflects leadership in joint music performance

Andrew Chang https://orcid.org/0000-0001-6745-4435, Steven R. Livingstone, Dan J. Bosnyak, and Laurel J. Trainor ljt@mcmaster.ca Authors Info & Affiliations

Significance

People perform tasks in coordination with others in daily life, but the mechanisms are not well understood. Using Granger causality models to examine string quartet dynamics, we demonstrated that musicians assigned as leaders affect other performers more than musicians assigned as followers. These effects were present during performance, when musicians could only hear each other, but were magnified when they could also see each other, indicating that both auditory and visual cues affect nonverbal social interactions. Furthermore, the overall degree of coupling between musicians was positively correlated with ratings of performance success. Thus, we have developed a method for measuring nonverbal interaction in complex situations and have shown that interaction dynamics are affected by social relations and perceptual cues.

Abstract

The cultural and technological achievements of the human species depend on complex social interactions. Nonverbal interpersonal coordination, or joint action, is a crucial element of social interaction, but the dynamics of nonverbal information flow among people are not well understood. We used joint music making in string quartets, a complex, naturalistic nonverbal behavior, as

a model system. Using motion capture, we recorded body sway simultaneously in four musicians, which reflected real-time interpersonal information sharing. We used Granger causality to analyze predictive relationships among the motion time series of the players to determine the magnitude and direction of information flow among the players. We experimentally manipulated which musician was the leader (followers were not informed who was leading) and whether they could see each other, to investigate how these variables affect information flow. We found that assigned leaders exerted significantly greater influence on others and were less influenced by others compared with followers. This effect was present, whether or not they could see each other, but was enhanced with visual information, indicating that visual as well as auditory information is used in musical coordination. Importantly, performers' ratings of the "goodness" of their performances were positively correlated with the overall degree of body sway coupling, indicating that communication through body sway reflects perceived performance success. These results confirm that information sharing in a nonverbal joint action task occurs through both auditory and visual cues and that the dynamics of information flow are affected by changing group relationships.

Sign up for PNAS alerts.

Get alerts for new articles, or get an alert when an article is cited.

MANAGE ALERTS

Coordinating actions with others in time and space—joint action—is essential for daily life. From opening a door for someone to conducting an orchestra, periods of attentional and physical synchrony are required to achieve a shared goal. Humans have been shaped by

evolution to engage in a high level of social interaction, reflected in high perceptual sensitivity to communicative features in voices and faces, the ability to understand the thoughts and beliefs of others, sensitivity to joint attention, and the ability to coordinate goal-directed actions with others (1–3). The social importance of joint action is demonstrated in that simply moving in synchrony with another increases interpersonal affiliation, trust, and/or cooperative behavior in infants and adults (e.g., refs. 4–9). The temporal predictability of music provides an ideal framework for achieving such synchronous movement, and it has been hypothesized that musical behavior evolved and remains adaptive today because it promotes cooperative social interaction and joint action (10–12). Indeed music is used in important situations where the goal is for people to feel a social bond, such as at religious ceremonies, weddings, funerals, parties, sporting events, political rallies, and in the military.

In social contexts, individuals often assume leader and follower roles. For example, to jointly lift a heavy object, one person may verbally indicate the upcoming joint movements. However, many forms of joint action are coordinated nonverbally, such as dancing a tango or performing music in an ensemble. Previous studies have examined nonverbal joint action between two coactors engaged in constrained laboratory tasks in terms of how they adapt to each other (2, 13–16). However, the effect of social roles, such as leader and follower, on group coordination is not well understood, nor are the nonverbal mechanisms by which leader–follower information is communicated, particularly in

ecologically realistic situations involving more than two coactors (2).

Music ensembles provide an ideal model to study leader–follower dynamics. Music exists in all human cultures as a social activity, and group coordination is a universal aspect of music (12, 17), which suggests that music and coordination functions may be linked in the human brain. Ensemble music performance shares many psychological principles with other forms of interpersonal coordination, including walking, dancing, and speaking (18). The principles of coordination in musical ensembles are thus expected to generalize to other interaction situations. Small music ensembles are well suited to studying joint action because they act as self-managed teams, with all performers contributing to the coordination of the team (19, 20).

The complex, nonverbal characteristics of ensemble playing require high sensorimotor, social, and motivational engagement to achieve shared technical and aesthetic goals without explicit verbal guidance, making it a tractable problem for initial study of joint action in ecologically realistic situations. Although musical scores provide some objective guidance for coordination, performers vary tempo, phrasing, articulation, and loudness dynamically to achieve joint musical expression (21–23). Thus, music ensembles provide an opportunity to investigate coordination and adaptation in a complex nonverbal interaction task in which agents work toward a shared goal: an aesthetic performance of music.

Previous studies on music and coordination have focused primarily on note-to-note temporal synchrony

(23), including how it is modulated by the partner (24–29) and the role of perceptual information (30–33). A few studies have attempted to describe temporal synchronization among musicians using mathematical models (34–36). However, we are not aware of any existing studies that have done so using an experimental manipulation of leadership for investigating interpersonal coordination beyond the note-to-note level. Unlike most previous studies, we examined leader–follower interaction dynamics at a more global level than note-to-note timing accuracy. Specifically, we measured body sway in two string quartets. Body sway is a global measurement of an individual's actions (14) not precisely time-locked to the musical notes or the finger and bow movements that are critical to musical sound generation. We used body sway as a reflection of interpersonal communication in achieving a joint aesthetic goal by coordinating performance aspects, such as phrasing, dynamics, timbre, and expressive timing between musicians. We used a bidirectional modeling technique—Granger causality (GC)—to study how musicians form a common musical expression. Leadership assignment was experimentally manipulated, as well as the presence of visual information and compositional style of the music performed, to study changes in information flow between musicians.

Body sway has been used to index underlying mechanisms of joint action as it becomes interpersonally coupled when individuals engage in a coordinated task (37, 38) and during conversation (14, 15). Body sway dynamics are thought to reflect real-time interpersonal

information sharing (14, 16). Previous studies of music ensembles suggest that body sway helps pianists coordinate tempo fluctuations (31) and that the body sway of performers playing the leading melody tend to precede the sway of those who play the accompaniment (30, 39). However, it remains unclear whether body sway is a motor byproduct of music performance, or reflects higher order aspects of joint action, such as leader–follower communication. In the present study, musicians' body sway was recorded with motion capture while we experimentally manipulated leader–follower roles through assignment of a "secret" leader on each trial, with the other musicians not told who the leader was. We hypothesized that body sway dynamics among performers would change as different performers were assigned the leadership role.

Auditory and visual information can contribute to coordinating joint actions, particularly in the absence of verbal communication (2). Even when visual information is not directly relevant, visual cues can improve interpersonal coordination (14, 40–42). In piano duos, being able to see the other performer increases temporal synchrony (31). Here, we examined how visual information affects leader–follower dynamics. We manipulated visual information by having the musicians face each other (visual present) or face away from one another (visual absent). We hypothesized that the influence of leader–follower dynamics would be higher in the visual-present condition, indicating that visual information, in addition to auditory information, is used to facilitate coordination.

We extended previous research on interpersonal

coordination by examining directional information flow among group members using Granger causality analysis. Granger causality is well-suited to the analysis of interdependent time-series data because it quantifies how well the history of one time series predicts the current status of a second time series, after taking into account how much the time series is predicted by its own previous history, in the form of a log-likelihood ratio referred to as GC (Fig. 1*B*) (43, 44). The larger the value of GC, the better the prediction and the more information is flowing from one time series to another. In the context of music ensembles, Granger causality will reveal both the direction and magnitude of information flow among performers. Previous studies primarily used cross-correlation to examine the similarity between movement time series. However, this measure is not sensitive to the direction of information flow between agents and may result in type I errors if time series are autocorrelated (45). A few previous observational studies on music ensembles used Granger causality to analyze bow movements, timbre variations, and body sway of performers (39, 46, 47). However, because these studies did not manipulate leadership assignment or visual information, it remains unclear whether previous findings reflect a motor byproduct of executing music scores or are related to higher order aspects of joint action, such as leader–follower communication.

Fig. 1.

Illustrations of the experimental design and Granger causality analyses. (A) Top-down view of the locations of performers on stage. (B) Example excerpt of recorded anterior–posterior body sway motion time series in four performers from the middle of a trial. The Granger causality (GC) of the body sway of violin 1 directionally coupling to (or predicting) violin 2, for example, was calculated by taking the log-likelihood ratio of the degree to which the prior body sway time series of violin 1 (predictor 1, shaded in light blue) contributes to predicting the current status of violin 2 (red dot), over and above the degree to which it is predicted by its own prior time series

(predictor 2, shaded in light yellow), while conditional on the prior time series of the other performers (predictors 3 and 4, shaded in gray). This calculation was repeated along the entire time axis to estimate GC. The length of the predictor window (shaded areas) was determined by the model order. The algorithm is conceptually expressed by the equation shown (see ref. 43 for mathematical details). (C) Categorization of the 12 directional relationship pairs between the four performers into three Paired-Roles. We calculated the GCs of directional body sway couplings (using the formula outlined in B) for all 12 directional pairs of performers (shown as arrows) in each quartet as shown on the left. For Single-Leader-Role trials, the 12 GCs were then categorized according to whether the Paired-Role was leader-to-follower (LF), follower-to-leader (FL), or follower-to-follower (FF). For example, if violin 2 was assigned as the single leader while the others were followers in a trial, the arrows (blue) coming out from violin 2 were categorized as LF, the arrows (red) pointing at violin 2 as FL, and the other arrows (green) as FF. For Ambiguous-Role trials, all 12 GCs were categorized as either all leaders (L-all) or all followers (F-all) (not shown in the figure). We treated the 12 unique directional pairs in a quartet on each trial as 12 unique samples for repeated-measures statistical analyses.

In the present study, we investigated information flow in two professional string quartets, manipulating leadership assignment and the presence of visual information. Performers' body sways were recorded with motion capture (Fig. 1*A*) while they performed works that varied in compositional style. Different performers were secretly assigned leader or follower roles on different trials (Fig. 1*C*). On all trials, musicians were told that one performer would be assigned the role of quartet leader. On Single-Leader-Role trials, only one of the four musicians was assigned as leader and the others as followers. We hypothesized that leaders would have greater influence on the body sway of followers than vice versa or than between followers, indexed by Granger causality. On Ambiguous-Role trials, either all four performers were assigned as leaders, or all four performers were assigned as followers, contrary to the expectations of the musicians. This assigning was done to examine whether information flow (indexed by Granger causality) between musicians would increase over the performance of a musical excerpt as the musicians established a joint musical interpretation. We

also examined the role of visual information by comparing Granger causality when the musicians faced each other (Seeing) or faced away (Nonseeing). We hypothesized that the influence of leader–follower dynamics would be higher in the visual present (Seeing) condition, indicating that visual information, in addition to auditory information, is used for predicting other musicians' intentions. We replicated the effects in two separate string quartets playing two different styles of music. One quartet played musical pieces from the Baroque period (quartet 1), in which the different parts are fairly equal in importance, whereas the second quartet played music from the Classical period (quartet 2). Because music of the Classical period more strongly assigns the leader role to one player (typically the first violinist) (33, 34), replicating the effect of leadership assignment on information flow in Classical period music would provide strong support for our conclusions.

Results

Quartet 1.

In quartet 1, the string quartet played Baroque music excerpts in which all four parts—violin 1, violin 2, viola, and cello—were fairly equal in importance.

Subjective identification of the leader.

In Single-Leader-Role conditions, one member of the quartet was assigned as the leader and the rest as followers. For Single-Leader-Role trials (including both Seeing and Nonseeing conditions), on average 2.94 ± 0.25 (mean ± SD) out of the three performers who were followers on each trial correctly identified who the leader was, and a Wilcoxon signed-rank test showed that this

result exceeded chance levels (one of three followers correctly identifying the leader) with $P < 0.001$. Thus, followers were highly accurate at identifying who the secretly assigned leader was on each trial, suggesting that experimentally manipulated leader–follower relationships were successfully built during music performances.

Analyses of body sway coupling in Single-Leader-Role conditions.

A two-way repeated-measures analysis of variance (rANOVA) was conducted on GC scores, with Paired-Role [leader-to-follower (LF), follower-to-leader (FL), and follower-to-follower (FF)] and Vision (Seeing, Nonseeing) as within-subjects factors.

The results of quartet 1 (Fig. 2A) showed a significant interaction [$F(2, 22) = 14.70$, $P < 0.001$, $f2 = 0.38$]. The simple effect analyses further showed that Paired-Role modulated GC under both Seeing [$F(2, 22) = 37.75$, $P < 0.001$, $f2 = 1.02$] and Nonseeing [$F(2, 22) = 26.03$, $P < 0.001$, $f2 = 0.70$] conditions.

Fig. 2.

GC of interpersonal body sway couplings. Comparisons are marked as *P < 0.05, **P < 0.01, ***P < 0.001. Error bars represent SE. (A and B) GCs of Single-Leader-Role trials are presented in (A) for quartet 1 and (B) for quartet 2. Both experiments showed a significant interaction between Paired-Role and Vision. Specifically, the LF (leader-to-follower) coupling was higher than both FL (follower-to-leader) and FF (follower-to-follower) couplings when performers could see each other, but this effect was attenuated when performers could not see each other. Also, the GC of LF coupling was higher when performers could see each other than when they could not. These results show that GC reflects leader–follower relationships, and seeing or not seeing others specifically mediates the FL coupling. (C and D) GC scores at the beginnings and ends of Ambiguous-Role trials are presented in (C) for quartet 1 and (D) for quartet 2. In quartet 1 (Baroque music), there was a significant three-way interaction (Time × Role × Vision). Specifically, GC increased from the first 30 s to the last 30 s of pieces performed when all performers were assigned as followers and they could see each other. The same analyses on quartet 2 (Classical music) did not show any significant effects.

Follow-up *t* tests among different levels of Paired-Role

showed that, in the Seeing condition, GC was higher in the LF condition than in the FF [$t(11) = 6.18$, $P < 0.001$] and FL [$t(11) = 7.33$, $P < 0.001$] conditions, but GC was not different between FF and FL conditions [$t(11) = 1.73$, $P = 0.111$]. In the Nonseeing condition, GC in the LF condition was higher than in the FF [$t(11) = 3.73$, $P = 0.003$] and FL [$t(11) = 7.54$, $P < 0.001$] conditions, and GC in the FF condition was higher than in the FL condition [$t(11) = 3.29$, $P = 0.007$].

Additionally, GC was higher in Seeing than Nonseeing under the LF condition [$t(11) = 7.69$, $P < 0.001$], but Vision did not significantly modulate GC under the FF [$t(11) = 1.92$, $P = 0.081$] or FL [$t(11) = 1.80$, $P = 0.099$] conditions.

In sum, the results showed an interaction between Paired-Role and Vision on GC of body sway coupling between performers. Specifically, the body sway of the leader predicted those of the followers (LF coupling) to a greater extent than the other directional couplings, follower-to-follower and follower-to-leader (FF and FL, respectively). Being able to see the other performers specifically facilitated the LF coupling but not the other pairings.

Analyses of body sway coupling in Ambiguous-Role conditions.

On the Ambiguous-Role trials, either all performers were assigned the leader role (L-all), or all were assigned the follower role (F-all). Our main hypothesis was that performers would nonverbally build a joint coordinated pattern over the course of a musical excerpt. A three-way rANOVA was performed with Time (Beginning,

End), Ambiguous-Role (L-all, F-all), and Vision (Seeing, Nonseeing) as factors.

Results of quartet 1 (Fig. 2C) showed a significant three-way interaction [$F(1, 11) = 6.85$, $P = 0.024$, $f_2 = 0.06$]. Under the L-all condition, the simple rANOVA did not show any significant effects (P values > 0.06). Under the F-all condition, there was a significant interaction between Vision and Time [$F(1, 11) = 10.65$, $P = 0.008$, $f_2 = 0.20$]. Paired t tests further showed that the GC was larger in the End than the Beginning of the trial in the Seeing condition [$t(11) = 2.86$, $P = 0.016$], but not in the Nonseeing condition [$t(11) = -2.14$, $P = 0.055$]. In sum, under Ambiguous-Role assignment, there was some evidence of learning over the course of a trial in the case where all performers were assigned follower roles, but there was no evidence of this learning when all performers were assigned leader roles. It is possible that three minutes is simply not long enough to see robust changes in coupling strength in musical performance under these conditions.

Correlations between body sway coupling and subjective ratings of performances.

We examined whether body sway coupling among quartet performers was associated with their subjective ratings of how well the quartet achieved the task goal (Fig. 3). Specifically, each performer rated three aspects of the performance after each trial: (*i*) the overall goodness of the ensemble performance, (*ii*) the temporal synchronization of the group, and (*iii*) the ease of coordination among performers.

Fig. 3.

Correlations between subjective ratings and group body sway couplings. (A, C, and E) Scatter plots of the results of quartet 1. (B, D, and F) Scatter plots of the results of quartet 2. On each trial, each performer subjectively rated the levels of goodness of performance (A and B), temporal synchronization (C and D), and ease of coordination (E and F), with five-point Likert scales, and we calculated the mean rating across the performers for each of these three aspects for each trial. To represent the group causal density (44) of body sway, we calculated the mean of the 12 GCs reflecting the pairwise influences for each trial, which reflects the overall causal interactivity sustained in a quartet. The Spearman rank correlation tests on all trials (n = 24) showed that the group body sway coupling was positively correlated with goodness of performance level in both experiments and also positively correlated with temporal synchronization level in quartet 1 but not in quartet 2.

Across trials, causal density, the mean GC of overall interpersonal body sway coupling (44), was positively correlated with the performers' subjective ratings of goodness [$r_s(23) = 0.49$, 95% CI (0.11, 0.75), P_{perm} (P value of permutation test) = 0.015] and synchronization [$r_s(23) = 0.58$, CI (0.23, 0.80), P_{perm} = 0.003], but it did not correlate significantly with performance difficulty [$r_s(23) = 0.39$, CI (–0.02, 0.69), P_{perm} = 0.059].

Quartet 2.

The procedure with quartet 1 was repeated in a second string quartet that played Classical music excerpts in which one instrument (most often the first violin) usually has the most important melodic part; however, we instructed the musicians to attempt to perform the piece using their assigned roles. We performed the same analyses as in quartet 1, summarized here (*SI Results* for detailed results).

Subjective identification of the leader.

On Single-Leader-Role trials, the leader was correctly identified on average by 2.44 ± 0.63 out of the three followers, which is higher than chance levels ($P <$ 0.001).

Analyses of body sway coupling in Single-Leader-Role

conditions.

The analyses of body sway coupling in Single-Leader-Role conditions (Fig. 2*B*) replicated the results of quartet 1 in that Paired-Role and Vision interactively modulated GC [$F(2, 22) = 4.43$, $P = 0.024$, $f_2 = 0.10$]. The *t* tests further showed that, in the Seeing condition, GC was higher in the LF than FL [$t(11) = 3.18$, $P = 0.009$] and FF [$t(11) = 2.36$, $P = 0.038$] conditions, and GC for the LF condition was higher in the Seeing than Nonseeing condition [$t(11) = 2.57$, $P = 0.026$].

Analyses of body sway coupling in Ambiguous-Role conditions.

On the Ambiguous-Role trials (Fig. 2*D*), the three-way rANOVA did not find any significant effect (*P* values > 0.128), and the *t* tests on changes in GC between the Beginning and the End of the performance were not significant within any Paired-Role × Vision condition [unsigned $t(11) < 2.14$, *P* values > 0.056], suggesting that, with Classical music, the performers were not able to improve the information flow within 3 min when faced with an ambiguous role assignment.

Correlations between body sway coupling and subjective ratings of performances.

The Spearman correlational analyses between total body sway coupling within a quartet and subjective ratings of performances (Fig. 3) showed that the GC causal density positively correlated with the subjective rating of how good the performance was across trials [$r_s(23) = 0.54$, CI (0.17, 0.77), $P_{perm} = 0.009$], as in quartet 1. Correlations with rated synchronization level

(P_{perm} = 0.108) and performance difficulty (P_{perm} = 0.079) were not significant.

Differences Between Baroque and Classical Music Styles.

We hypothesized that, although the basic findings would be similar across quartets and musical styles, because Classical music (played in quartet 2) tends to have one instrument playing the melody and the others playing accompaniment, unlike the Baroque music in quartet 1, the quartets would adopt somewhat different coordinative strategies. This unbalanced default leadership hypothesis is supported in that (*i*) a Mann–Whitney *U* test showed that more followers correctly identified the assigned leader when Baroque music was played compared with Classical music (z = 2.45, P = 0.007), and (*ii*) the Kruskal–Wallis H tests showed that the baseline GCs (i.e., without regard to who was the assigned leader) among the performers in each quartet were more unequal when playing Classical music [$\chi^2(3)$ = 18.2, $P < 0.001$, f^2 = 0.52] than Baroque music [$\chi^2(3)$ = 7.98, P = 0.046, f^2 = 0.13]. See *SI Results* and Fig. S1 for detailed results.

Fig. S1.

*The baseline-GC of each performer without regard to leadership role is presented in (A) for playing Baroque music in quartet 1 and (B) for playing Classical music in quartet 2. The greater difference across instruments for quartet 2 compared with quartet 1 likely reflects the different musical styles of the pieces played by the two quartets, with the quartet playing Baroque music having intrinsically more equal parts across instruments compared with the quartet playing Classical music, in which one of the violins tends to have the melody at a particular point in time. Thus, this difference likely indicates that the musical structure of the pieces plays a role in social leadership interactions in addition to effects of assigned leadership role. Significant comparisons are marked as *P < 0.05, **P < 0.01. Error bars represent SE.*

Discussion

The results showed that anterior–posterior body sway couplings among string quartet performers reflected

nonverbal interpersonal coordination, leadership roles, and performers' subjective evaluations of their coordinative performance. Assigned leaders influenced followers more than followers influenced leaders or than one follower influenced another, and this effect was larger when performers could see each other than when they could not. Moreover, the degree of total coupling within an ensemble was positively associated with the rated goodness of the performance, suggesting that body sway reflects the communication of information critical to performance success. We replicated these effects in two string quartets playing music of different compositional styles. The findings extend our understanding of interpersonal coordination by examining interactions between more than two people in a naturalistic complex coordinative task through experimental manipulation of leadership assignment and visual information. The use of Granger causality enabled the asymmetric coordination structure among coactors to be disentangled, revealing "who leads whom."

Our methods enabled examination of nonverbal joint action in a naturalistic setting with four coactors. Through this approach, we were able to examine both overall coordination and the coordination structure associated with leader–follower, follower–leader, and follower–follower roles. We also showed that visual information selectively facilitates leader-to-follower coupling among four coactors in a string quartet, suggesting that visual communication is most useful for coordinating with a leader. Although previous studies examined how perceptual or motor manipulations affect precise movement synchronization between coactors at

the note-to-note level (e.g., ref. 34), we used body sway to examine interaction at the more global level of creating a joint aesthetic expression, including factors such as synchronization, phrasing, and dynamics. Furthermore, although previous studies provided descriptions of leadership dynamics in music ensembles (20, 39, 46), our study experimentally manipulated leadership orthogonally to characteristics of the individual musical parts played by each instrument. This manipulation allowed us to show that information flow changes with leadership assignment and that body sway is not a motor byproduct of music performance, but reflects nonverbal communication between members of a group.

An important contribution of the present study is the use of Granger causality over more traditional cross-correlation approaches (45). The use of Granger causality analysis enabled us to examine body sway relationships after partialing out predictions within each performer. This handling of auto-correlated series is critical because within-performer movements may be similar across performers due to the use of a common musical score. Furthermore, we found that cross-correlation analyses of our data were not able to reveal leader–follower effects evident with the Granger causality approach (*SI Results* and Fig. S2). Importantly, the Granger causality approach enabled us to conclude that the coupling relations reflected actual directional information flow between performers.

Fig. S2.

Mean body sway temporal asynchrony of leader preceding follower, as identified by the maximum unsigned cross-correlation coefficient. The y axis reflects the maximum cross-correlation, with scores above 0 indicating that the leader preceded the follower and scores

Critical to the idea that body sway coupling reflects interpersonal coordination (14), we found that total interpersonal body sway coupling in a performance correlated with the performers' rated goodness of that ensemble performance. Previous studies showed that, when two participants converse during a coordinative task, their body sways tend to unintentionally and spontaneously couple (14, 15). The present study extends this finding to a nonverbal musical context and shows that body sway interactions are related to the rated outcome success of the joint action. Although high-level ensemble performers who have played together for many years are experienced and critical evaluators of the quality of their performances, it would be interesting to investigate relations between total interpersonal body sway coupling and objective measures of performance quality (e.g., acoustic analyses) and ratings by expert and nonexpert audience members.

Although musical performance has often been characterized as involving auditory–motor coordination, basic research on individuals has shown that visual information can help anticipation of auditory information (48, 49), which could be useful for auditory-guided motor coordination. Supporting this view, previous piano duet studies showed that anticipatory gazing at the head motion of a coactor improved accuracy of

synchronization during expressive tempo variations (31, 32) and that seeing a partner facilitated temporal synchronization in a jazz ensemble (50). Our findings further show that visual information is particularly helpful for leader-to-follower coordination, suggesting that followers use visual information to anticipatively coordinate their own actions to the leader's performance. We also found that, when no one was assigned as the leader in a musical style without strong intrinsic leader roles, body sway couplings among performers increased from the beginning to the end of each piece when they could see each other, suggesting that exchanged visual information helped build up interpersonal auditory–motor coordination in real time. Together, these findings indicate that visual information facilitates interpersonal information flow over and above auditory information as a channel for nonverbal communication and improves auditory–motor coordination among coactors.

Despite the overall similarity of the results for quartets 1 and 2, the few differences that emerged suggest that characteristics of the task—specifically, how strongly the musical score suggests a specific leader—modulate group coordination. Previous studies in nonmusical domains indicated that the nature of a coordinative task can modulate interpersonal coordinative structure (e.g., ref. 51), which is consistent with our findings that the secretly assigned leader was more accurately identified for Baroque than Classical performances. Baseline body sway couplings without regard to leadership role were more equal across musician dyads for Baroque than Classical music (*SI Results* and Fig. S1), suggesting

that the relative importance of the four parts intrinsic to the musical score affected interpersonal coordination. Furthermore, total interpersonal body sway coordination was related only to rated temporal synchronization between performers for Baroque but not for Classical music. It is possible that different musical styles shift the coordinative emphases from timing synchronization to other common aesthetic goals, such as coordinated loudness or timbre, which were not rated explicitly in the current study. Further studies are needed to directly examine these issues.

Another question requiring study is how interpersonally exchanged information between ensemble members guides motor coordination within each performer. When listening to an auditory beat, low frequency oscillations in the EEG phase lock with the incoming beats (52). Furthermore, the power of beta band (~20 Hz) oscillations in both auditory and motor regions fluctuates at the tempo of the auditory rhythm and reflects temporal prediction (53–55). These oscillatory activities might also play a role in the prediction of the intentions and actions of others (21, 56, 57).

Given that music is universal across human societies and is used extensively to promote nonverbal social coordination, it is important to understand how musicians achieve a joint aesthetic goal. Furthermore, the findings of the current study using a music ensemble as a model can be generalized and applied to other forms of interpersonal coordination. Evidence to date indicates that music performance shares the same principles with many other forms of interpersonal coordination (e.g., walking, dancing, and speaking):

People tend to coordinate with each other's movements in joint actions where there is a common goal, and interpersonal coordination involves perceptual, motor, and social factors (see ref. 18 for a review). Thus, understanding how leader–follower roles, sensory information, and structural elements of the situation affect interpersonal coordination in music performance is critical for understanding joint action in general. Furthermore, our method of using Granger causality to uncover bidirectional information flow under experimental manipulation in a naturalistic situation can be applied to understanding group behavior in various situations, such as the important ability to detect who does or does not belong to a particular social group (58), and determining best crowd control procedures for emergency evacuation (59), which are important topics in psychology, computer vision, and public safety.

In conclusion, the present study showed that manipulation of leadership roles and visual information interactively modulated interpersonal coordination in string quartets across styles of music played, as reflected by interpersonally coupled body sways indexed by Granger causality. The coupling degree was positively associated with the success of coordinative performance, and leader-to-follower influence was highest when performers could see each other. Importantly, these results were obtained in a complex naturalistic environment with more than two individuals, while, at the same time, exercising high experimental control of the factors of interest.

Methods

Participants.

Two internationally recognized professional music ensembles participated: the Cecilia String Quartet (four females; mean age = 30.5 y; range = 30–34), and the Afiara Quartet (three male, one female; mean age = 32 y; range = 29–33). Participants were right-handed except for the second violin of the Cecilia Quartet and the cello of the Afiara Quartet. Participants had normal hearing and were neurologically healthy by self-report. Consent was obtained from each participant, and they received reimbursement. The McMaster University Research Ethics Board approved all procedures.

Stimuli and Apparatus.

The data were collected in the McMaster Large Interactive Virtual Environment laboratory (LIVELab) (LIVELab.mcmaster.ca). Quartet 1 (Cecilia Quartet) performed 12 different chorales (Table S1) from the Baroque period (~1600 to 1750), composed by J. S. Bach, each between 2 and 3 min long. Quartet 2 (Afiara Quartet) also performed 12 different pieces (Table S2), each between 2.5 and 5 min long, from the Classical period (~1730 to 1820), composed by J. Haydn and W. A. Mozart. The Bach chorales consisted of four parallel melodies, which were relatively equal in importance compared with music from the Classical period, in which one instrument typically has the main melody whereas the others play a supportive accompaniment (34). The musicians did not rehearse the pieces, nor had they previously performed the pieces together. They did not verbally discuss the pieces although they had their individual parts ahead of the experiment.

Table S1.

The trial order of conditions and the Baroque music pieces played in quartet 1

Block-trial order	Leader	Vision	Bach chorale no.
1-1	Violin 2	Seeing	9
1-2	Cello	Seeing	5
1-3	Viola	Seeing	8
1-4	Violin 1	Seeing	4
1-5	All followers	Seeing	7
1-6	All leaders	Seeing	14
2-1	Violin 2	Nonseeing	5
2-2	Cello	Nonseeing	4
2-3	Viola	Nonseeing	9
2-4	Violin 1	Nonseeing	14
2-5	All followers	Nonseeing	8
2-6	All leaders	Nonseeing	7
3-1	All leaders	Nonseeing	2
3-2	All followers	Nonseeing	15
3-3	Violin 1	Nonseeing	11
3-4	Viola	Nonseeing	3
3-5	Cello	Nonseeing	1
3-6	Violin 2	Nonseeing	13
4-1	All leaders	Seeing	15
4-2	All followers	Seeing	3

Block-trial order	Leader	Vision	Music piece
4-3	Violin 1	Seeing	2
4-4	Viola	Seeing	13
4-5	Cello	Seeing	11
4-6	Violin 2	Seeing	1

EXPAND FOR MORE

Table S2.

The trial order of conditions and the Classical music pieces played in quartet 2

Block-trial order	Leader	Vision	Music piece
1-1	Violin 1	Nonseeing	Mozart: K. 158, mvt. 3
1-2	Viola	Nonseeing	Haydn: Op. 20, No. 1, mvt. 2
1-3	Cello	Nonseeing	Haydn: Op. 17, No. 4, mvt. 2
1-4	Violin 2	Nonseeing	Haydn: Op. 17, No. 2, mvt. 2
1-5	All followers	Nonseeing	Mozart: K. 169, mvt. 3
1-6	All leaders	Nonseeing	Haydn: Op. 17, No. 1, mvt. 2
2-1	Violin 1	Seeing	Haydn: Op. 20, No. 1, mvt. 2
2-2	Viola	Seeing	Haydn: Op. 17, No. 2, mvt. 2
2-3	Cello	Seeing	Mozart: K. 158, mvt. 3
2-4	Violin 2	Seeing	Haydn: Op. 17, No. 1, mvt. 2
2-5	All followers	Seeing	Haydn: Op. 17, No. 4, mvt. 2
2-6	All leaders	Seeing	Mozart: K. 169, mvt. 3
3-1	All leaders	Seeing	Mozart: K. 156, mvt. 3
3-2	All followers	Seeing	Mozart: K. 170, mvt. 2

3-3	Violin 2	Seeing	Haydn: Op. 20, No. 6, mvt. 3
3-4	Cello	Seeing	Mozart: K. 171, mvt. 3
3-5	Viola	Seeing	Mozart: K. 168, mvt. 3
3-6	Violin 1	Seeing	Mozart: K. 172, mvt. 3
4-1	All leaders	Nonseeing	Mozart: K. 170, mvt. 2
4-2	All followers	Nonseeing	Mozart: K. 171, mvt. 3
4-3	Violin 2	Nonseeing	Mozart: K. 156, mvt. 3
4-4	Cello	Nonseeing	Mozart: K. 172, mvt. 3
4-5	Viola	Nonseeing	Haydn: Op. 20, No. 6, mvt. 3
4-6	Violin 1	Nonseeing	Mozart: K. 168, mvt. 3

EXPAND FOR MORE

K., Köchel number; mvt., movement; Op., Opus.

An optical motion capture system (24 Oqus 5+ cameras and an Oqus 210c video camera; Qualisys) recorded the head movements of participants at 179 Hz. Four retroreflective markers (3 mm) were placed symmetrically on headbands. Each participant wore the headband around the forehead, located above the nose, eyes, and occipital bone. Because string quartets sit in a semicircle, we recorded the motion of the head as an index of the global upper-body sway (14, 15). Additional EEG and EKG electrodes were also placed on the chest and head of each participant (data not reported here). It was confirmed by the performers that these placements did not constrain their body movements and that they were able to perform as usual.

Design and Procedure.

A full factorial Role (violin 1 leader, violin 2 leader, viola leader, cello leader, all leader, all follower) × Vision (Seeing, Nonseeing) experimental design was used, collapsing over musical pieces. In the Seeing condition, performers faced the center of the quartet whereas, in the Nonseeing condition, each faced 180 degrees away from the center, with all participants unable to see one another (Fig. 1A). The orders of Vision and Role levels were orthogonally counterbalanced across blocks. Each of the 12 musical pieces was performed twice, once in each Vision condition, yielding 24 trials. Each piece was played in two different Role conditions so the exact same trial was never repeated (Tables S1 and S2 show the complete design).

Performers were given confidential sheets that assigned them as either leader or follower on each trial. Performers were informed that there was one leader and three followers in all trials. Four Role levels were Single-Leader-Role conditions, in that one performer was the leader and the others followers (leader was either violin 1, violin 2, viola, or cello), and two were Ambiguous-Role conditions (all four were leaders, all four were followers). Performers were instructed to perform at their best within the role assigned to them. All trials were performed on the same day for quartet 1, and the first and the second halves of the trials were performed 2 d apart for quartet 2.

To prevent the assigned leader from explicitly initiating the performance, trials began with three metronome tones (interonset interval, 875 ms for all pieces played by quartet 1; and 500 ms for pieces played by quartet 2)

via speakers. Performers were instructed to follow the tempo and start on the first downbeat after the last tone. Because some pieces were shorter in duration than others, if necessary, each piece was performed repeatedly until the trial reached 2 min for quartet 1, and 2.5 min for quartet 2.

A potential concern was that experimentally manipulating leadership might distort natural body sway, but the evidence suggests that such distortion was not the case. First, performers were not told that body sway was the variable of interest. Second, Mann–Whitney U tests showed that leader/follower role assignment did not significantly modulate any of range, variation, or total moving distance of body sway within any performer of any quartet (P values > 0.379), suggesting that body sway was not unnaturally exaggerated or attenuated by experimentally manipulated leadership. So, manipulating leadership seems to have influenced only the relationships between players' movements, not the movements themselves.

After each trial, each performer rated three aspects of the group's performance using a five-point Likert scale (–2 to 2): (*i*) the overall goodness of the ensemble performance, (*ii*) the temporal synchronization (taking all levels, from note-to-note to musical phrase, into consideration) of the group, and (*iii*) the ease of coordination among performers. Given the high level of the musicians and the fact that each quartet had worked together for many years with regular intense rehearsal and concert schedules, we expected that they would be sensitive evaluators of these variables. Performers who

were not assigned as the leader on a particular trial were also asked to identify who they thought was the assigned leader using a forced-choice categorical response measure.

Data Processing and Granger Causality.

For each trial, the recorded motion trajectories were de-noised, spatially averaged, down-sampled, *z*-score–normalized, and projected to the anterior–posterior body orientation (*SI Methods* for details) to produce four body sway time series, one for each performer (Fig. 1*B*). The Matlab Multivariate Granger Causality Toolbox (43) was used to estimate the magnitude of Granger causality (GC) between each pair of body sway time series among all four performers in each quartet (Fig. 1*B*). GC is a statistical estimation of the magnitude of how much one time series is predicted by the history of another time series, taking into account how much it is predicted by its own previous history, in the form of a log-likelihood ratio. The larger the value of GC, the better the prediction and the more information is flowing from one time series to another. It is important to note that we estimated each GC between two time series conditional on the remaining two time series because, in this way, any potential common influence on other variables was partialed out (43). In this way, 12 unique GCs were obtained from each trial, corresponding to the degree to which each of violin 1, violin 2, viola, and cello predicted each of three other performers (Fig. 1*C* and *SI Methods* for details).

Statistical Analyses.

The 12 directional couplings ($n = 12$) between the

various pairs of performers were considered independent (43) and treated as a random factor in a within-subjects analysis (Fig. 1*C*). The Single-Leader-Role trials and the Ambiguous-Role trials were analyzed separately. For the Single-Leader-Role analyses, the factor Paired-Role had three levels [leader-to-follower (LF), follower-to-leader (FL), and follower-to-follower (FF)], and Vision had two levels (Seeing and Nonseeing). For each of the 12 directional couplings for each trial, we first estimated the GC, and then we took the mean GCs of the trials belonging to each Vision (Seeing, Nonseeing) × Paired-Role (LF, FL, FF) level as an estimate for each condition. For the Ambiguous-Role analyses, Paired-Role had two levels (Leader-all and Follower-all), Vision had two levels (Seeing and Nonseeing), and Time had two levels (Beginning and End). The Time level was extracted from the first and last 30-s epochs from each performance (trial). Again, we took the mean GCs of the trials belonging to each Vision × Paired-Role × Time level for each directional coupling.

We performed repeated-measures analysis of variance (rANOVA) on GC scores for analyses of body sway coupling in Single-Leader-Role and Ambiguous-Role conditions. Follow-up post hoc analyses were performed for each significant *F*-test.

For analyzing correlations between body sway coupling and subjective ratings of performances across all 24 trials, we correlated the mean of the four performers' subjective ratings (one missing easy/difficulty rating on a trial by a performer of quartet 1 was excluded from averaging) with the group causal density (44) of body

sways, calculated as the mean of the 12 GCs reflecting all pairwise influences in a quartet, for each trial. For each correlation analysis, we estimated the Spearman rank correlation coefficient (r_s) on the nonnormally distributed data, using Fisher's z transformation to estimate the 95% confidence interval (CI) of r_s, and used a permutation test (5,000 times) to calculate the P value (P_{perm}) to adjust for the ties in the ranked data.

We used χ^2 tests to check the normality assumption for parametric tests, the Mauchly test to check the sphericity assumption, and the Brown–Forsythe test to check the homogeneity of variances assumption. When the assumption of normality was violated ($P < 0.05$), nonparametric test equivalents were used (specified in *Results*). Every statistical test was performed two-tailed, if applicable. We set $\alpha = 0.05$, and each Bonferroni-adjusted α was used for each post hoc comparison series as a conservative control for type I error; we report the tests with Bonferroni-adjusted $\alpha < P < 0.05$ as trends.

SI Methods

Details of Data Processing and Granger Causality.

Motion capture data processing.

Motion trajectories were exported from Qualisys Track Manager for processing and analysis in MATLAB. The first 120 s of each piece were analyzed for quartet 1 (Cecilia Quartet), and the first 150 s of each piece were analyzed for quartet 2 (Afiara Quartet). Missing data due to recording noise was found in only 1 of 384 trajectories (0.06% of the epoch length) of quartet 1, and 3 of 384 trajectories (0.11 to 0.17% of the epoch length) of

quartet 2, and were filled with spline interpolation. Each trajectory was down-sampled to 8 Hz by spatially averaging the samples within each nonoverlapped 125-ms window because Granger causality analysis prefers a low model order for capturing a given physical time length of the movement trajectory (43). This rate is sufficient for capturing most of the head movements, confirmed by visual inspection. No filtering or temporal smoothing was applied to the data because temporal convolution distorts the estimation of Granger causality (43). To estimate the anterior–posterior body sway, we spatially averaged the positions of the four motion capture markers on the head of each performer in the x–y plane (collapsing altitude) for each time frame, and the anterior–posterior orientation was referenced to the center of the half circle formed by the locations of the performers (Fig. 1A). Finally, each time series was normalized (z-score) to equalize the magnitude of the sway motion among performers. This procedure produced four normalized body sway time series, one for each performer for each trial (Fig. 1B).

Granger causality.

We followed the procedure and used the algorithms of the Multivariate Granger Causality (MVGC) Toolbox (43) in MATLAB to estimate the magnitude of Granger causality (GC) between each pair of body sway time series among all four performers in each quartet. First, the MVGC toolbox confirmed that each time series passed the stationary assumption for Granger causality analysis, with the spectral radius less than 1. Second, the optimal model order (the length of history included)

was determined by the Akaike information criterion on each trial. The optimal model order is a balance between maximizing goodness of fit and minimizing the number of coefficients (length of the time series) being estimated. The model order used was 13 (1,625 ms, 1.86 beats) for quartet 1, and 9 (1,125 ms, 2.25 beats) for quartet 2 because these were the largest optimal model orders across trials within each quartet. Model order was a fixed parameter within quartet (i.e., did not vary by trial), which avoided model order artificially affecting GCs on different trials, and the largest model order across trials covered all optimal model orders across trials. Finally, 12 unique GCs were obtained from each trial, corresponding to the degrees to which each of violin 1, violin 2, viola, and cello predicted each of three other performers. It is important to note that each GC estimated, representing how well the time series of one performer predicted that of another performer, was made conditional on how well the time series of that performer predicted those of the other two performers. This conditional estimation is optional for estimating GC, and we chose to do so because this way partials out any potential common dependencies on other variables (43), which fit our purpose of quantifying the pairwise information flows among performers of a quartet.

SI Results

Details of Results for Quartet 2.

Analyses of body sway coupling in Single-Leader-Role conditions.

As for quartet 1, we used an rANOVA to investigate whether Paired-Role (LF, FL, FF) and Vision (Seeing,

Nonseeing) modulated the coupling strength of anterior–posterior body sway motions among performers (Fig. 1B). The results showed an interaction between Vision and Paired-Role [$F(2, 22) = 4.43$, $P = 0.024$, $f2 = 0.10$]. Follow-up post hoc simple effect analyses showed a simple effect of Paired-Role in the Seeing condition [$F(2, 22) = 5.78$, $P = 0.010$, $f2 = 0.13$]; specifically, post hoc paired t tests showed that GC was higher in the LF than FL condition [$t(11) = 3.18$, $P = 0.009$] and a trend for GC to be higher in LF than FF condition [$t(11) = 2.36$, $P = 0.038$, Bonferroni-adjusted α = 0.05/3], but GC in the FF condition was not different from in the FL [$t(11) = 0.59$, $P = 0.564$] condition. In contrast, there was no effect of Paired-Role in the Nonseeing condition [$F(2, 22) = 1.18$, $P = 0.325$]. Orthogonal post hoc paired t tests showed that the GC was not modulated by Vision for either FL [$t(11) = -0.41$, $P = 0.687$] or FF [$t(11) = -1.17$, $P = 0.268$] conditions. There was, however, a trend for GC to be higher in Seeing than Nonseeing for the LF condition [$t(11) = 2.57$, $P = 0.026$, Bonferroni-adjusted α = 0.05/3].

In sum, despite playing music differing in style, the quartets of both experiments 1 and 2 showed an interaction between Paired-Role and Vision on GC of body sway coupling between performers. Specifically, the body sway of the leader predicted those of the followers (LF coupling) to a greater extent than for other directional couplings. Also, seeing other performers specifically facilitated the LF coupling but not other pairings. However, these effects seemed to be weaker in the case of Classical music (quartet 2) than Baroque music (quartet 1), which is investigated further below.

Analyses of body sway coupling in Ambiguous-Role conditions.

As for quartet 1, we extracted the first and last 30-s epochs from each trial and compared whether the GC of body sway couplings improved from the beginning to end of each performance (Fig. 2D). A three-way rANOVA performed with Time (Beginning, End), Ambiguous-Role (L-all, F-all), and Vision (Seeing, Nonseeing) as factors revealed no significant main effects or interaction effects (P values > 0.12).

Correlations between body sway coupling and subjective ratings of performances.

As for quartet 1, we examined whether body sway coupling among quartet performers reflected their subjective ratings of how well the quartet achieved the task goals (Fig. 3). The Spearman rank correlation results showed that the causal density positively correlated with the subjective rating of how good the performance was [$r_s(23) = 0.54$, CI (0.17, 0.77), $P_{perm} = 0.009$], but not with how synchronized [$r_s(23) = 0.34$, CI (–0.07, 0.65), $P_{perm} = 0.108$] or how easy [$r_s(23) = 0.36$, CI (–0.05, 0.67), $P_{perm} = 0.079$] the performance was.

Differences Between Quartets.

We hypothesized that, because Classical music tends to have one instrument playing the melody and the others accompaniment (33–35, 60, 61) whereas, in Baroque music, the parts are more equal in importance, the quartets would adopt somewhat different coordinative strategies. To estimate the extent to which each performer acted as a leader when assigned a follower

role, termed Leadership Baseline, we took the mean GCs of body sway coupling of each performer in predicting each of the other three performers [e.g., baseline-$GC_{Violin1}$ = ($GC_{Violin1 \rightarrow Violin2}$ + $GC_{Violin1 \rightarrow Viola}$ + $GC_{Violin1 \rightarrow cello}$)/3], when that performer was assigned as a follower in Single-Leader-Role trials. We then compared the baseline-GC among performers in each quartet (Fig. S1).

Kruskal–Wallis H tests (nonparametric alternative of a one-way between-subject ANOVA) showed that the performers had different baseline-GC in both quartets [quartet 1, $\chi_2(3)$ = 7.98, P = 0.046, f_2 = 0.13; quartet 2, $\chi_2(3)$ = 18.2, $P < 0.001$, f_2 = 0.52]. In the quartet playing Baroque music (quartet 1) (Fig. S1A), however, post hoc Mann–Whitney U tests (Bonferroni-adjusted α = 0.05/6) did not show any significant baseline-GC differences between players (P values > 0.035), but there were trends for the baseline-GC of violin 1 to be higher than those of violin 2 (z = 2.11, P = 0.035) and the viola (z = 1.99, P = 0.046), and for the baseline-GC of the cello to be higher than that of violin 2 (z = 2.11, P = 0.035). In the quartet playing Classical music (quartet 2) (Fig. S1B), on the other hand, the post hoc U tests showed that the baseline-GC of violin 1 was higher than that of the viola (z = 2.97, P = 0.003) and that the baseline-GC of violin 2 was higher than those of the viola (z = 3.20, P = 0.001) and cello (z = 2.86, P = 0.004) although the other comparison failed to reach threshold (P values > 0.212). These results confirmed our hypothesis that the default leadership was more unequal when playing Classical compared with Baroque music.

Analyses with Cross-Correlation.

Cross-correlation estimates the similarity between two time series as a function of a shifting time step and has been used for quantifying the preceding-lag relationship between two time series variables. To empirically compare Granger causality and cross-correlation, we performed cross-correlation analyses on the same preprocessed data to which we had applied Granger causality. If cross-correlation is able to reflect the leading–following relationship between performers in the Single-Leader-Role trials, the cross-correlation analysis on body sway would show the highest similarity in the time step where leader precedes follower.

For each trial of each quartet, the cross-correlation was estimated between each pair of the four performers' anterior–posterior body sway time series. To match the Granger causality analyses, the cross-correlation coefficients were calculated for lags up to plus or minus the model order of each quartet. Twelve samples (the random factor) were obtained under each of the Seeing and Nonseeing conditions for each quartet, corresponding to the mean asynchrony by which each performer (violin 1, violin 2, viola, or cello) led each of the other three performers. If cross-correlation reflects leading–following relationships between performers, the maximum unsigned cross-correlation coefficient (highest similarity) would occur at the temporal asynchrony where the leader's time series precedes the follower's. The cross-correlation results are shown in Fig. S2. One-sample t tests showed that the asynchrony between leader and follower's body sways were not significantly different from 0 in either Seeing or Nonseeing conditions

of either quartet [unsigned $t(11) < 1.35$, P values > 0.205]. The paired t tests further showed that the asynchrony was not different between Seeing and Nonseeing conditions of either quartet [unsigned $t(11) < 0.93$, P values > 0.375]. In sum, the results showed that the cross-correlational analysis was not powerful enough to reflect the leading–following relationship or the effect of visual information, as was revealed by Granger causality analyses.